Présentation et accès aux bases de données OMIM, GDB, EMBL/GenBank, SWISS-PROT, PDB et au système d'information IMGT

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

17 Octobre 2006

Rôles des bases de données

1 - Collecter les informations

Dans la littérature
Auprès des biologistes
Dans les autres bases de données

2 - Stocker et organiser les données

Dans une logique cohérente
(outils informatiques)

3 - Distribuer l'information

Distribuer des données contrôlées
Assurer une large diffusion

4 - Faciliter l'exploitation des données

Interfaces conviviales
Définition des critères de recherche
Recherche, comparaison de données

Quelques sources d'information

1 - Des exemples de serveurs WWW :

IGH	http://www.igh.cnrs.fr
PASTEUR	http://bioweb.pasteur.fr/intro-uk.html
PBIL	http://pbil.univ-lyon1.fr/
EBI	http://www.ebi.ac.uk
DKFZ	http://genome.dkfz-heidelberg.de
ExPASy	http://www.expasy.ch
NCBI	http://www.ncbi.nlm.nih.gov/
TIGR	http://www.tigr.org

Quelques sources d'information

2 - La littérature

Le premier *Nucleic Acids Research* de l'année
est dédié aux bases de données en Biologie

3 - DBCAT: The Public Catalog of Databases

maintenu par INFOBIOGEN
(activités interrompues en Juillet 2006

4 - Catalogue de logiciels

Biocatalog: http://www.ebi.ac.uk/biocat
(606 entrées - fin 2004)

SoftwareSeek: http://genamics.com/software/index.htm

Bases de données de gènes et cartographies

OMIM **NCBI, USA** (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM/)

Catalogue des gènes humains et des désordres génétiques et cliniques

GDB **The Genome Database,** RTI, Caroline du nord (http://www.gdb.org/)

Cartographie des gènes humains (gènes, polymorphismes, mutations, sondes, cartes génétiques, GenBank)

Entrez gene **NCBI, USA**

(http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene/)

Collections de liens (locus, cartographie, medline, OMIM, Entrez, GenBank, Unigene, polymorphisme) associés à chacun des gènes étudiés

GeneCards TM, **Weizmann Institute, Israël**

(http://www.genecards.org/index.shtml/)

Base de données de gènes automatiquement générée qui intègre les informations sur les gènes, leurs produits, et leurs implications dans les maladies

GENATLAS **Paris,** France (http://www.dsi.univ-paris5.fr/genatlas/)

Collection d'informations et de données (gènes, maladies, marqueurs)

OMIM Online Mendelian Inheritance in Man
Base de données de gènes et maladies genetiques

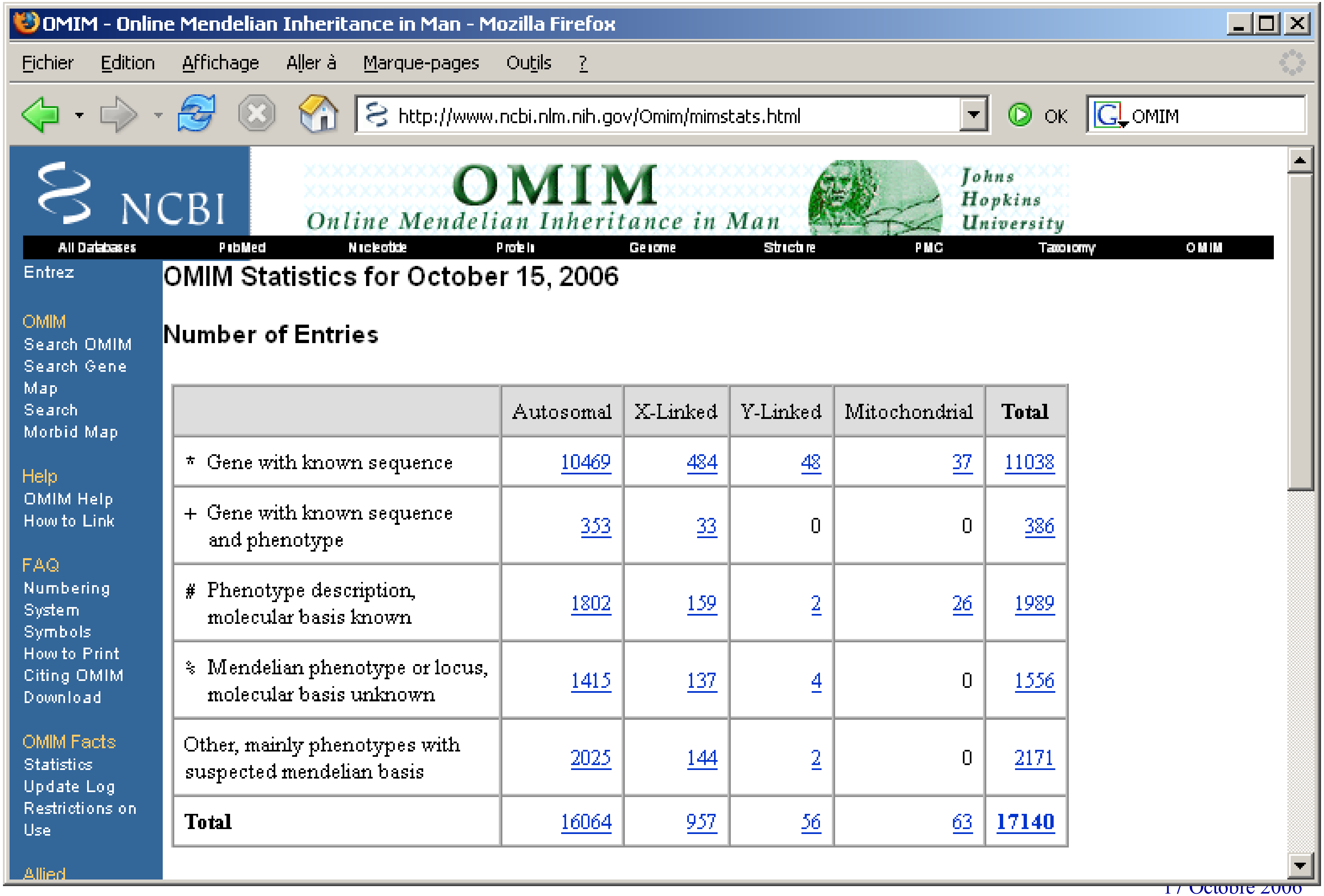
Créée en 1968 par Dr. Victor A. McKusick

Localisation Maintenue à l'Université Johns Hopkins, Baltimore,
Maryland (USA)
L’interface WWW est développée par le NCBI
(National Center for Biotechnology Information)

Regroupe: - une base de données : catalogue de maladies génétiques et de gènes associés
- des cartes cytogénétiques des gènes décrits dans la base de données (recherche par gènes, localisation chromosomiques ou par maladies)

Contient **actuellement 17140 entrées (Octobre 2006)**

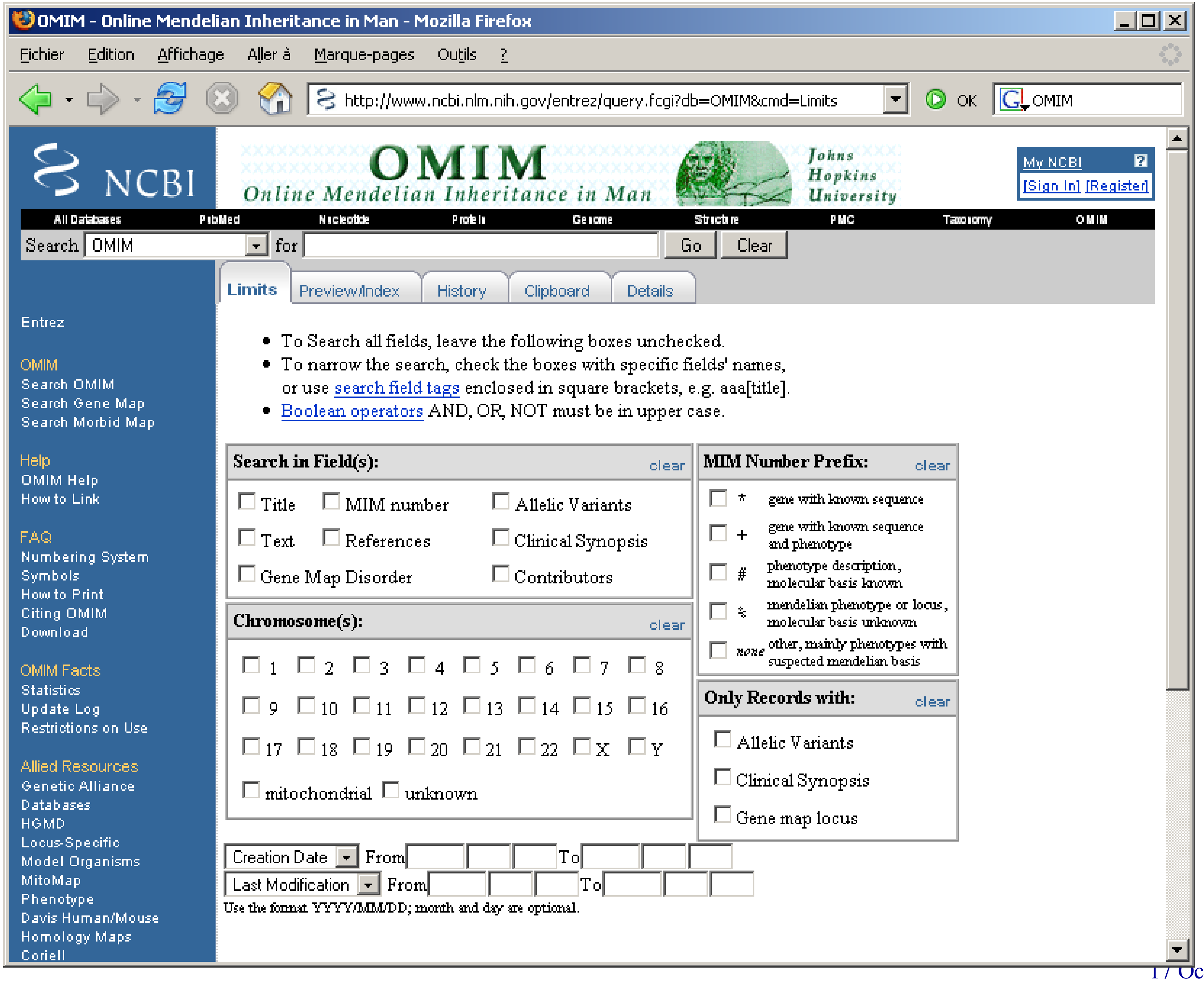
Distribution **http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM**



OMIM Statistics for October 15, 2006

Number of Entries

	Autosomal	X-Linked	Y-Linked	Mitochondrial	**Total**
* Gene with known sequence	10469	484	48	37	11038
+ Gene with known sequence and phenotype	353	33	0	0	386
# Phenotype description, molecular basis known	1802	159	2	26	1989
% Mendelian phenotype or locus, molecular basis unknown	1415	137	4	0	1556
Other, mainly phenotypes with suspected mendelian basis	2025	144	2	0	2171
Total	16064	957	56	63	**17140**



- To Search all fields, leave the following boxes unchecked.
- To narrow the search, check the boxes with specific fields' names, or use search field tags enclosed in square brackets, e.g. aaa[title].
- Boolean operators AND, OR, NOT must be in upper case.

Search in Field(s): clear

☐ Title ☐ MIM number ☐ Allelic Variants
☐ Text ☐ References ☐ Clinical Synopsis
☐ Gene Map Disorder ☐ Contributors

Chromosome(s): clear

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8
☐ 9 ☐ 10 ☐ 11 ☐ 12 ☐ 13 ☐ 14 ☐ 15 ☐ 16
☐ 17 ☐ 18 ☐ 19 ☐ 20 ☐ 21 ☐ 22 ☐ X ☐ Y
☐ mitochondrial ☐ unknown

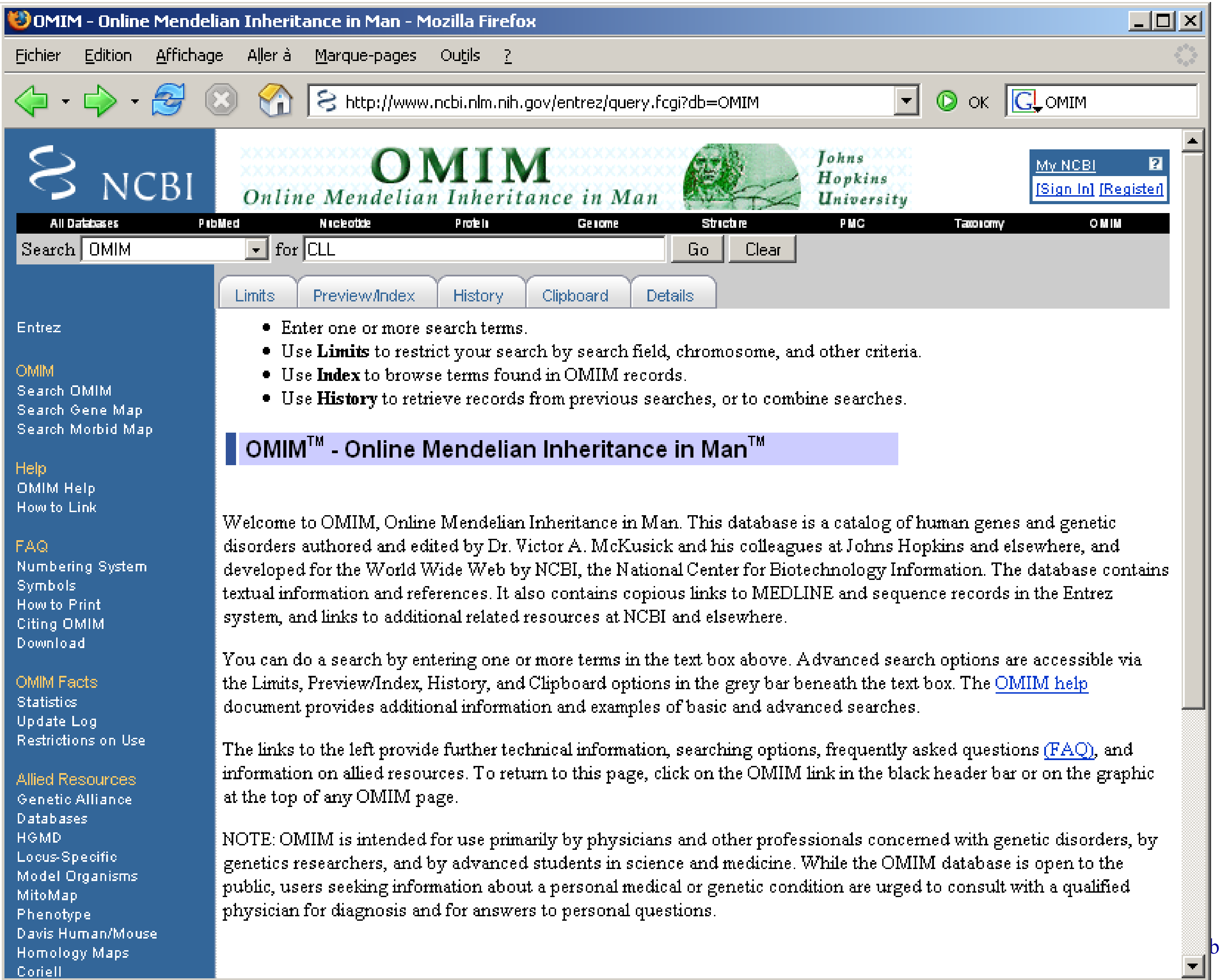
MIM Number Prefix: clear

☐ * gene with known sequence
☐ + gene with known sequence and phenotype
☐ # phenotype description, molecular basis known
☐ % mendelian phenotype or locus, molecular basis unknown
☐ *none* other, mainly phenotypes with suspected mendelian basis

Only Records with: clear

☐ Allelic Variants
☐ Clinical Synopsis
☐ Gene map locus

Creation Date From ☐ To ☐
Last Modification From ☐ To ☐
Use the format YYYY/MM/DD; month and day are optional.



http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM

Search: OMIM

Search OMIM for CLL

- Enter one or more search terms.
- Use **Limits** to restrict your search by search field, chromosome, and other criteria.
- Use **Index** to browse terms found in OMIM records.
- Use **History** to retrieve records from previous searches, or to combine searches.

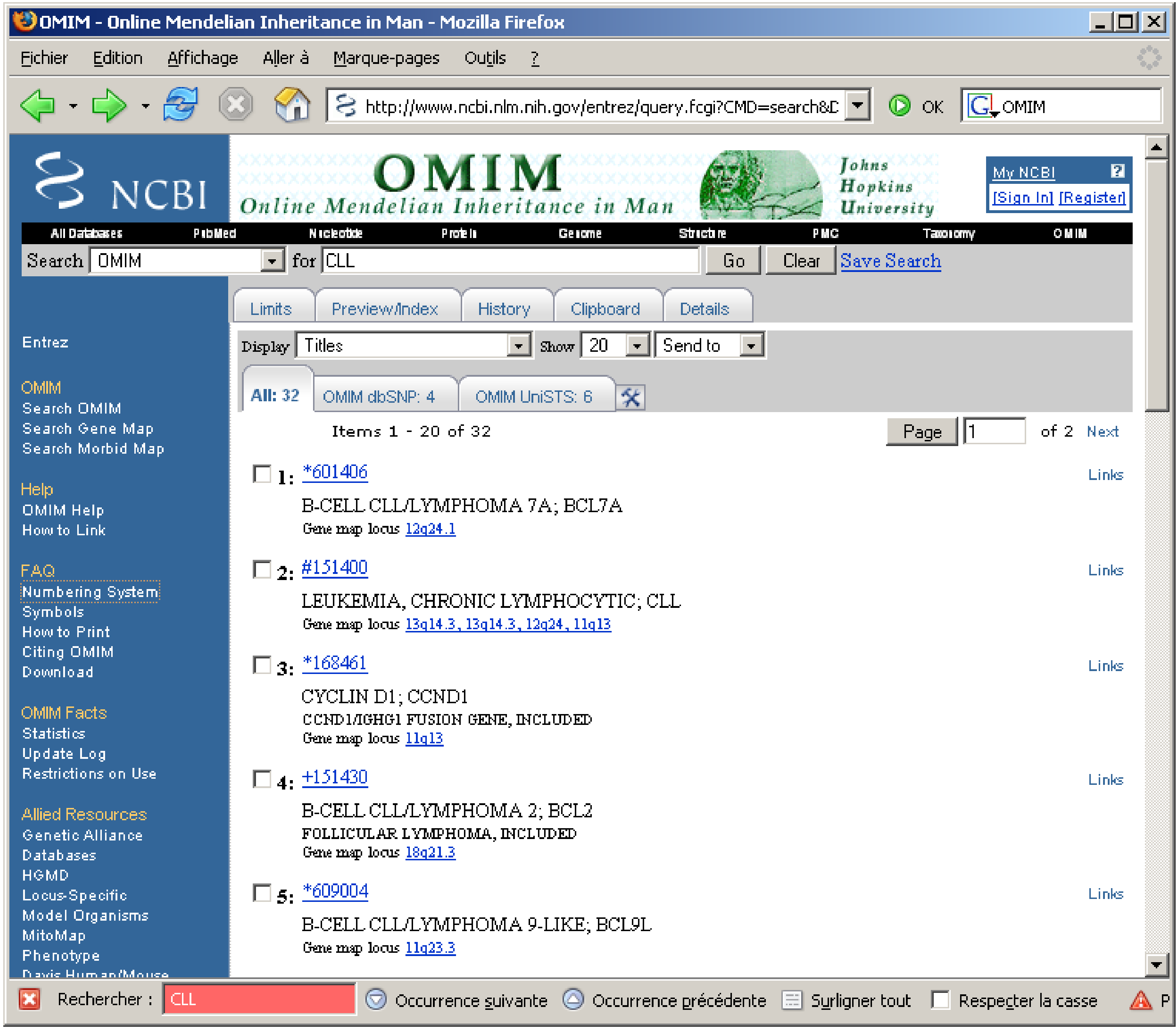
OMIM™ - Online Mendelian Inheritance in Man™

Welcome to OMIM, Online Mendelian Inheritance in Man. This database is a catalog of human genes and genetic disorders authored and edited by Dr. Victor A. McKusick and his colleagues at Johns Hopkins and elsewhere, and developed for the World Wide Web by NCBI, the National Center for Biotechnology Information. The database contains textual information and references. It also contains copious links to MEDLINE and sequence records in the Entrez system, and links to additional related resources at NCBI and elsewhere.

You can do a search by entering one or more terms in the text box above. Advanced search options are accessible via the Limits, Preview/Index, History, and Clipboard options in the grey bar beneath the text box. The OMIM help document provides additional information and examples of basic and advanced searches.

The links to the left provide further technical information, searching options, frequently asked questions (FAQ), and information on allied resources. To return to this page, click on the OMIM link in the black header bar or on the graphic at the top of any OMIM page.

NOTE: OMIM is intended for use primarily by physicians and other professionals concerned with genetic disorders, by genetics researchers, and by advanced students in science and medicine. While the OMIM database is open to the public, users seeking information about a personal medical or genetic condition are urged to consult with a qualified physician for diagnosis and for answers to personal questions.



http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&D

Search OMIM for CLL

Display Titles Show 20

All: 32 OMIM dbSNP: 4 OMIM UniSTS: 6

Items 1 - 20 of 32

Page 1 of 2

☐ 1: *601406
B-CELL CLL/LYMPHOMA 7A; BCL7A
Gene map locus 12q24.1

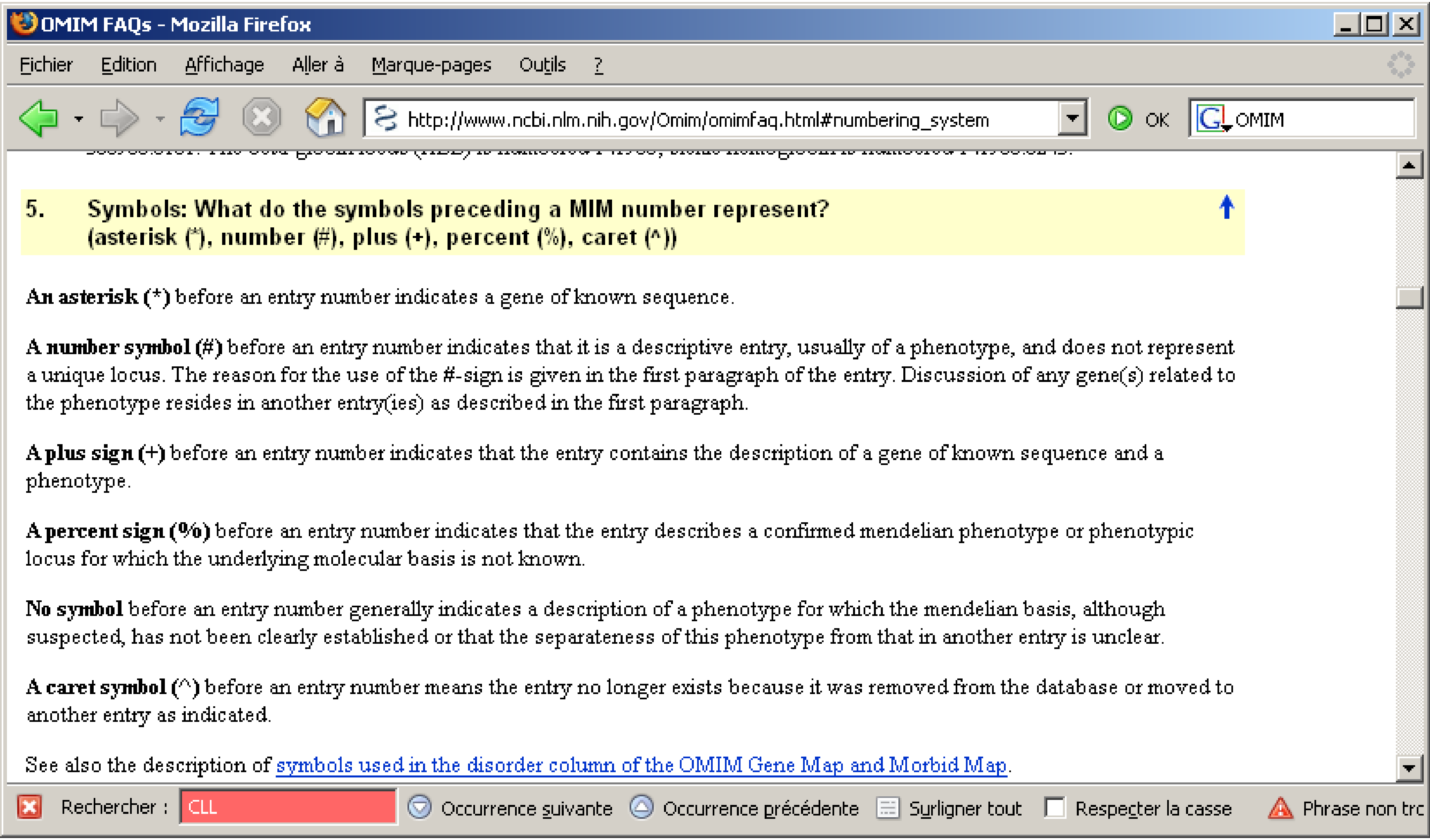
☐ 2: #151400
LEUKEMIA, CHRONIC LYMPHOCYTIC; CLL
Gene map locus 13q14.3, 13q14.3, 12q24, 11q13

☐ 3: *168461
CYCLIN D1; CCND1
CCND1/IGHG1 FUSION GENE, INCLUDED
Gene map locus 11q13

☐ 4: +151430
B-CELL CLL/LYMPHOMA 2; BCL2
FOLLICULAR LYMPHOMA, INCLUDED
Gene map locus 18q21.3

☐ 5: *609004
B-CELL CLL/LYMPHOMA 9-LIKE; BCL9L
Gene map locus 11q23.3

Rechercher : CLL



http://www.ncbi.nlm.nih.gov/Omim/omimfaq.html#numbering_system

5. Symbols: What do the symbols preceding a MIM number represent? (asterisk (*), number (#), plus (+), percent (%), caret (^))

An asterisk (*) before an entry number indicates a gene of known sequence.

A number symbol (#) before an entry number indicates that it is a descriptive entry, usually of a phenotype, and does not represent a unique locus. The reason for the use of the #-sign is given in the first paragraph of the entry. Discussion of any gene(s) related to the phenotype resides in another entry(ies) as described in the first paragraph.

A plus sign (+) before an entry number indicates that the entry contains the description of a gene of known sequence and a phenotype.

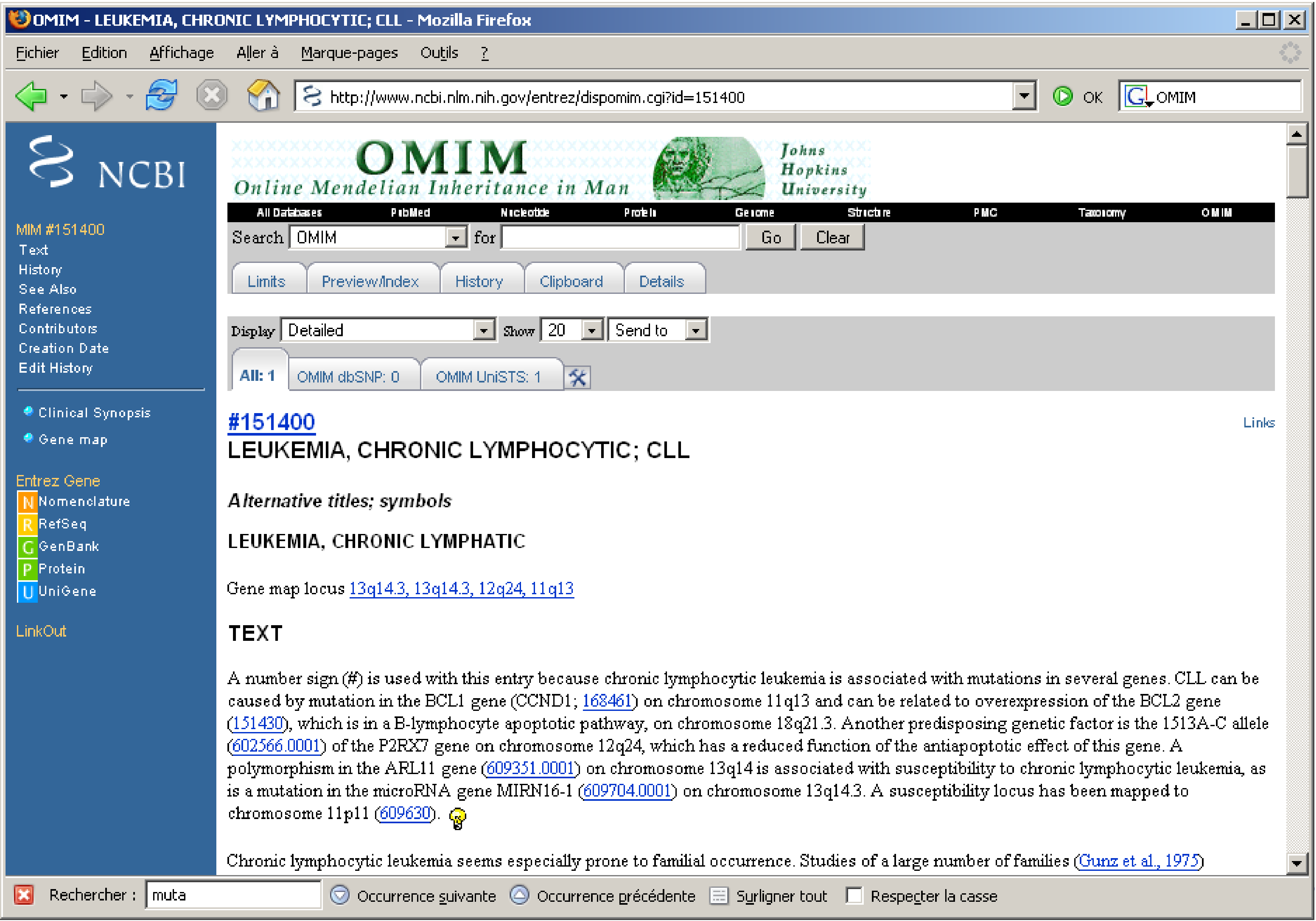
A percent sign (%) before an entry number indicates that the entry describes a confirmed mendelian phenotype or phenotypic locus for which the underlying molecular basis is not known.

No symbol before an entry number generally indicates a description of a phenotype for which the mendelian basis, although suspected, has not been clearly established or that the separateness of this phenotype from that in another entry is unclear.

A caret symbol (^) before an entry number means the entry no longer exists because it was removed from the database or moved to another entry as indicated.

See also the description of symbols used in the disorder column of the OMIM Gene Map and Morbid Map.

Rechercher : CLL



#151400

LEUKEMIA, CHRONIC LYMPHOCYTIC; CLL

Alternative titles; symbols

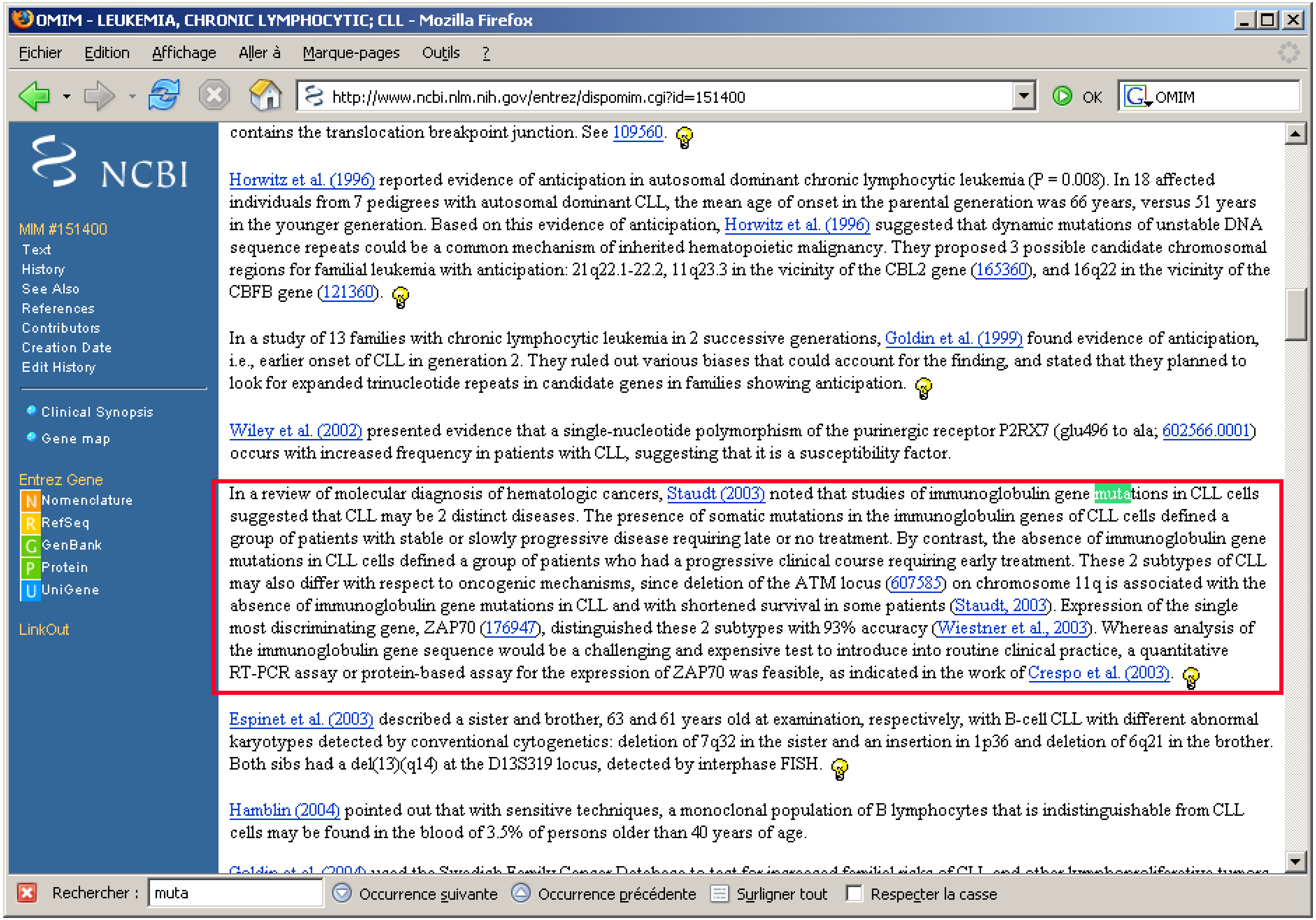
LEUKEMIA, CHRONIC LYMPHATIC

Gene map locus 13q14.3, 13q14.3, 12q24, 11q13

TEXT

A number sign (#) is used with this entry because chronic lymphocytic leukemia is associated with mutations in several genes. CLL can be caused by mutation in the BCL1 gene (CCND1; 168461) on chromosome 11q13 and can be related to overexpression of the BCL2 gene (151430), which is in a B-lymphocyte apoptotic pathway, on chromosome 18q21.3. Another predisposing genetic factor is the 1513A-C allele (602566.0001) of the P2RX7 gene on chromosome 12q24, which has a reduced function of the antiapoptotic effect of this gene. A polymorphism in the ARL11 gene (609351.0001) on chromosome 13q14 is associated with susceptibility to chronic lymphocytic leukemia, as is a mutation in the microRNA gene MIRN16-1 (609704.0001) on chromosome 13q14.3. A susceptibility locus has been mapped to chromosome 11p11 (609630).

Chronic lymphocytic leukemia seems especially prone to familial occurrence. Studies of a large number of families (Gunz et al., 1975)



contains the translocation breakpoint junction. See 109560.

Horwitz et al. (1996) reported evidence of anticipation in autosomal dominant chronic lymphocytic leukemia (P = 0.008). In 18 affected individuals from 7 pedigrees with autosomal dominant CLL, the mean age of onset in the parental generation was 66 years, versus 51 years in the younger generation. Based on this evidence of anticipation, Horwitz et al. (1996) suggested that dynamic mutations of unstable DNA sequence repeats could be a common mechanism of inherited hematopoietic malignancy. They proposed 3 possible candidate chromosomal regions for familial leukemia with anticipation: 21q22.1-22.2, 11q23.3 in the vicinity of the CBL2 gene (165360), and 16q22 in the vicinity of the CBFB gene (121360).

In a study of 13 families with chronic lymphocytic leukemia in 2 successive generations, Goldin et al. (1999) found evidence of anticipation, i.e., earlier onset of CLL in generation 2. They ruled out various biases that could account for the finding, and stated that they planned to look for expanded trinucleotide repeats in candidate genes in families showing anticipation.

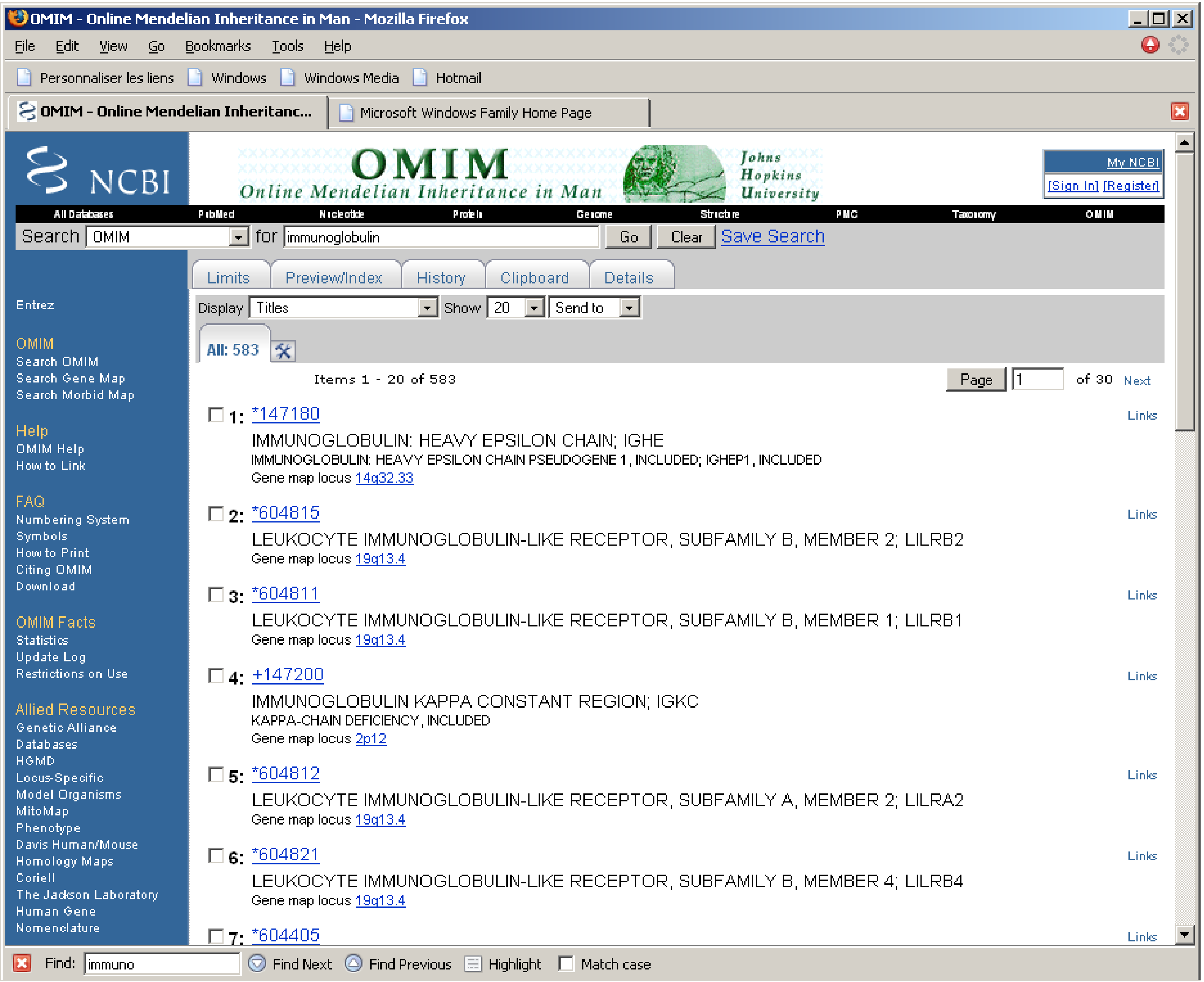
Wiley et al. (2002) presented evidence that a single-nucleotide polymorphism of the purinergic receptor P2RX7 (glu496 to ala; 602566.0001) occurs with increased frequency in patients with CLL, suggesting that it is a susceptibility factor.

In a review of molecular diagnosis of hematologic cancers, Staudt (2003) noted that studies of immunoglobulin gene mutations in CLL cells suggested that CLL may be 2 distinct diseases. The presence of somatic mutations in the immunoglobulin genes of CLL cells defined a group of patients with stable or slowly progressive disease requiring late or no treatment. By contrast, the absence of immunoglobulin gene mutations in CLL cells defined a group of patients who had a progressive clinical course requiring early treatment. These 2 subtypes of CLL may also differ with respect to oncogenic mechanisms, since deletion of the ATM locus (607585) on chromosome 11q is associated with the absence of immunoglobulin gene mutations in CLL and with shortened survival in some patients (Staudt, 2003). Expression of the single most discriminating gene, ZAP70 (176947), distinguished these 2 subtypes with 93% accuracy (Wiestner et al., 2003). Whereas analysis of the immunoglobulin gene sequence would be a challenging and expensive test to introduce into routine clinical practice, a quantitative RT-PCR assay or protein-based assay for the expression of ZAP70 was feasible, as indicated in the work of Crespo et al. (2003).

Espinet et al. (2003) described a sister and brother, 63 and 61 years old at examination, respectively, with B-cell CLL with different abnormal karyotypes detected by conventional cytogenetics: deletion of 7q32 in the sister and an insertion in 1p36 and deletion of 6q21 in the brother. Both sibs had a del(13)(q14) at the D13S319 locus, detected by interphase FISH.

Hamblin (2004) pointed out that with sensitive techniques, a monoclonal population of B lymphocytes that is indistinguishable from CLL cells may be found in the blood of 3.5% of persons older than 40 years of age.

Goldin et al. (2004) used the Swedish Family-Cancer Database to test for increased familial risks of CLL and other lymphoproliferative tumors



Search OMIM for immunoglobulin

Display Titles Show 20

All: 583

Items 1 - 20 of 583

Page 1 of 30

☐ 1: *147180
IMMUNOGLOBULIN: HEAVY EPSILON CHAIN; IGHE
IMMUNOGLOBULIN: HEAVY EPSILON CHAIN PSEUDOGENE 1, INCLUDED; IGHEP1, INCLUDED
Gene map locus 14q32.33

☐ 2: *604815
LEUKOCYTE IMMUNOGLOBULIN-LIKE RECEPTOR, SUBFAMILY B, MEMBER 2; LILRB2
Gene map locus 19q13.4

☐ 3: *604811
LEUKOCYTE IMMUNOGLOBULIN-LIKE RECEPTOR, SUBFAMILY B, MEMBER 1; LILRB1
Gene map locus 19q13.4

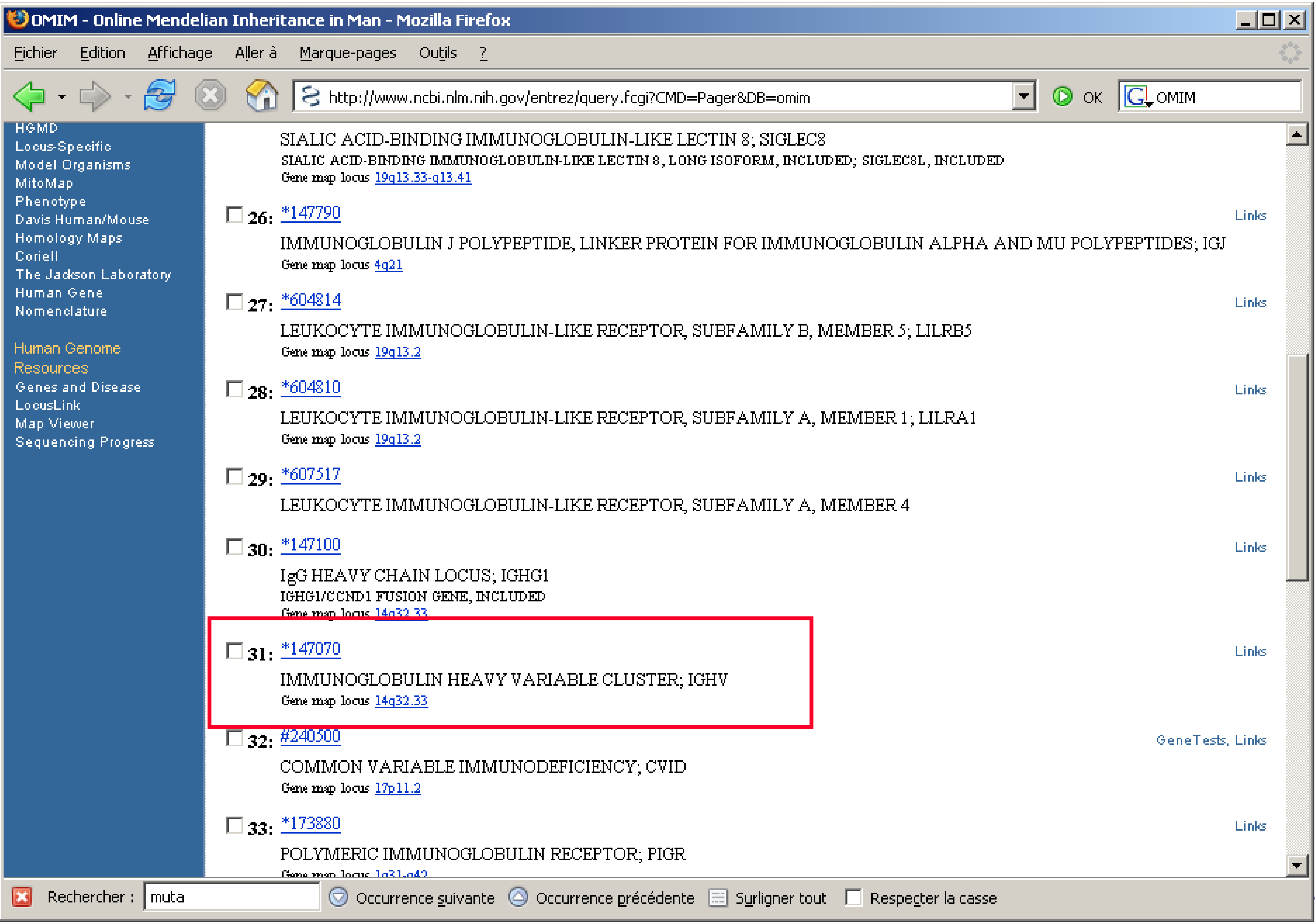
☐ 4: +147200
IMMUNOGLOBULIN KAPPA CONSTANT REGION; IGKC
KAPPA-CHAIN DEFICIENCY, INCLUDED
Gene map locus 2p12

☐ 5: *604812
LEUKOCYTE IMMUNOGLOBULIN-LIKE RECEPTOR, SUBFAMILY A, MEMBER 2; LILRA2
Gene map locus 19q13.4

☐ 6: *604821
LEUKOCYTE IMMUNOGLOBULIN-LIKE RECEPTOR, SUBFAMILY B, MEMBER 4; LILRB4
Gene map locus 19q13.4

☐ 7: *604405

Find: immuno



http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Pager&DB=omim

Rechercher : muta

SIALIC ACID-BINDING IMMUNOGLOBULIN-LIKE LECTIN 8; SIGLEC8
SIALIC ACID-BINDING IMMUNOGLOBULIN-LIKE LECTIN 8, LONG ISOFORM, INCLUDED; SIGLEC8L, INCLUDED
Gene map locus 19q13.33-q13.41

☐ **26:** *147790 Links
IMMUNOGLOBULIN J POLYPEPTIDE, LINKER PROTEIN FOR IMMUNOGLOBULIN ALPHA AND MU POLYPEPTIDES; IGJ
Gene map locus 4q21

☐ **27:** *604814 Links
LEUKOCYTE IMMUNOGLOBULIN-LIKE RECEPTOR, SUBFAMILY B, MEMBER 5; LILRB5
Gene map locus 19q13.2

☐ **28:** *604810 Links
LEUKOCYTE IMMUNOGLOBULIN-LIKE RECEPTOR, SUBFAMILY A, MEMBER 1; LILRA1
Gene map locus 19q13.2

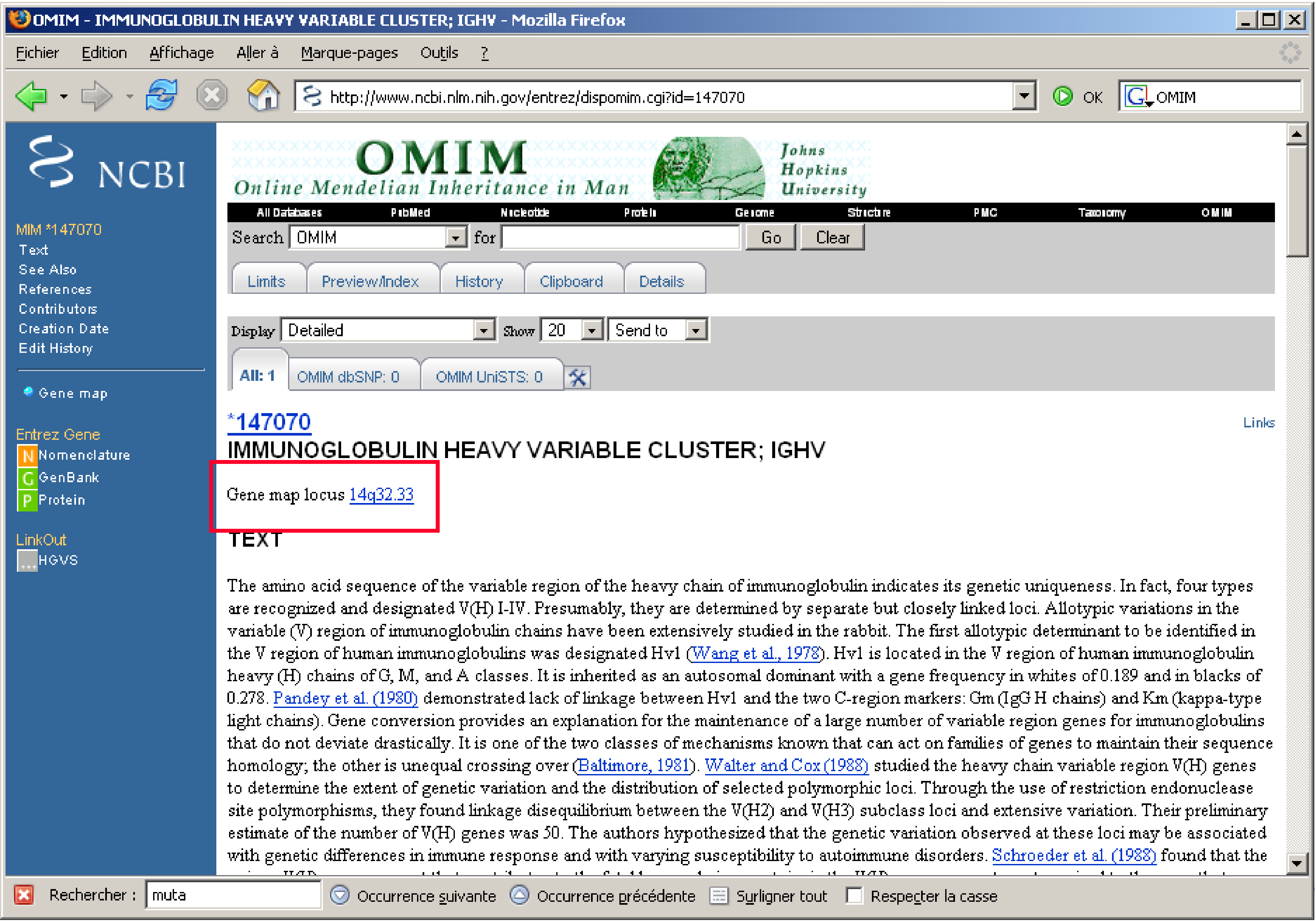
☐ **29:** *607517 Links
LEUKOCYTE IMMUNOGLOBULIN-LIKE RECEPTOR, SUBFAMILY A, MEMBER 4

☐ **30:** *147100 Links
IgG HEAVY CHAIN LOCUS; IGHG1
IGHG1/CCND1 FUSION GENE, INCLUDED
Gene map locus 14q32.33

☐ **31:** *147070 Links
IMMUNOGLOBULIN HEAVY VARIABLE CLUSTER; IGHV
Gene map locus 14q32.33

☐ **32:** #240500 GeneTests, Links
COMMON VARIABLE IMMUNODEFICIENCY; CVID
Gene map locus 17p11.2

☐ **33:** *173880 Links
POLYMERIC IMMUNOGLOBULIN RECEPTOR; PIGR
Gene map locus 1q31-q42



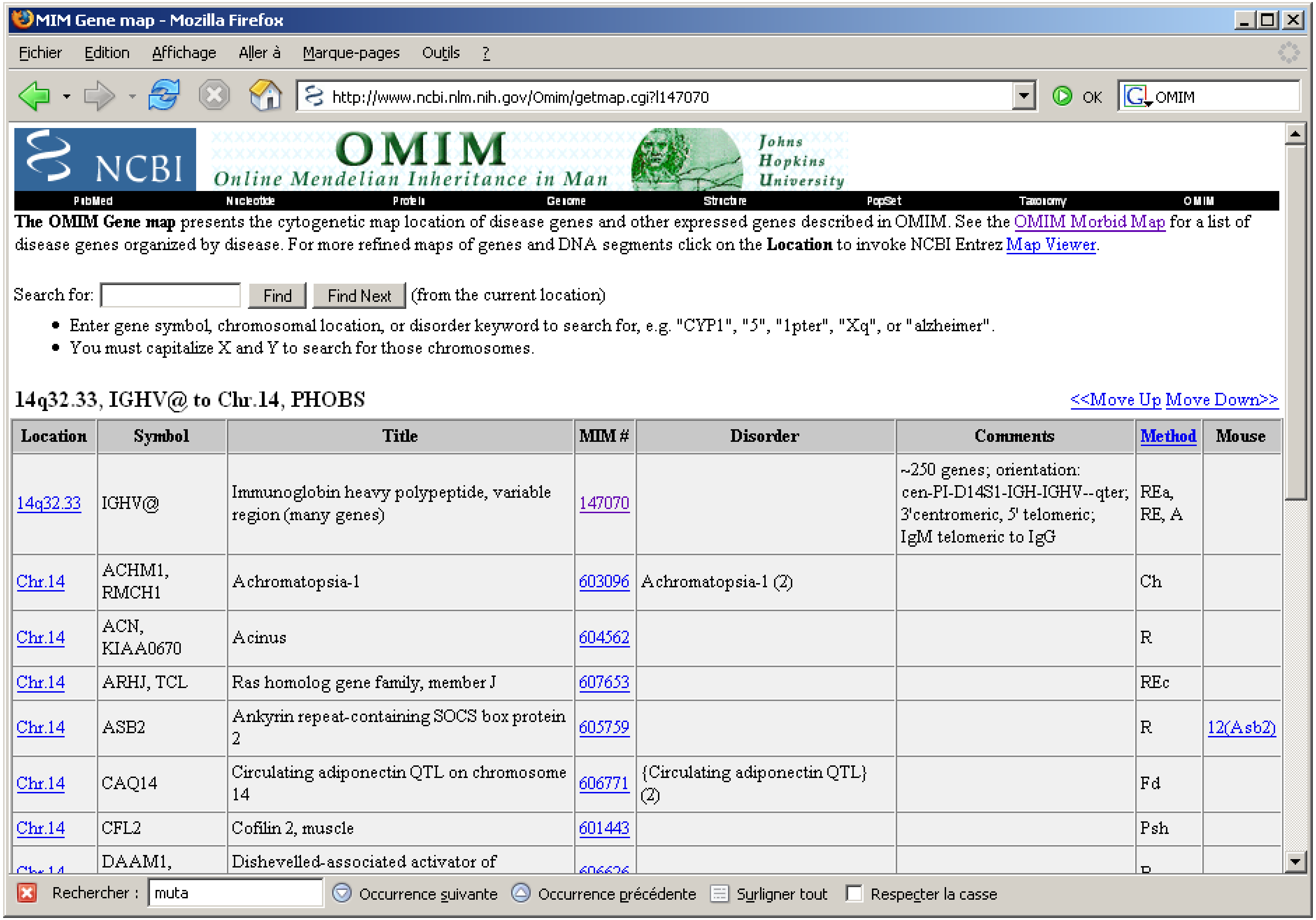
*147070

IMMUNOGLOBULIN HEAVY VARIABLE CLUSTER; IGHV

Gene map locus 14q32.33

TEXT

The amino acid sequence of the variable region of the heavy chain of immunoglobulin indicates its genetic uniqueness. In fact, four types are recognized and designated V(H) I-IV. Presumably, they are determined by separate but closely linked loci. Allotypic variations in the variable (V) region of immunoglobulin chains have been extensively studied in the rabbit. The first allotypic determinant to be identified in the V region of human immunoglobulins was designated Hv1 (Wang et al., 1978). Hv1 is located in the V region of human immunoglobulin heavy (H) chains of G, M, and A classes. It is inherited as an autosomal dominant with a gene frequency in whites of 0.189 and in blacks of 0.278. Pandey et al. (1980) demonstrated lack of linkage between Hv1 and the two C-region markers: Gm (IgG H chains) and Km (kappa-type light chains). Gene conversion provides an explanation for the maintenance of a large number of variable region genes for immunoglobulins that do not deviate drastically. It is one of the two classes of mechanisms known that can act on families of genes to maintain their sequence homology; the other is unequal crossing over (Baltimore, 1981). Walter and Cox (1988) studied the heavy chain variable region V(H) genes to determine the extent of genetic variation and the distribution of selected polymorphic loci. Through the use of restriction endonuclease site polymorphisms, they found linkage disequilibrium between the V(H2) and V(H3) subclass loci and extensive variation. Their preliminary estimate of the number of V(H) genes was 30. The authors hypothesized that the genetic variation observed at these loci may be associated with genetic differences in immune response and with varying susceptibility to autoimmune disorders. Schroeder et al. (1988) found that the



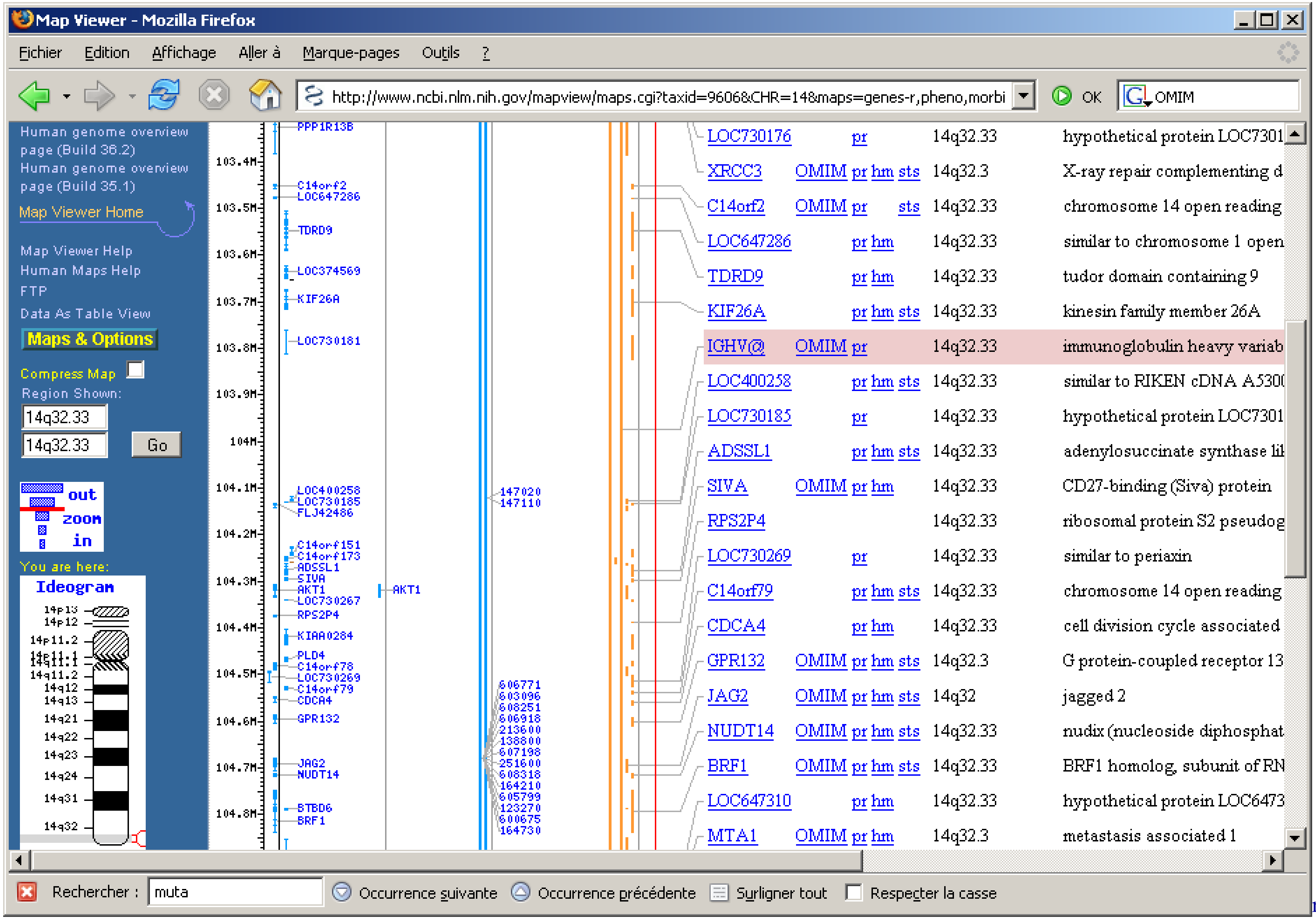
The OMIM Gene map presents the cytogenetic map location of disease genes and other expressed genes described in OMIM. See the OMIM Morbid Map for a list of disease genes organized by disease. For more refined maps of genes and DNA segments click on the **Location** to invoke NCBI Entrez Map Viewer.

Search for: ☐ Find Find Next (from the current location)

- Enter gene symbol, chromosomal location, or disorder keyword to search for, e.g. "CYP1", "5", "1pter", "Xq", or "alzheimer".
- You must capitalize X and Y to search for those chromosomes.

14q32.33, IGHV@ to Chr.14, PHOBS <<Move Up Move Down>>

Location	Symbol	Title	MIM #	Disorder	Comments	Method	Mouse
14q32.33	IGHV@	Immunoglobin heavy polypeptide, variable region (many genes)	147070		~250 genes; orientation: cen-PI-D14S1-IGH-IGHV--qter; 3'centromeric, 5' telomeric; IgM telomeric to IgG	REa, RE, A	
Chr.14	ACHM1, RMCH1	Achromatopsia-1	603096	Achromatopsia-1 (2)		Ch	
Chr.14	ACN, KIAA0670	Acinus	604562			R	
Chr.14	ARHJ, TCL	Ras homolog gene family, member J	607653			REc	
Chr.14	ASB2	Ankyrin repeat-containing SOCS box protein 2	605759			R	12(Asb2)
Chr.14	CAQ14	Circulating adiponectin QTL on chromosome 14	606771	{Circulating adiponectin QTL} (2)		Fd	
Chr.14	CFL2	Cofilin 2, muscle	601443			Psh	
Chr.14	DAAM1,	Dishevelled-associated activator of	606626			R	

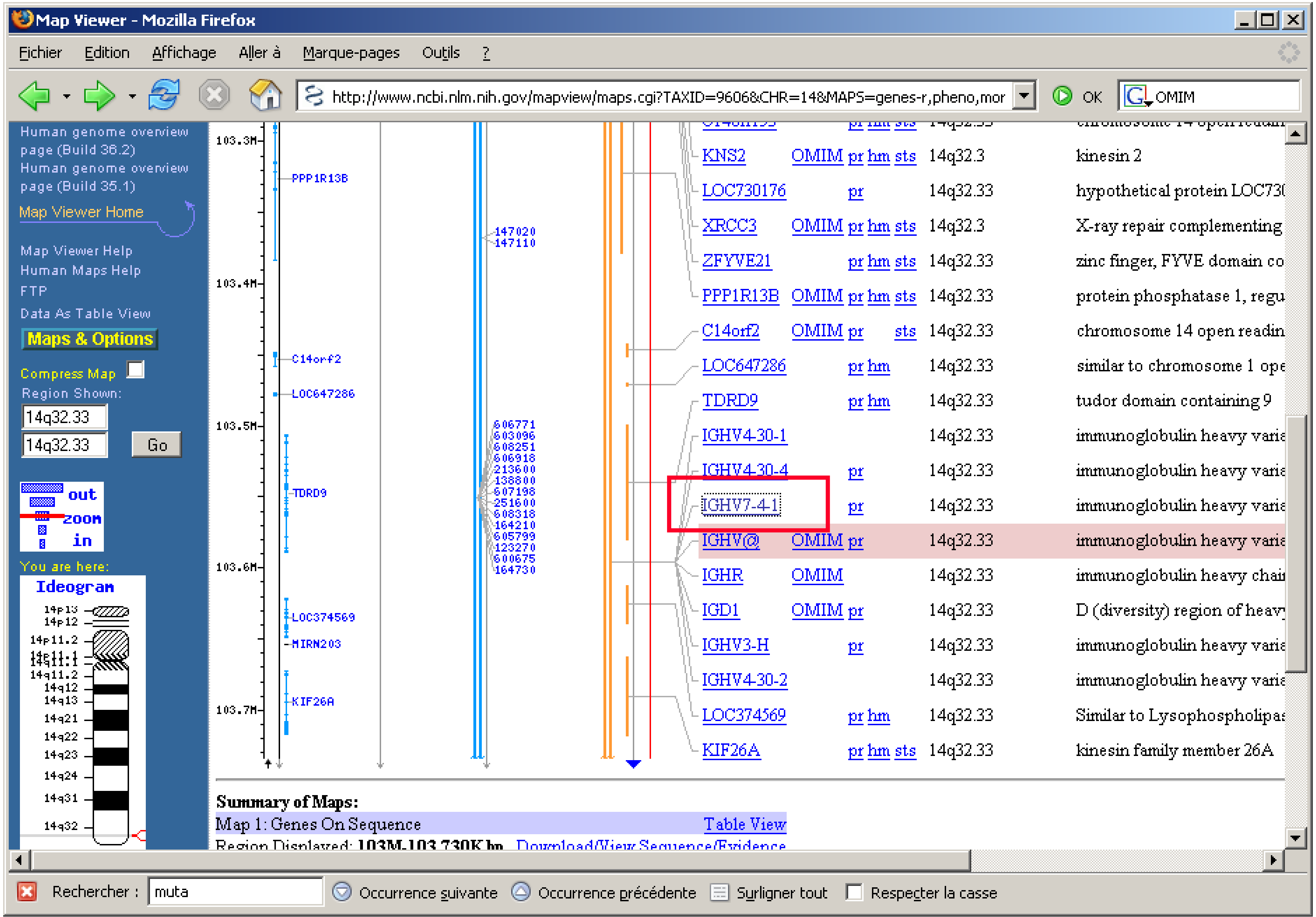


http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?taxid=9606&CHR=14&maps=genes-r,pheno,morbi

Region Shown:
14q32.33
14q32.33

LOC730176		pr	14q32.33	hypothetical protein LOC7301
XRCC3	OMIM	pr hm sts	14q32.3	X-ray repair complementing d
C14orf2	OMIM	pr sts	14q32.33	chromosome 14 open reading
LOC647286		pr hm	14q32.33	similar to chromosome 1 open
TDRD9		pr hm	14q32.33	tudor domain containing 9
KIF26A		pr hm sts	14q32.33	kinesin family member 26A
IGHV@	OMIM	pr	14q32.33	immunoglobulin heavy variab
LOC400258		pr hm sts	14q32.33	similar to RIKEN cDNA A530
LOC730185		pr	14q32.33	hypothetical protein LOC7301
ADSSL1		pr hm sts	14q32.33	adenylosuccinate synthase lil
SIVA	OMIM	pr hm	14q32.33	CD27-binding (Siva) protein
RPS2P4			14q32.33	ribosomal protein S2 pseudog
LOC730269		pr	14q32.33	similar to periaxin
C14orf79		pr hm sts	14q32.33	chromosome 14 open reading
CDCA4		pr hm	14q32.33	cell division cycle associated
GPR132	OMIM	pr hm sts	14q32.3	G protein-coupled receptor 13
JAG2	OMIM	pr hm sts	14q32	jagged 2
NUDT14	OMIM	pr hm sts	14q32.33	nudix (nucleoside diphosphat
BRF1	OMIM	pr hm sts	14q32.33	BRF1 homolog, subunit of RN
LOC647310		pr hm	14q32.33	hypothetical protein LOC6473
MTA1	OMIM	pr hm	14q32.3	metastasis associated 1

Rechercher : muta

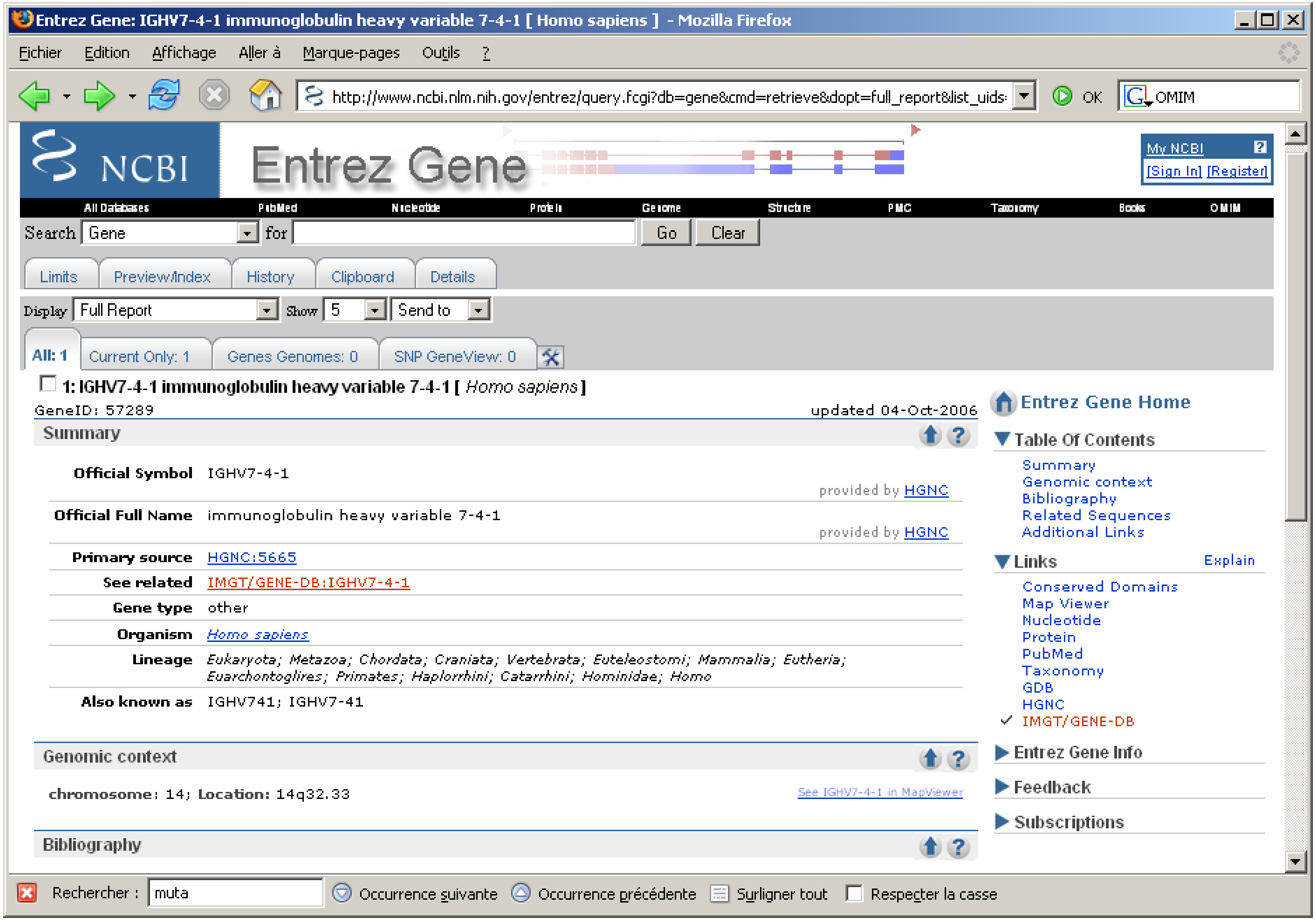


KNS2	OMIM pr hm sts	14q32.3	kinesin 2
LOC730176	pr	14q32.33	hypothetical protein LOC73
XRCC3	OMIM pr hm sts	14q32.3	X-ray repair complementing
ZFYVE21	pr hm sts	14q32.33	zinc finger, FYVE domain co
PPP1R13B	OMIM pr hm sts	14q32.33	protein phosphatase 1, regu
C14orf2	OMIM pr sts	14q32.33	chromosome 14 open readin
LOC647286	pr hm	14q32.33	similar to chromosome 1 ope
TDRD9	pr hm	14q32.33	tudor domain containing 9
IGHV4-30-1		14q32.33	immunoglobulin heavy varia
IGHV4-30-4	pr	14q32.33	immunoglobulin heavy varia
IGHV7-4-1	pr	14q32.33	immunoglobulin heavy varia
IGHV@	OMIM pr	14q32.33	immunoglobulin heavy varia
IGHR	OMIM	14q32.33	immunoglobulin heavy chai
IGD1	OMIM pr	14q32.33	D (diversity) region of heav
IGHV3-H	pr	14q32.33	immunoglobulin heavy varia
IGHV4-30-2		14q32.33	immunoglobulin heavy varia
LOC374569	pr hm	14q32.33	Similar to Lysophospholipa
KIF26A	pr hm sts	14q32.33	kinesin family member 26A

Summary of Maps:

Map 1: Genes On Sequence Table View

Region Displayed: 103M-103,730K bp Download/View Sequence/Evidence



☐ 1: IGHV7-4-1 immunoglobulin heavy variable 7-4-1 [*Homo sapiens*]

GeneID: 57289 updated 04-Oct-2006

Summary

Official Symbol	IGHV7-4-1 provided by HGNC
Official Full Name	immunoglobulin heavy variable 7-4-1 provided by HGNC
Primary source	HGNC:5665
See related	IMGT/GENE-DB:IGHV7-4-1
Gene type	other
Organism	*Homo sapiens*
Lineage	*Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo*
Also known as	IGHV741; IGHV7-41

Genomic context

chromosome: 14; **Location:** 14q32.33

See IGHV7-4-1 in MapViewer

Bibliography

Entrez Gene: IGHV7-4-1 immunoglobulin heavy variable 7-4-1 [Homo sapiens] - Mozilla Firefox

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=retrieve&dopt=full_report&lis

OMIM

chromosome: 14; **Location:** 14q32.33

See IGHV7-4-1 in MapViewer

Bibliography

Related Articles in Pubmed

PubMed links

GeneRIFs: Gene References Into Function

What's a GeneRIF?

Submit: New GeneRIF

Related Sequences

Nucleotide		Protein
Genomic	L10057.1	AAA36058.1
Genomic	Z12323.1	CAA78193.1

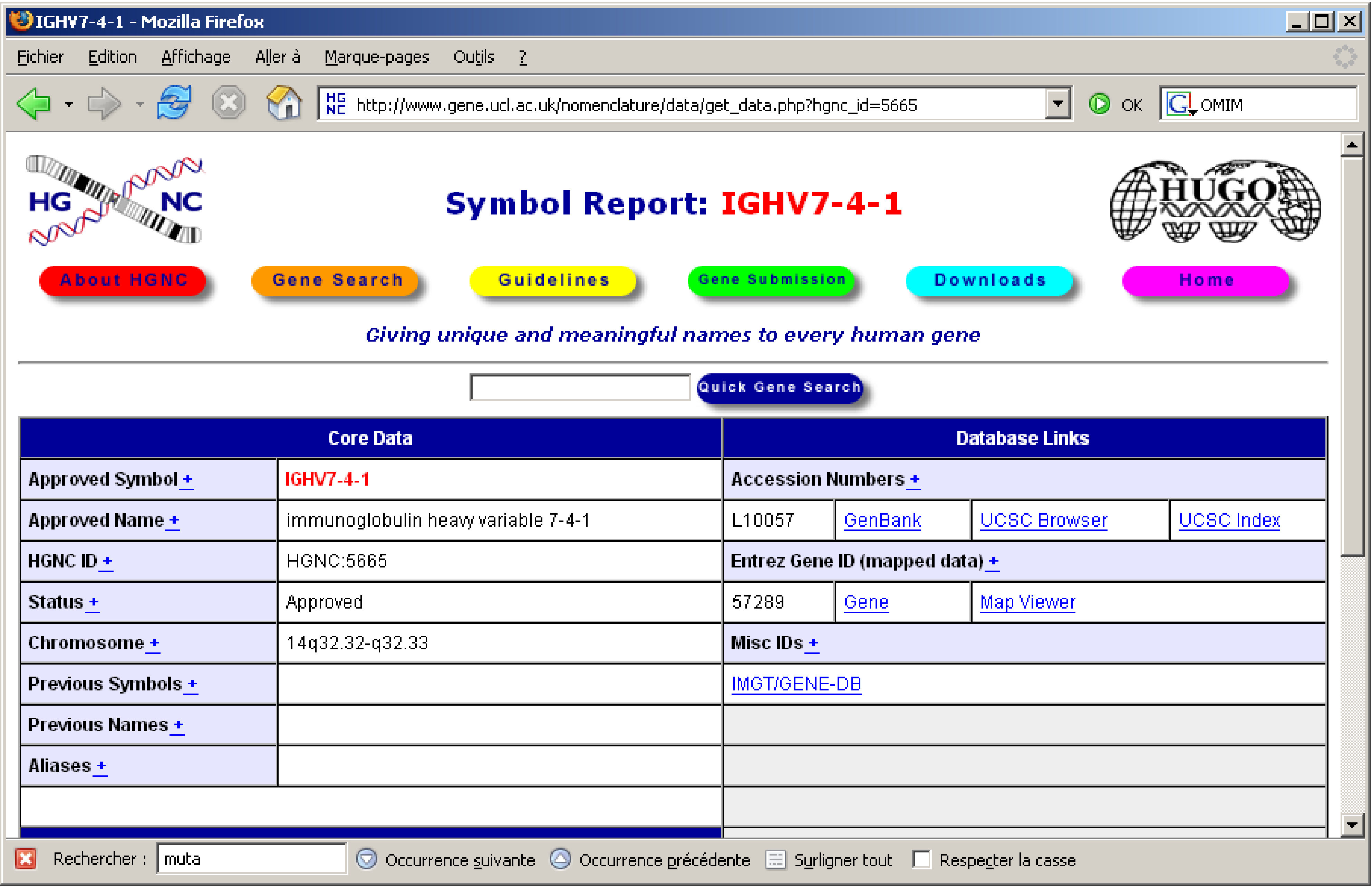
Additional Links

- IMGT Repertoire for individual human immunoglobulin and T cell receptor genes IGHV7-4-1

Display: Full Report
Show: 5
Send to

Restrictions on Use | Write to the Help Desk
NCBI | NLM | NIH

Rechercher : muta



http://www.gene.ucl.ac.uk/nomenclature/data/get_data.php?hgnc_id=5665

OMIM

Symbol Report: IGHV7-4-1

Giving unique and meaningful names to every human gene

Core Data		Database Links			
Approved Symbol +	IGHV7-4-1	**Accession Numbers +**			
Approved Name +	immunoglobulin heavy variable 7-4-1	L10057	GenBank	UCSC Browser	UCSC Index
HGNC ID +	HGNC:5665	**Entrez Gene ID (mapped data) +**			
Status +	Approved	57289	Gene	Map Viewer	
Chromosome +	14q32.32-q32.33	**Misc IDs +**			
Previous Symbols +		IMGT/GENE-DB			
Previous Names +					
Aliases +					

Rechercher : muta

☐ Respecter la casse

2006

GDB Genome DataBank

Créée en 1990 **à l'Université Johns Hopkins, Baltimore, Maryland (USA)**

Localisation maintenue de1999 à 2002 par le BiSC, Toronto
depuis novembre 2003 par le RTI (Caroline du nord)

Regroupe :
- les gènes,
- les clones,
- les marqueurs cytogénétiques, les cartographies
- RFLP, les polymorphismes
- les allèles

Contient les données relatives à la structure du génome humain, sa diversité et à son évolution

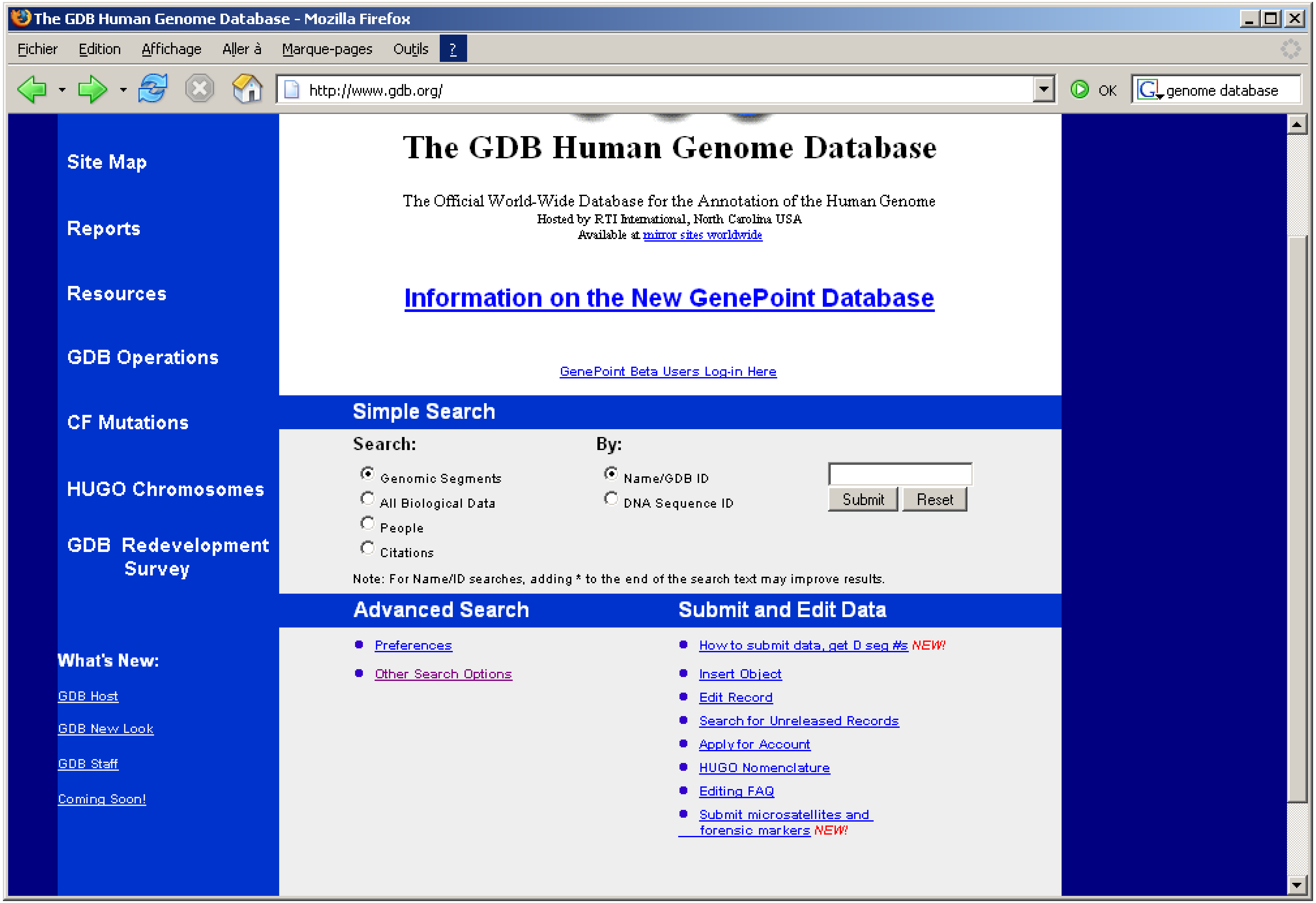
Outils de soumission des données

GenePoint : pour l'annotation du génome

Système interactif pour commenter les données

Distribution

Web: http://www.gdb.org/



Site Map

Reports

Resources

GDB Operations

CF Mutations

HUGO Chromosomes

GDB Redevelopment Survey

What's New:

GDB Host

GDB New Look

GDB Staff

Coming Soon!

The GDB Human Genome Database

The Official World-Wide Database for the Annotation of the Human Genome
Hosted by RTI International, North Carolina USA
Available at mirror sites worldwide

Information on the New GenePoint Database

GenePoint Beta Users Log-in Here

Simple Search

Search:

● Genomic Segments
○ All Biological Data
○ People
○ Citations

By:

● Name/GDB ID
○ DNA Sequence ID

Submit Reset

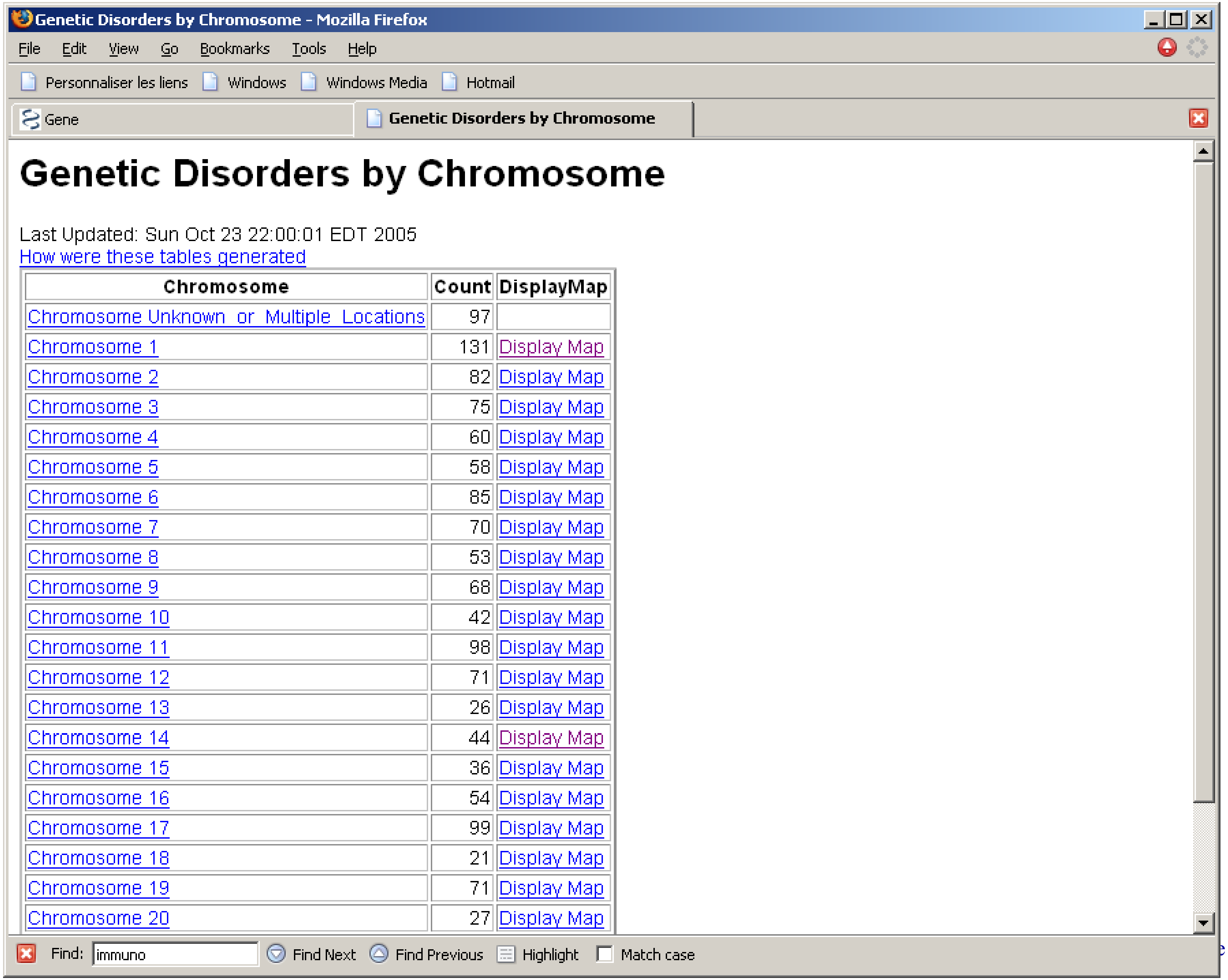
Note: For Name/ID searches, adding * to the end of the search text may improve results.

Advanced Search

- Preferences
- Other Search Options

Submit and Edit Data

- How to submit data, get D seg #s *NEW!*
- Insert Object
- Edit Record
- Search for Unreleased Records
- Apply for Account
- HUGO Nomenclature
- Editing FAQ
- Submit microsatellites and forensic markers *NEW!*

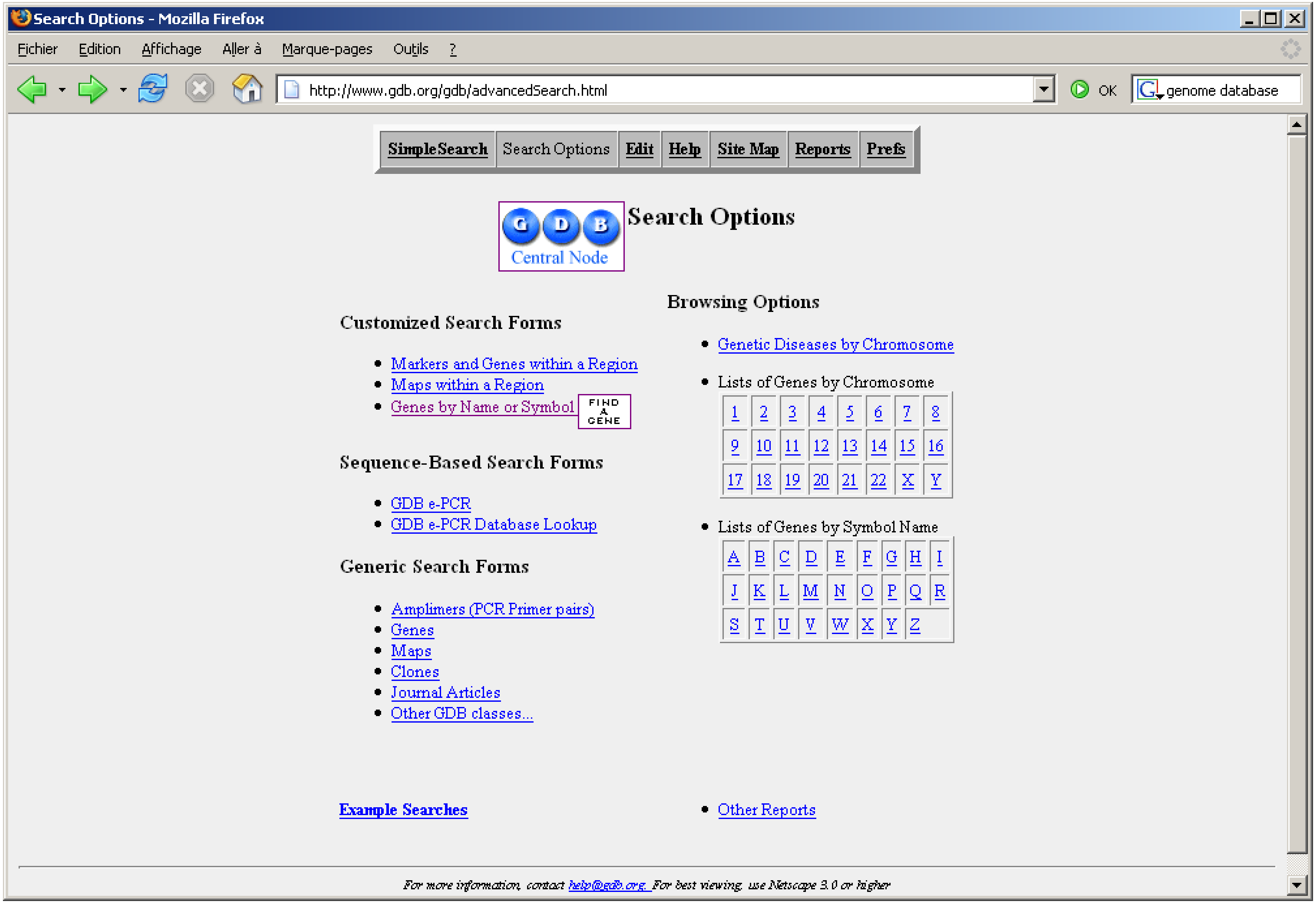


Find: immuno

Genetic Disorders by Chromosome

Last Updated: Sun Oct 23 22:00:01 EDT 2005
How were these tables generated

Chromosome	Count	DisplayMap
Chromosome Unknown or Multiple Locations	97	
Chromosome 1	131	Display Map
Chromosome 2	82	Display Map
Chromosome 3	75	Display Map
Chromosome 4	60	Display Map
Chromosome 5	58	Display Map
Chromosome 6	85	Display Map
Chromosome 7	70	Display Map
Chromosome 8	53	Display Map
Chromosome 9	68	Display Map
Chromosome 10	42	Display Map
Chromosome 11	98	Display Map
Chromosome 12	71	Display Map
Chromosome 13	26	Display Map
Chromosome 14	44	Display Map
Chromosome 15	36	Display Map
Chromosome 16	54	Display Map
Chromosome 17	99	Display Map
Chromosome 18	21	Display Map
Chromosome 19	71	Display Map
Chromosome 20	27	Display Map



Search Options

Customized Search Forms

- Markers and Genes within a Region
- Maps within a Region
- Genes by Name or Symbol

Sequence-Based Search Forms

- GDB e-PCR
- GDB e-PCR Database Lookup

Generic Search Forms

- Amplimers (PCR Primer pairs)
- Genes
- Maps
- Clones
- Journal Articles
- Other GDB classes...

Example Searches

Browsing Options

- Genetic Diseases by Chromosome
- Lists of Genes by Chromosome

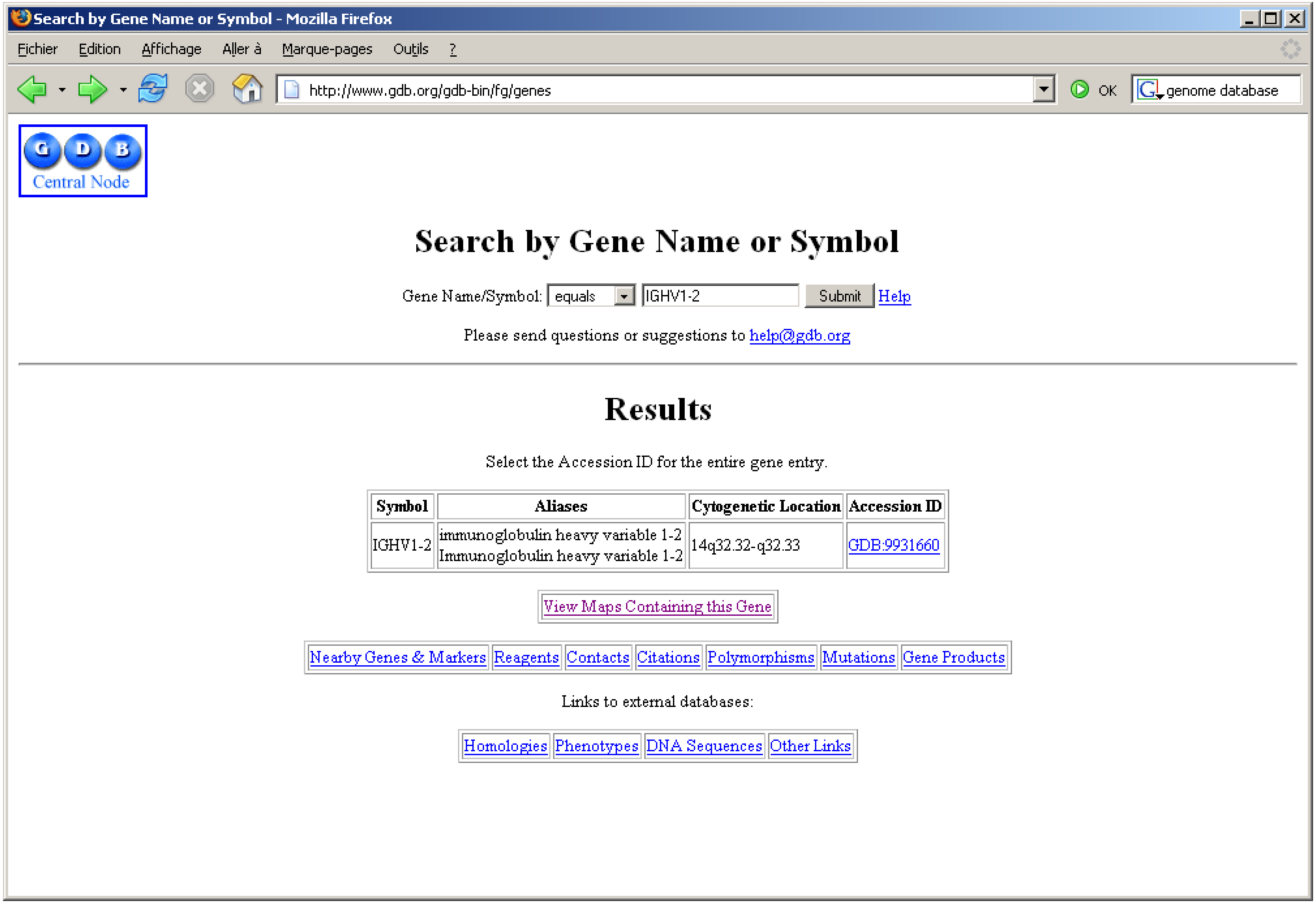
1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	X	Y

- Lists of Genes by Symbol Name

A	B	C	D	E	F	G	H	I
J	K	L	M	N	O	P	Q	R
S	T	U	V	W	X	Y	Z	

- Other Reports

For more information, contact help@gdb.org. For best viewing, use Netscape 3.0 or higher



Search by Gene Name or Symbol

Gene Name/Symbol: equals IGHV1-2 Submit Help

Please send questions or suggestions to help@gdb.org

Results

Select the Accession ID for the entire gene entry.

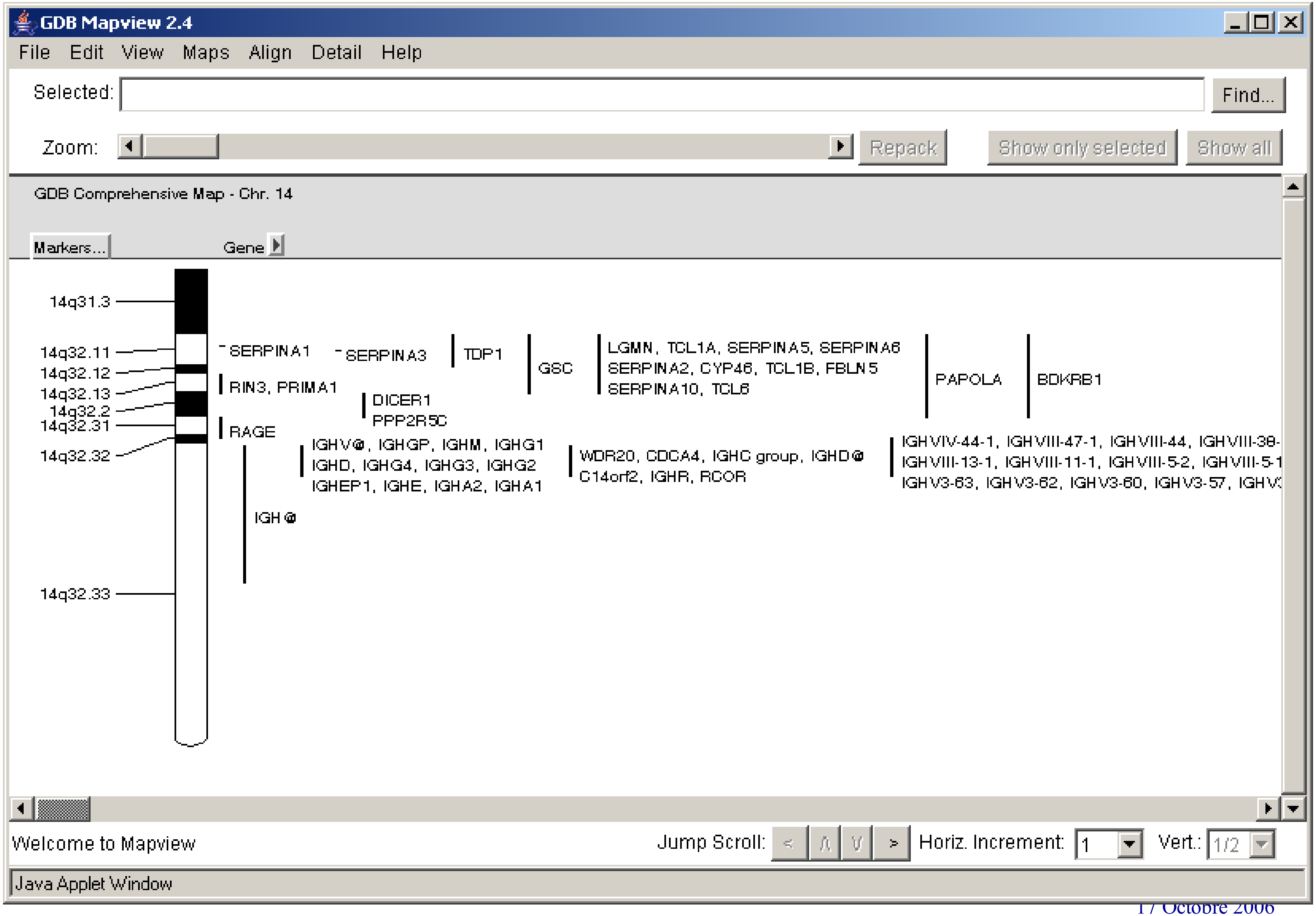
Symbol	Aliases	Cytogenetic Location	Accession ID
IGHV1-2	immunoglobulin heavy variable 1-2 Immunoglobulin heavy variable 1-2	14q32.32-q32.33	GDB:9931660

View Maps Containing this Gene

Nearby Genes & Markers | Reagents | Contacts | Citations | Polymorphisms | Mutations | Gene Products

Links to external databases:

Homologies | Phenotypes | DNA Sequences | Other Links



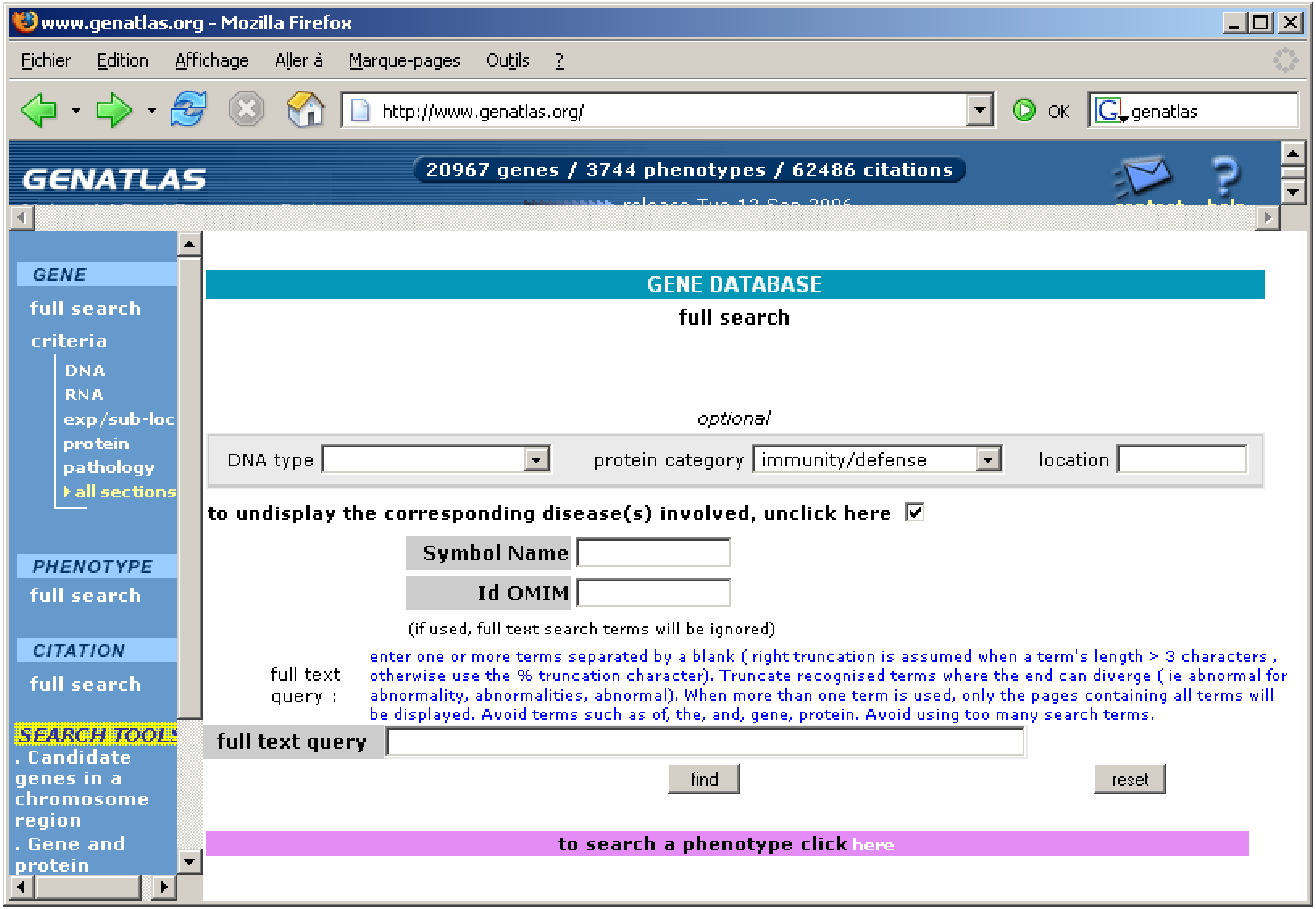
GDB Comprehensive Map - Chr. 14

14q31.3
14q32.11
14q32.12
14q32.13
14q32.2
14q32.31
14q32.32
14q32.33

SERPINA1
SERPINA3
TDP1
GSC
LGMN, TCL1A, SERPINA5, SERPINA6
SERPINA2, CYP46, TCL1B, FBLN5
SERPINA10, TCL6
PAPOLA
BDKRB1
RIN3, PRIMA1
DICER1
PPP2R5C
RAGE
IGHV@, IGHGP, IGHM, IGHG1
IGHD, IGHG4, IGHG3, IGHG2
IGHEP1, IGHE, IGHA2, IGHA1
WDR20, CDCA4, IGHC group, IGHD@
C14orf2, IGHR, RCOR
IGHVIV-44-1, IGHVIII-47-1, IGHVIII-44, IGHVIII-38-
IGHVIII-13-1, IGHVIII-11-1, IGHVIII-5-2, IGHVIII-5-1
IGHV3-63, IGHV3-62, IGHV3-60, IGHV3-57, IGHV
IGH@

Welcome to Mapview

Horiz. Increment: 1
Vert.: 1/2



http://www.genatlas.org/

20967 genes / 3744 phenotypes / 62486 citations

. Candidate genes in a chromosome region

. Gene and protein

GENE DATABASE

full search

optional

DNA type: (blank)

protein category: immunity/defense

location: (blank)

to undisplay the corresponding disease(s) involved, unclick here ☑

Symbol Name (blank)

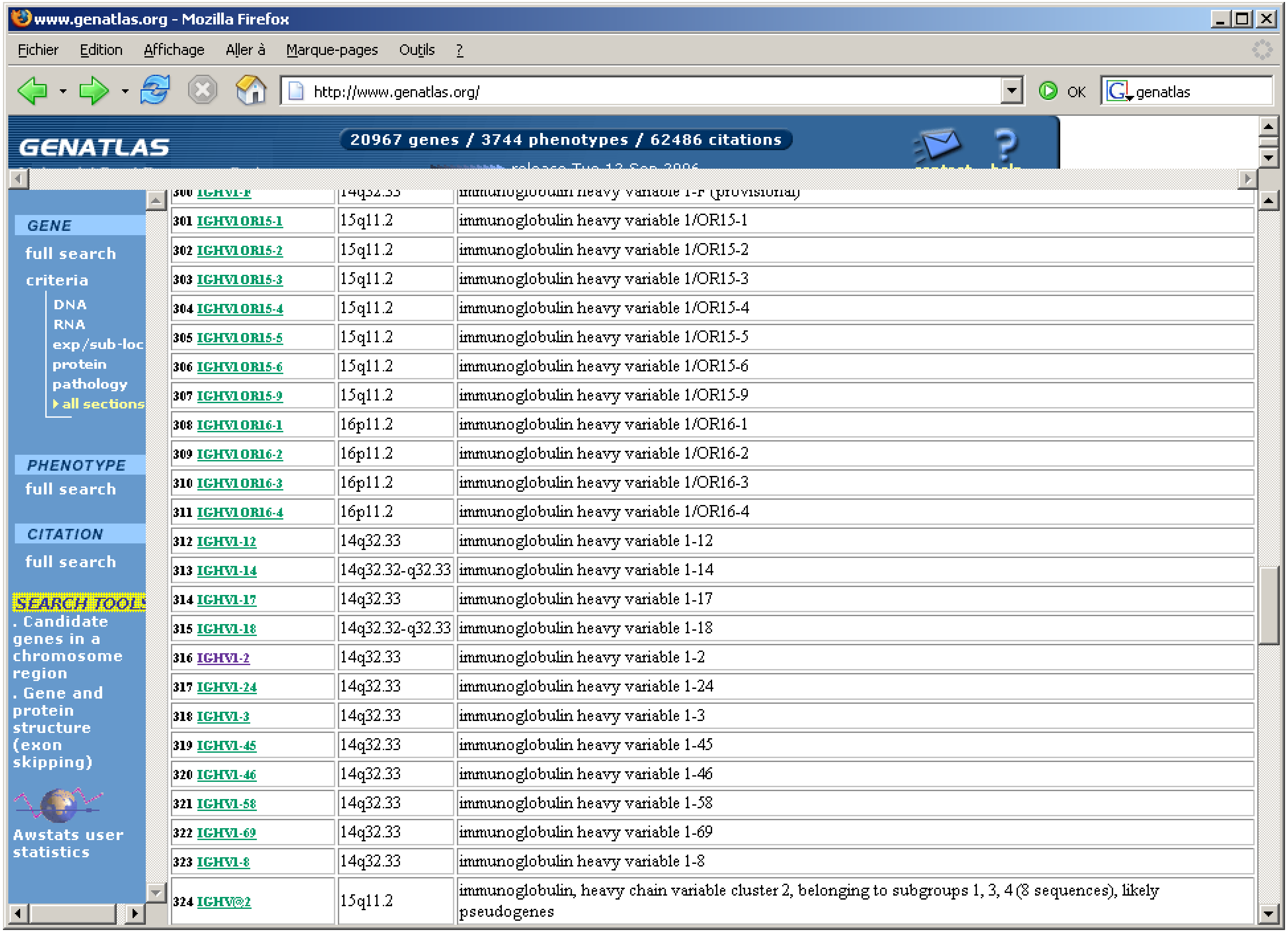
Id OMIM (blank)

(if used, full text search terms will be ignored)

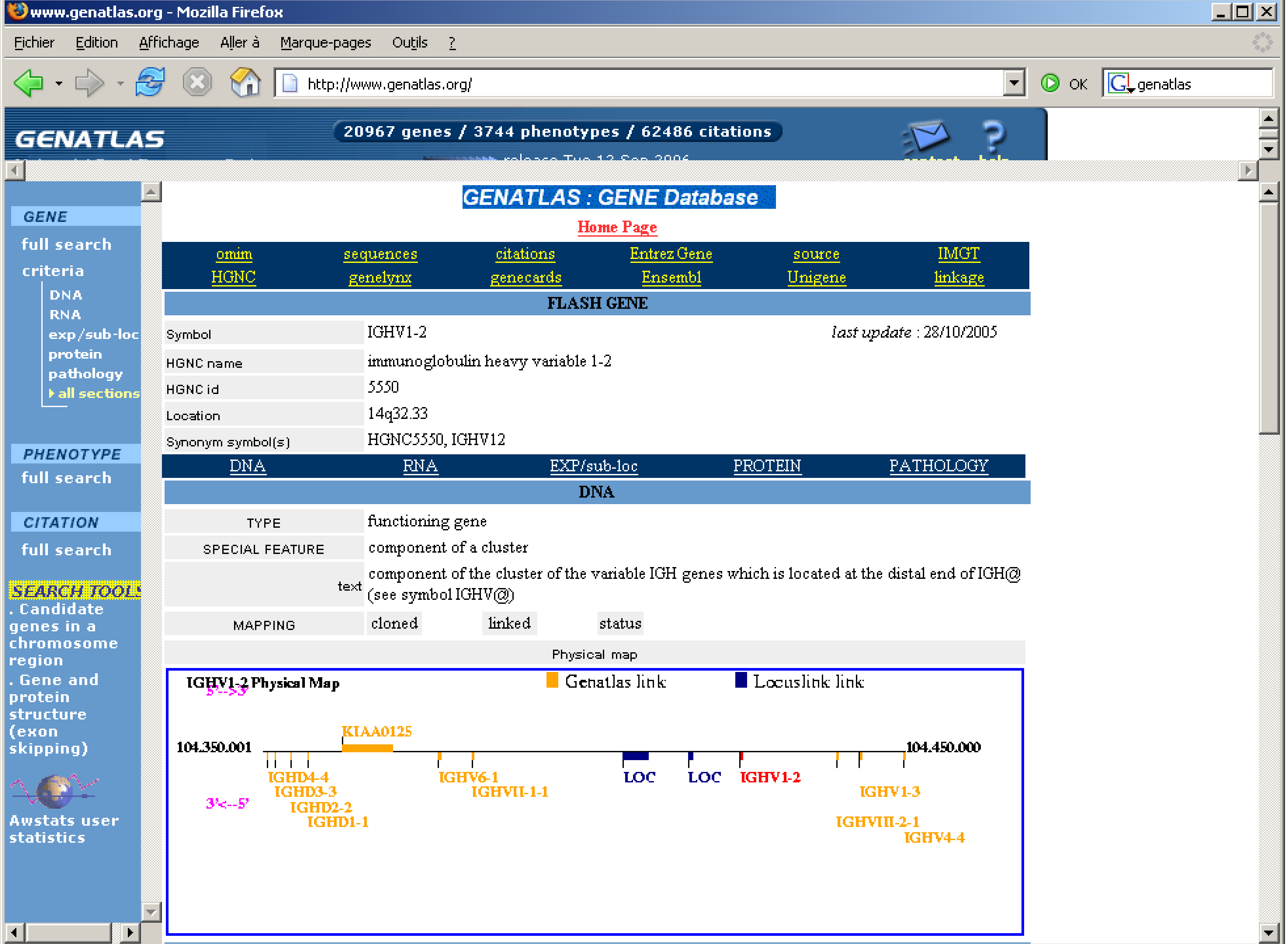
full text query : enter one or more terms separated by a blank (right truncation is assumed when a term's length > 3 characters , otherwise use the % truncation character). Truncate recognised terms where the end can diverge (ie abnormal for abnormality, abnormalities, abnormal). When more than one term is used, only the pages containing all terms will be displayed. Avoid terms such as of, the, and, gene, protein. Avoid using too many search terms.

full text query (blank)

to search a phenotype click here



300 IGHV1-F	14q32.33	immunoglobulin heavy variable 1-F (provisional)
301 IGHV1OR15-1	15q11.2	immunoglobulin heavy variable 1/OR15-1
302 IGHV1OR15-2	15q11.2	immunoglobulin heavy variable 1/OR15-2
303 IGHV1OR15-3	15q11.2	immunoglobulin heavy variable 1/OR15-3
304 IGHV1OR15-4	15q11.2	immunoglobulin heavy variable 1/OR15-4
305 IGHV1OR15-5	15q11.2	immunoglobulin heavy variable 1/OR15-5
306 IGHV1OR15-6	15q11.2	immunoglobulin heavy variable 1/OR15-6
307 IGHV1OR15-9	15q11.2	immunoglobulin heavy variable 1/OR15-9
308 IGHV1OR16-1	16p11.2	immunoglobulin heavy variable 1/OR16-1
309 IGHV1OR16-2	16p11.2	immunoglobulin heavy variable 1/OR16-2
310 IGHV1OR16-3	16p11.2	immunoglobulin heavy variable 1/OR16-3
311 IGHV1OR16-4	16p11.2	immunoglobulin heavy variable 1/OR16-4
312 IGHV1-12	14q32.33	immunoglobulin heavy variable 1-12
313 IGHV1-14	14q32.32-q32.33	immunoglobulin heavy variable 1-14
314 IGHV1-17	14q32.33	immunoglobulin heavy variable 1-17
315 IGHV1-18	14q32.32-q32.33	immunoglobulin heavy variable 1-18
316 IGHV1-2	14q32.33	immunoglobulin heavy variable 1-2
317 IGHV1-24	14q32.33	immunoglobulin heavy variable 1-24
318 IGHV1-3	14q32.33	immunoglobulin heavy variable 1-3
319 IGHV1-45	14q32.33	immunoglobulin heavy variable 1-45
320 IGHV1-46	14q32.33	immunoglobulin heavy variable 1-46
321 IGHV1-58	14q32.33	immunoglobulin heavy variable 1-58
322 IGHV1-69	14q32.33	immunoglobulin heavy variable 1-69
323 IGHV1-8	14q32.33	immunoglobulin heavy variable 1-8
324 IGHV@2	15q11.2	immunoglobulin, heavy chain variable cluster 2, belonging to subgroups 1, 3, 4 (8 sequences), likely pseudogenes



GENATLAS : GENE Database

Home Page

omim	sequences	citations	Entrez Gene	source	IMGT
HGNC	genelynx	genecards	Ensembl	Unigene	linkage

FLASH GENE

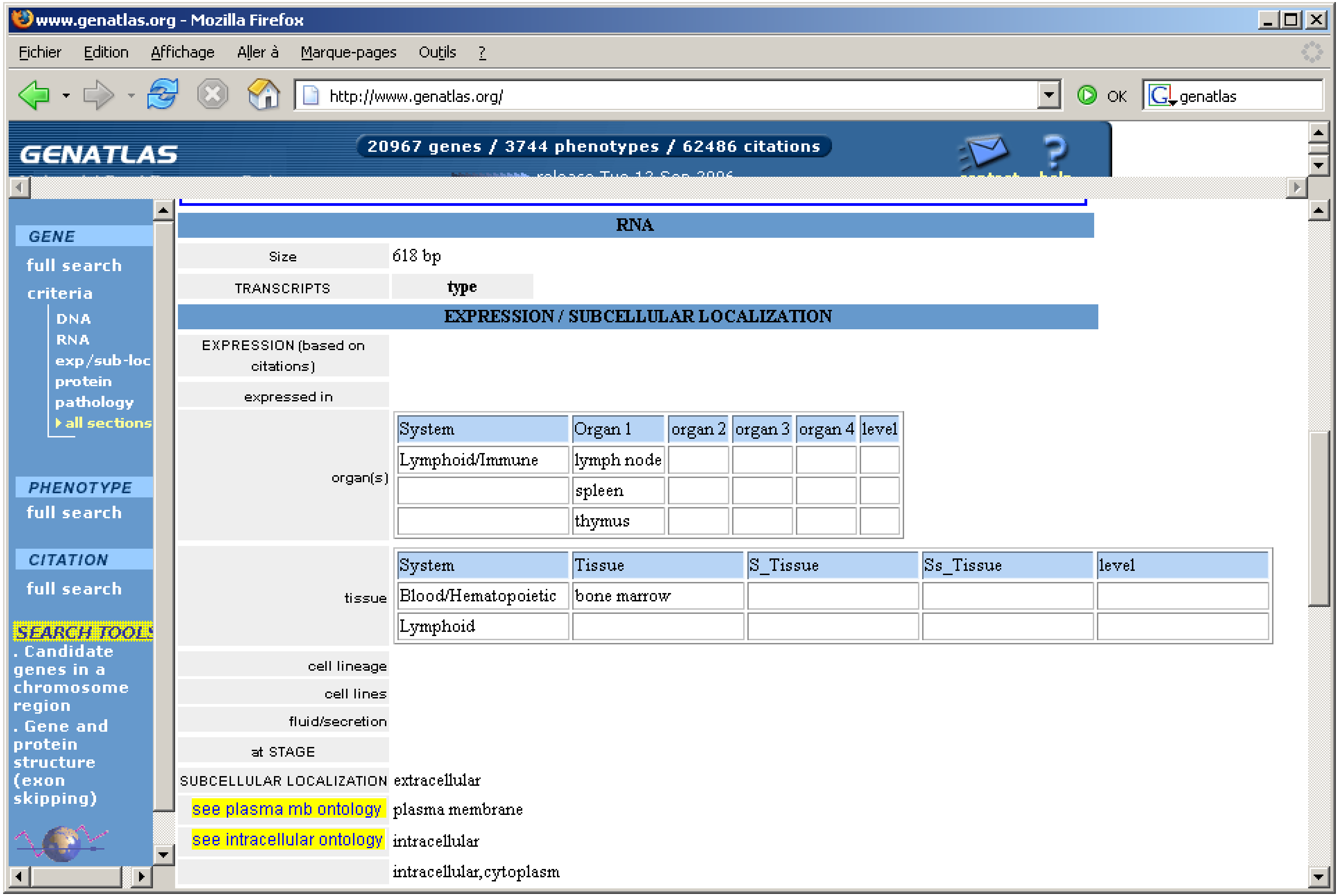
Symbol	IGHV1-2	*last update* : 28/10/2005
HGNC name	immunoglobulin heavy variable 1-2	
HGNC id	5550	
Location	14q32.33	
Synonym symbol(s)	HGNC5550, IGHV12	

DNA RNA EXP/sub-loc PROTEIN PATHOLOGY

DNA

TYPE	functioning gene
SPECIAL FEATURE	component of a cluster
text	component of the cluster of the variable IGH genes which is located at the distal end of IGH@ (see symbol IGHV@)
MAPPING	cloned linked status

Physical map



RNA

Size	618 bp
TRANSCRIPTS	**type**

EXPRESSION / SUBCELLULAR LOCALIZATION

EXPRESSION (based on citations)

expressed in

organ(s)

System	Organ 1	organ 2	organ 3	organ 4	level
Lymphoid/Immune	lymph node				
	spleen				
	thymus				

tissue

System	Tissue	S_Tissue	Ss_Tissue	level
Blood/Hematopoietic	bone marrow			
Lymphoid				

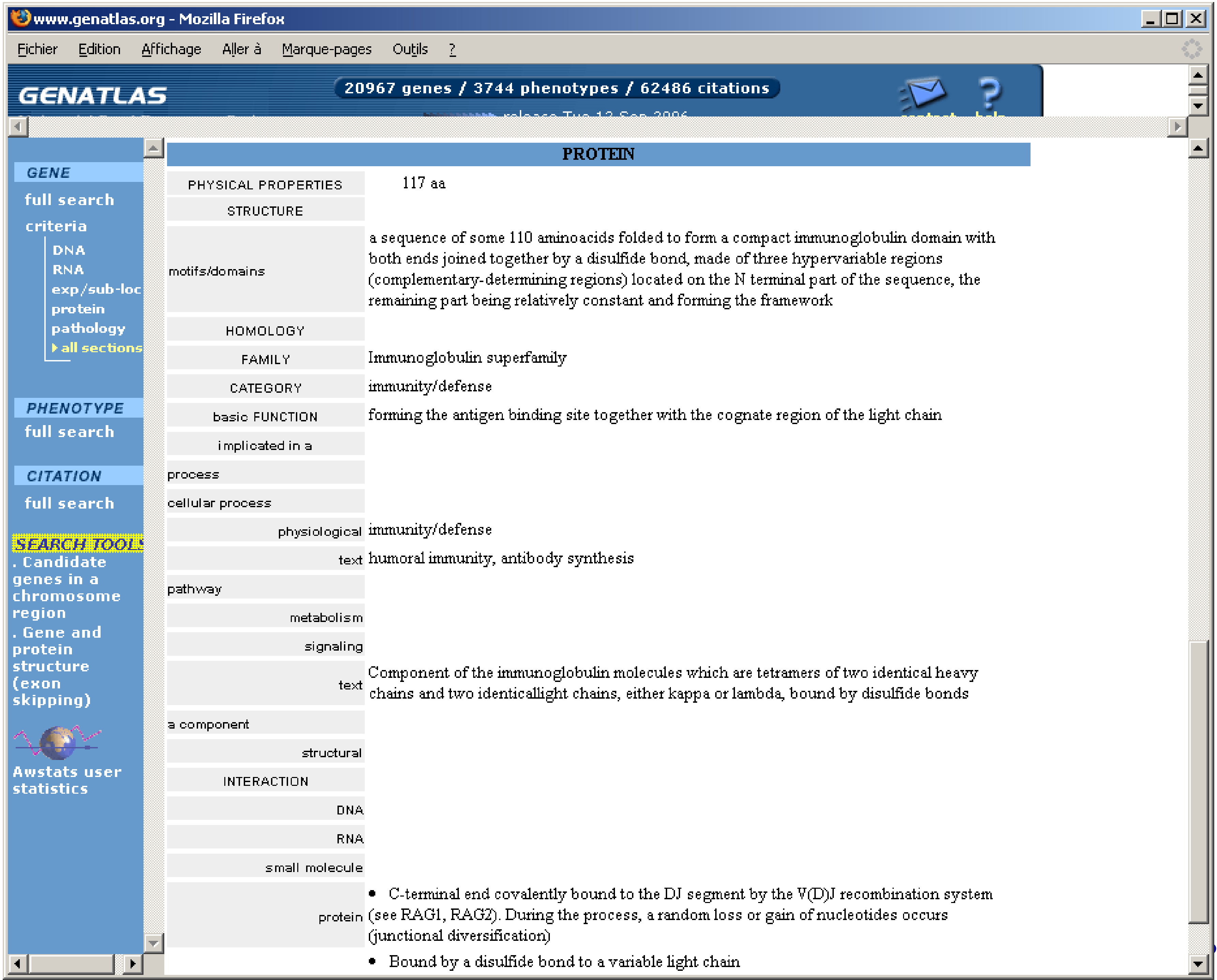
cell lineage

cell lines

fluid/secretion

at STAGE

SUBCELLULAR LOCALIZATION	extracellular
see plasma mb ontology	plasma membrane
see intracellular ontology	intracellular
	intracellular,cytoplasm



PROTEIN

PHYSICAL PROPERTIES	117 aa
STRUCTURE	
motifs/domains	a sequence of some 110 aminoacids folded to form a compact immunoglobulin domain with both ends joined together by a disulfide bond, made of three hypervariable regions (complementary-determining regions) located on the N terminal part of the sequence, the remaining part being relatively constant and forming the framework
HOMOLOGY	
FAMILY	Immunoglobulin superfamily
CATEGORY	immunity/defense
basic FUNCTION	forming the antigen binding site together with the cognate region of the light chain
implicated in a	
process	
cellular process	
physiological	immunity/defense
text	humoral immunity, antibody synthesis
pathway	
metabolism	
signaling	
text	Component of the immunoglobulin molecules which are tetramers of two identical heavy chains and two identicallight chains, either kappa or lambda, bound by disulfide bonds
a component	
structural	
INTERACTION	
DNA	
RNA	
small molecule	
protein	• C-terminal end covalently bound to the DJ segment by the V(D)J recombination system (see RAG1, RAG2). During the process, a random loss or gain of nucleotides occurs (junctional diversification) • Bound by a disulfide bond to a variable light chain

Les bases de données généralistes
de séquences nucléotidiques et protéiques

Elles couvrent tous les secteurs de la biologie, toutes les espèces

1 - Avantages

- Ce sont les bases de données les plus complètes
- Elles sont très largement distribuées
- Elles permettent aux auteurs de prendre date
- Ce sont des outils indispensables à la diffusion rapide des résultats scientifiques

2 - Inconvénients

- Elles contiennent beaucoup de redondance
- La qualité des données est sous la responsabilité des auteurs
- Encore beaucoup d'erreurs malgré les contrôles
- Evolution lente

Les bases de données généralistes de séquences nucléotidiques

GenBank	**NCBI, USA**
http://www.ncbi.nlm.nih.gov/Genbank/index.html	
EMBL	**EBI, UK**
http://www.ebi.ac.uk/embl/	
DDBJ	**NIG, Japon**
http://www.ddbj.nig.ac.jp	

EMBL/GenBank/DDJB

International Nucleotide Sequence Database Collaboration

http://www.insdc.org/

- Ces 3 BD contiennent les **mêmes informations** à quelques jours prêts (différences dans la format et la syntaxe)
- Elles contiennent toutes les séquences nucléotidiques de:
  - projets de séquençage de génomes
  - Projets individuel de recherche
  - Patent offices
- Les données non confidentielles sont échangées tous les jours

EMBL/GenBank/DDJB

- Grande hétérogénéité dans la longueur des séquences
- Tailles: il n'y a plus de maximum (>300.000pb)

  Les plus petites de l'ordre de 10 nt

- **Toutes les séquences sont conservées**

  (beaucoup de redondance)

- Beaucoup d'erreurs dans les séquences, les annotations, la détermination des CDS

- Hétérogénité de la qualité des annotations qui sont principalement fournies par les auteurs

GenBank

Banque de séquences d’ADN et d’ARN

Créée en 1978

Localisation

National Center for Biotechnology Information,
NLM, NIH, Bethesda, MD, USA

Août 2006
(version 155)

> 61.10^6 entrées (+ de 65.10^9 de nucléotides)
Plus de 205.000 espèces sont représentées
(+ de 3000 nouvelles espèces par mois)

Origine des séquences

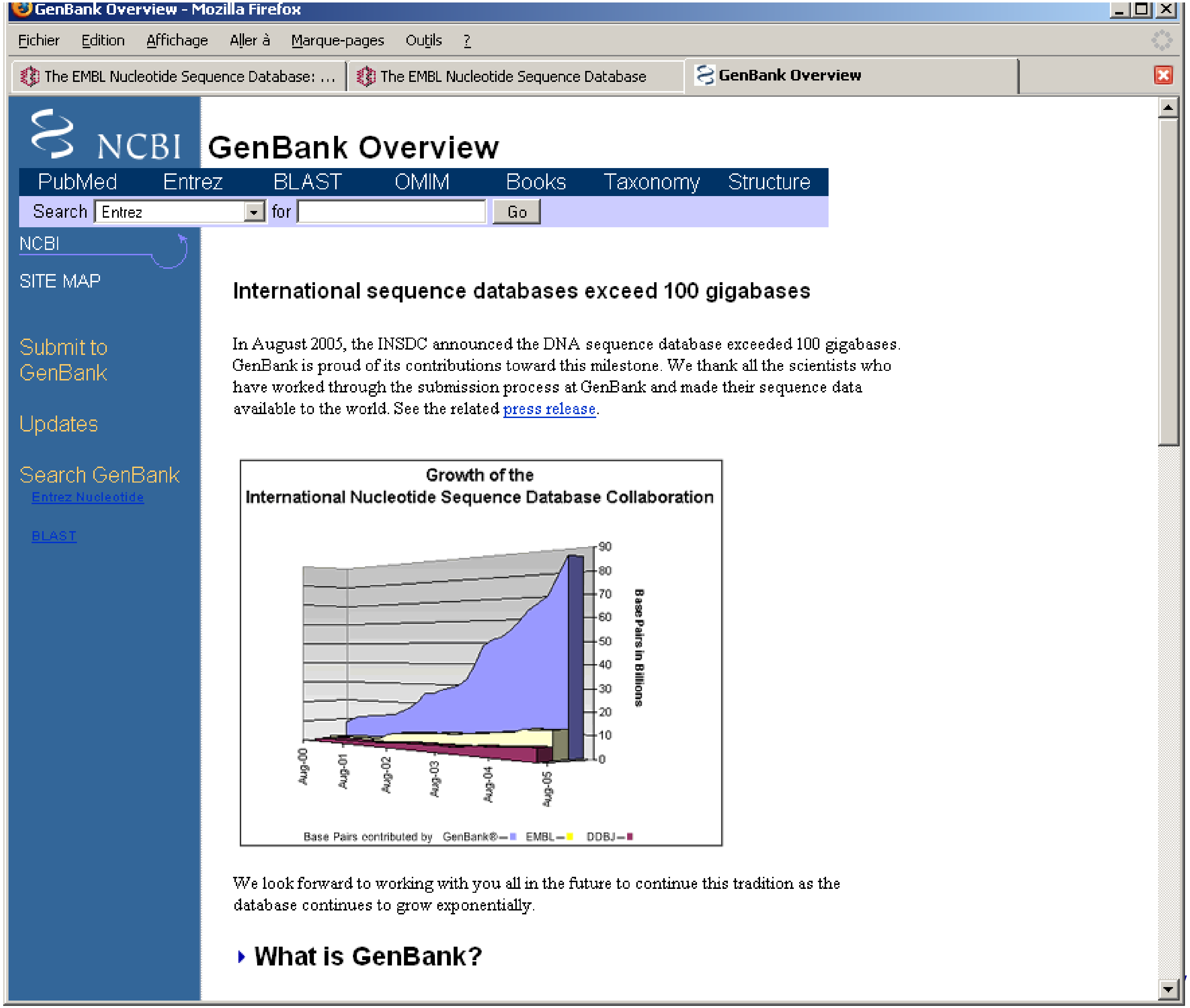
Analyse de la littérature mondiale
Soumission par les auteurs
Collaboration avec EMBL et DDBJ

Outils de soumission des séquences

Sequin, Bankit, tbl2asn

Consultation, distribution (fichiers à plat)

Web (http://www.ncbi.nlm.nih.gov), Entrez, FTP



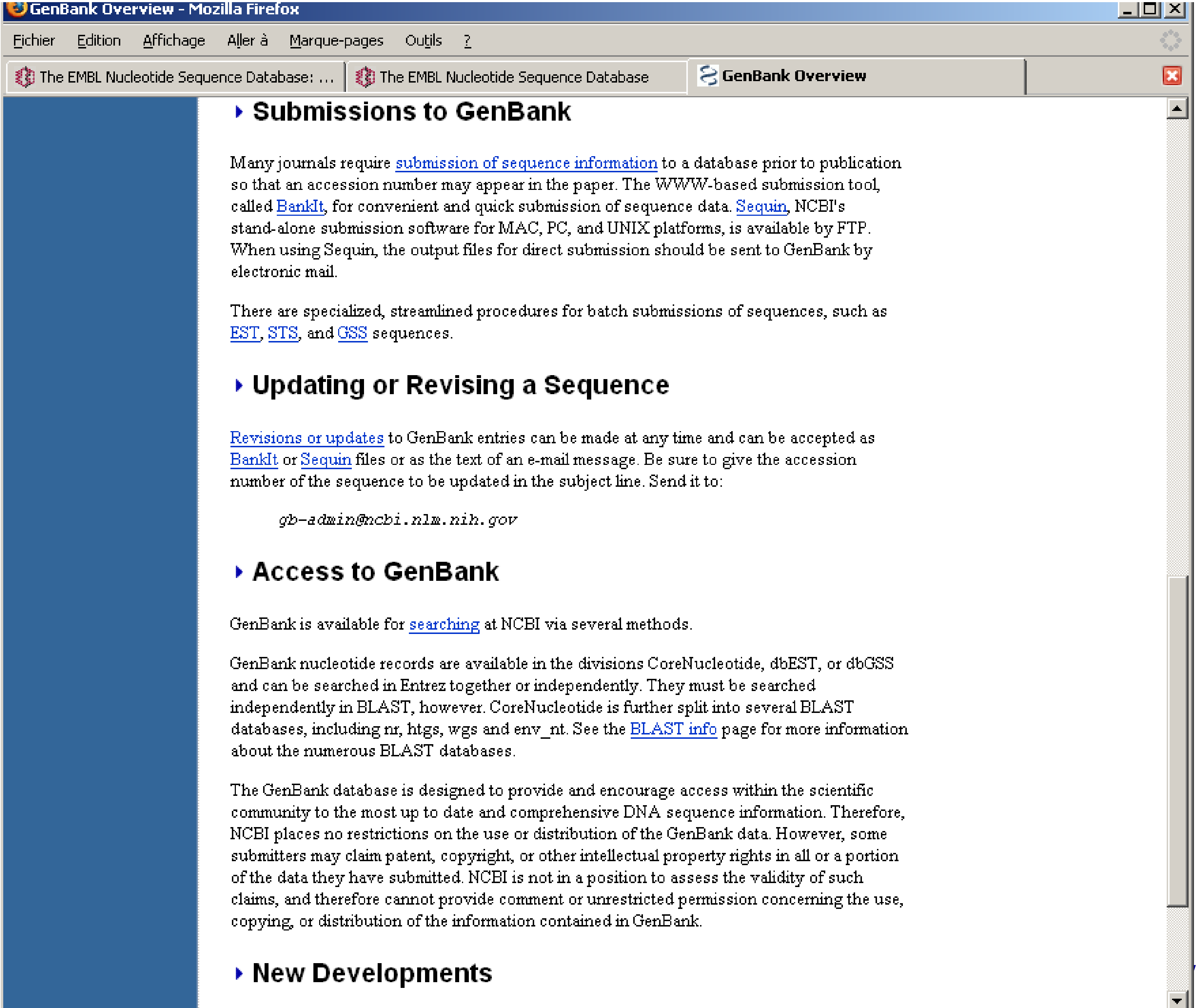
GenBank Overview

International sequence databases exceed 100 gigabases

In August 2005, the INSDC announced the DNA sequence database exceeded 100 gigabases. GenBank is proud of its contributions toward this milestone. We thank all the scientists who have worked through the submission process at GenBank and made their sequence data available to the world. See the related press release.

We look forward to working with you all in the future to continue this tradition as the database continues to grow exponentially.

What is GenBank?



Submissions to GenBank

Many journals require submission of sequence information to a database prior to publication so that an accession number may appear in the paper. The WWW-based submission tool, called BankIt, for convenient and quick submission of sequence data. Sequin, NCBI's stand-alone submission software for MAC, PC, and UNIX platforms, is available by FTP. When using Sequin, the output files for direct submission should be sent to GenBank by electronic mail.

There are specialized, streamlined procedures for batch submissions of sequences, such as EST, STS, and GSS sequences.

Updating or Revising a Sequence

Revisions or updates to GenBank entries can be made at any time and can be accepted as BankIt or Sequin files or as the text of an e-mail message. Be sure to give the accession number of the sequence to be updated in the subject line. Send it to:

gb-admin@ncbi.nlm.nih.gov

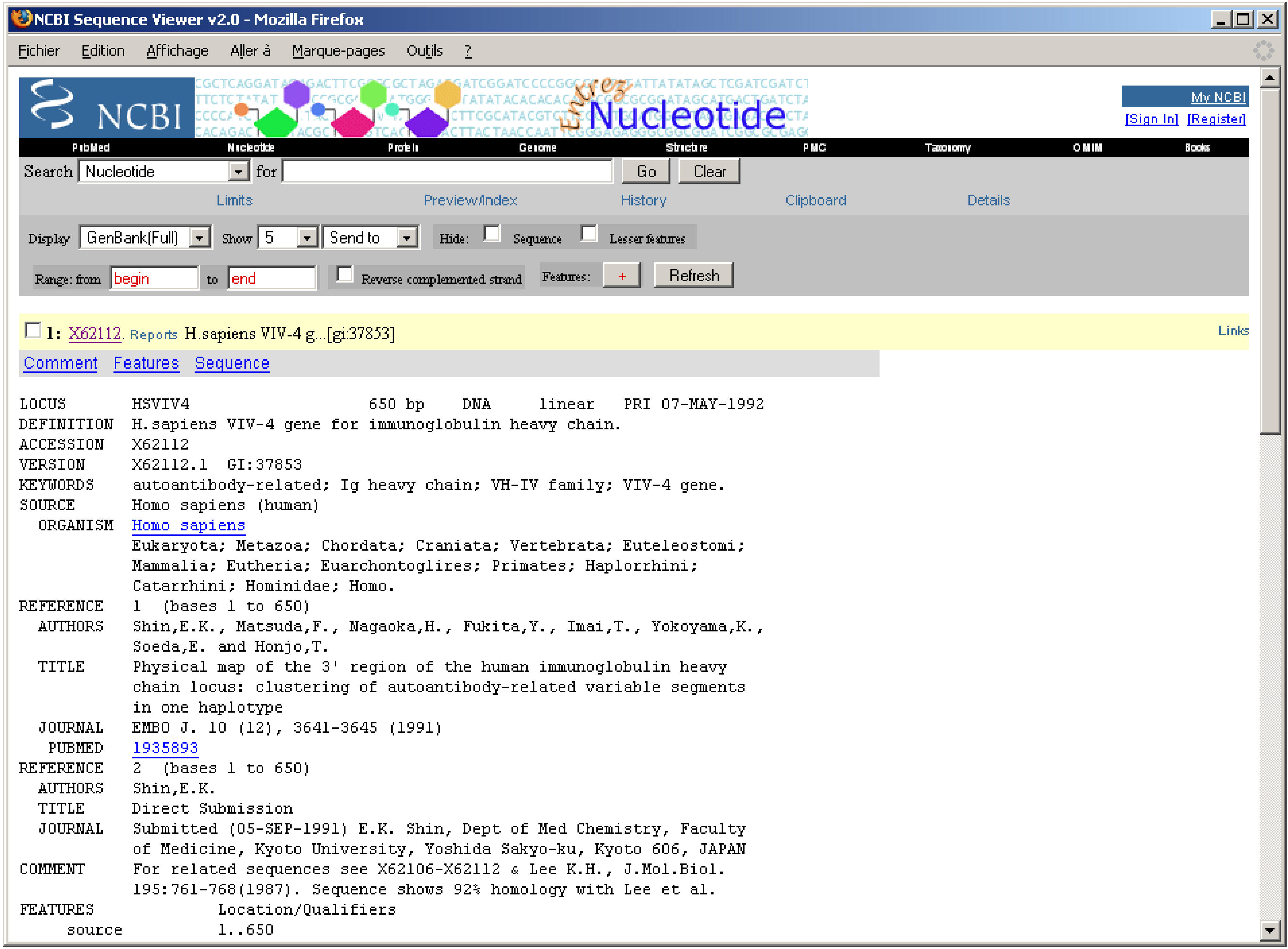
Access to GenBank

GenBank is available for searching at NCBI via several methods.

GenBank nucleotide records are available in the divisions CoreNucleotide, dbEST, or dbGSS and can be searched in Entrez together or independently. They must be searched independently in BLAST, however. CoreNucleotide is further split into several BLAST databases, including nr, htgs, wgs and env_nt. See the BLAST info page for more information about the numerous BLAST databases.

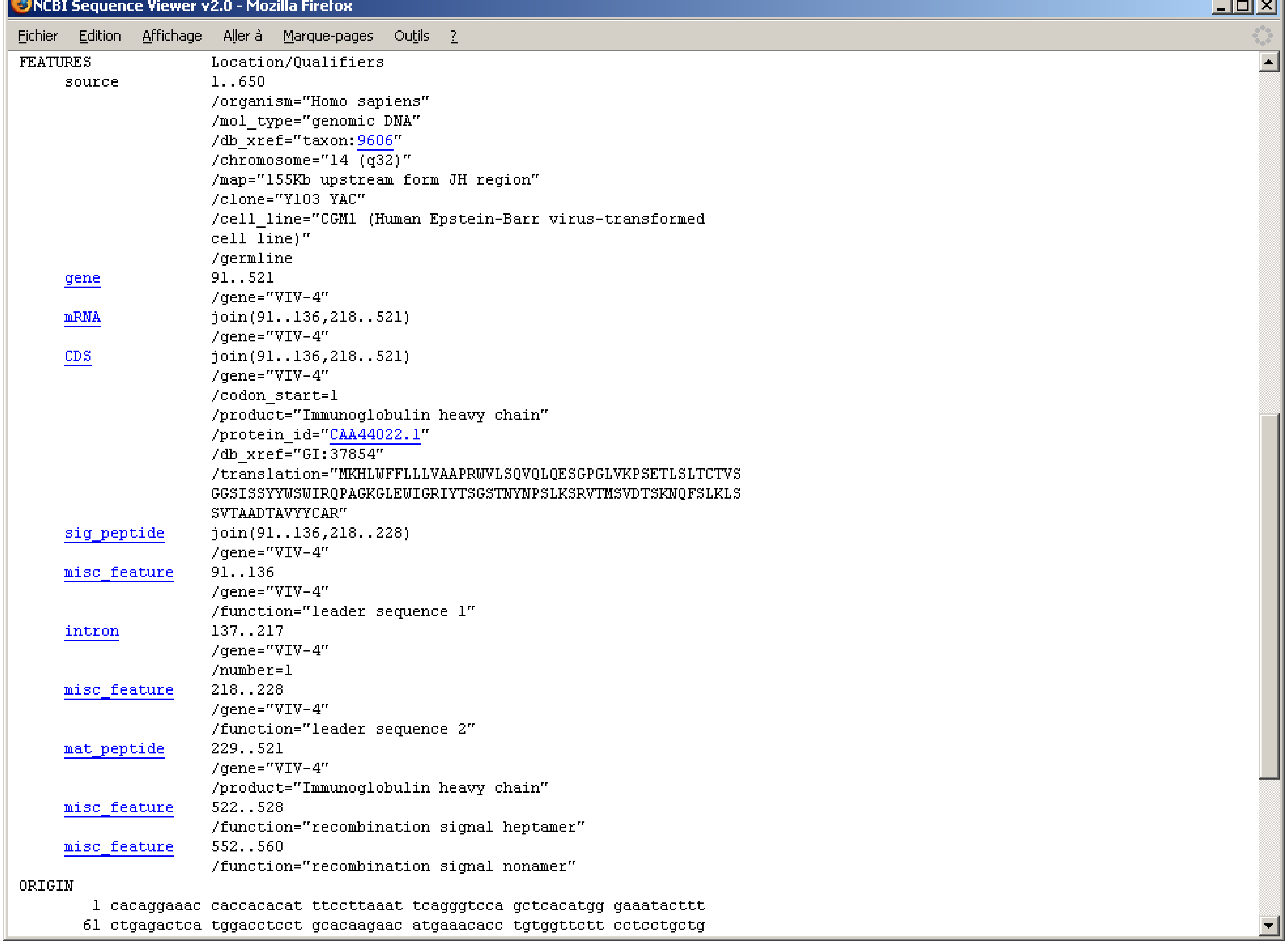
The GenBank database is designed to provide and encourage access within the scientific community to the most up to date and comprehensive DNA sequence information. Therefore, NCBI places no restrictions on the use or distribution of the GenBank data. However, some submitters may claim patent, copyright, or other intellectual property rights in all or a portion of the data they have submitted. NCBI is not in a position to assess the validity of such claims, and therefore cannot provide comment or unrestricted permission concerning the use, copying, or distribution of the information contained in GenBank.

New Developments



☐ **1:** X62112. Reports H.sapiens VIV-4 g...[gi:37853]

```
LOCUS       HSVIV4                   650 bp    DNA     linear   PRI 07-MAY-1992
DEFINITION  H.sapiens VIV-4 gene for immunoglobulin heavy chain.
ACCESSION   X62112
VERSION     X62112.1  GI:37853
KEYWORDS    autoantibody-related; Ig heavy chain; VH-IV family; VIV-4 gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 650)
  AUTHORS   Shin,E.K., Matsuda,F., Nagaoka,H., Fukita,Y., Imai,T., Yokoyama,K.,
            Soeda,E. and Honjo,T.
  TITLE     Physical map of the 3' region of the human immunoglobulin heavy
            chain locus: clustering of autoantibody-related variable segments
            in one haplotype
  JOURNAL   EMBO J. 10 (12), 3641-3645 (1991)
   PUBMED   1935893
REFERENCE   2  (bases 1 to 650)
  AUTHORS   Shin,E.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-SEP-1991) E.K. Shin, Dept of Med Chemistry, Faculty
            of Medicine, Kyoto University, Yoshida Sakyo-ku, Kyoto 606, JAPAN
COMMENT     For related sequences see X62106-X62112 & Lee K.H., J.Mol.Biol.
            195:761-768(1987). Sequence shows 92% homology with Lee et al.
FEATURES             Location/Qualifiers
     source          1..650
```



```
FEATURES             Location/Qualifiers
     source          1..650
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14 (q32)"
                     /map="155Kb upstream form JH region"
                     /clone="Y103 YAC"
                     /cell_line="CGM1 (Human Epstein-Barr virus-transformed
                     cell line)"
                     /germline
     gene            91..521
                     /gene="VIV-4"
     mRNA            join(91..136,218..521)
                     /gene="VIV-4"
     CDS             join(91..136,218..521)
                     /gene="VIV-4"
                     /codon_start=1
                     /product="Immunoglobulin heavy chain"
                     /protein_id="CAA44022.1"
                     /db_xref="GI:37854"
                     /translation="MKHLWFFLLLVAAPRWVLSQVQLQESGPGLVKPSETLSLTCTVS
                     GGSISSYYWSWIRQPAGKGLEWIGRIYTSGSTNYNPSLKSRVTMSVDTSKNQFSLKLS
                     SVTAADTAVYYCAR"
     sig_peptide     join(91..136,218..228)
                     /gene="VIV-4"
     misc_feature    91..136
                     /gene="VIV-4"
                     /function="leader sequence 1"
     intron          137..217
                     /gene="VIV-4"
                     /number=1
     misc_feature    218..228
                     /gene="VIV-4"
                     /function="leader sequence 2"
     mat_peptide     229..521
                     /gene="VIV-4"
                     /product="Immunoglobulin heavy chain"
     misc_feature    522..528
                     /function="recombination signal heptamer"
     misc_feature    552..560
                     /function="recombination signal nonamer"
ORIGIN
        1 cacaggaaac caccacacat ttccttaaat tcagggtcca gctcacatgg gaaatacttt
       61 ctgagactca tggacctcct gcacaagaac atgaaacacc tgtggttctt cctcctgctg
```

EMBL Nucleotide Sequence Database
Banque de séquences d'ADN et d'ARN

Créée en 1980

Localisation European Bioinformatics Institute, Hinxton, UK

Septembre 2006 (version 88) > 80.10^6 entrées (+ de 146.10^9 de nucléotides)

Origine des séquences

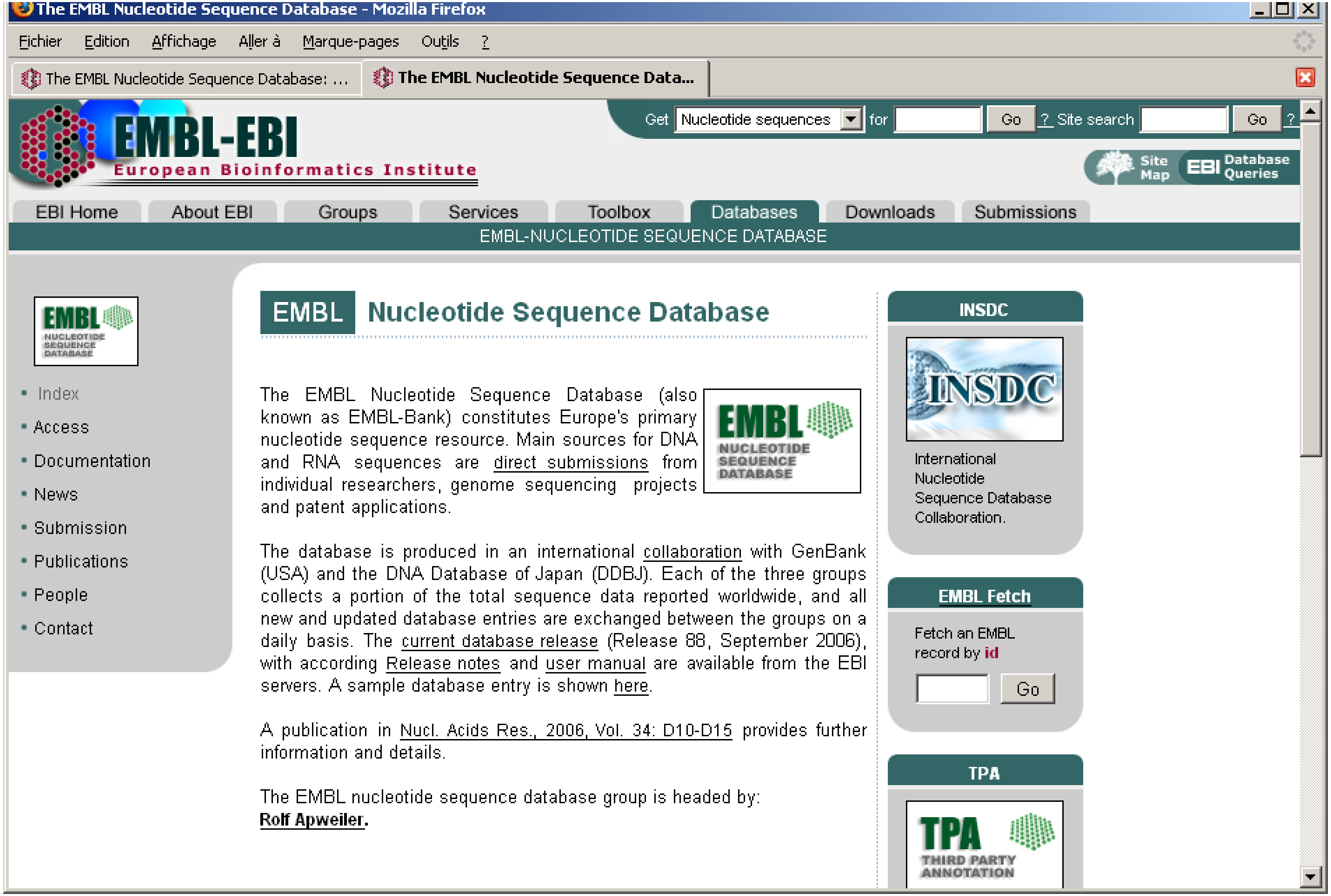
Analyse de la littérature européenne
Soumission par les auteurs
Collaboration avec GenBank et DDBJ

Outils de soumission des séquences

Sequin, Webin (http://www.ebi.ac.uk/embl/Submission/)

Consultation, distribution (fichiers à plat)

Web (http://www.ebi.ac.uk), SRS, FTP



EMBL Nucleotide Sequence Database

The EMBL Nucleotide Sequence Database (also known as EMBL-Bank) constitutes Europe's primary nucleotide sequence resource. Main sources for DNA and RNA sequences are direct submissions from individual researchers, genome sequencing projects and patent applications.

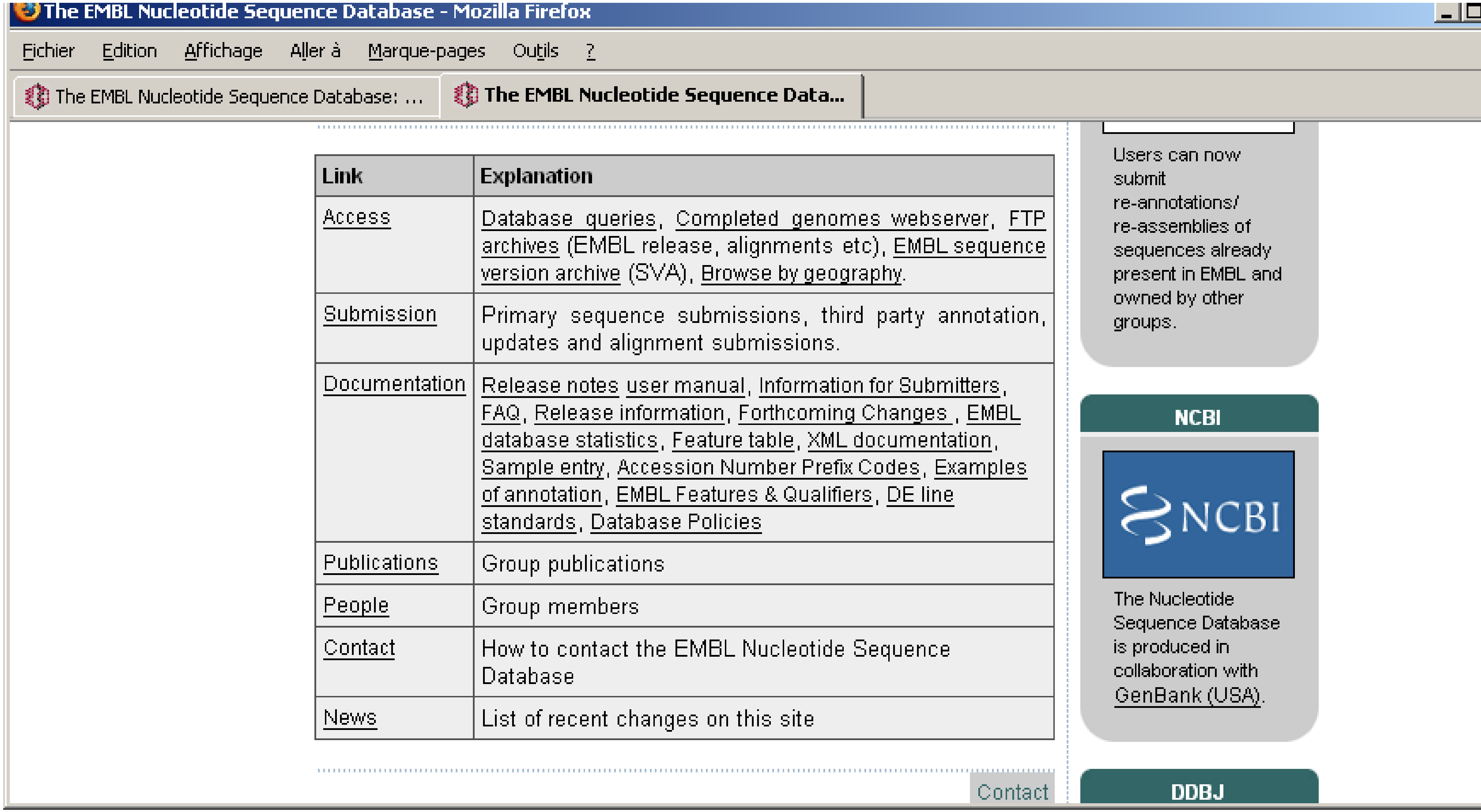
The database is produced in an international collaboration with GenBank (USA) and the DNA Database of Japan (DDBJ). Each of the three groups collects a portion of the total sequence data reported worldwide, and all new and updated database entries are exchanged between the groups on a daily basis. The current database release (Release 88, September 2006), with according Release notes and user manual are available from the EBI servers. A sample database entry is shown here.

A publication in Nucl. Acids Res., 2006, Vol. 34: D10-D15 provides further information and details.

The EMBL nucleotide sequence database group is headed by:
Rolf Apweiler.

International Nucleotide Sequence Database Collaboration.

Fetch an EMBL record by **id**

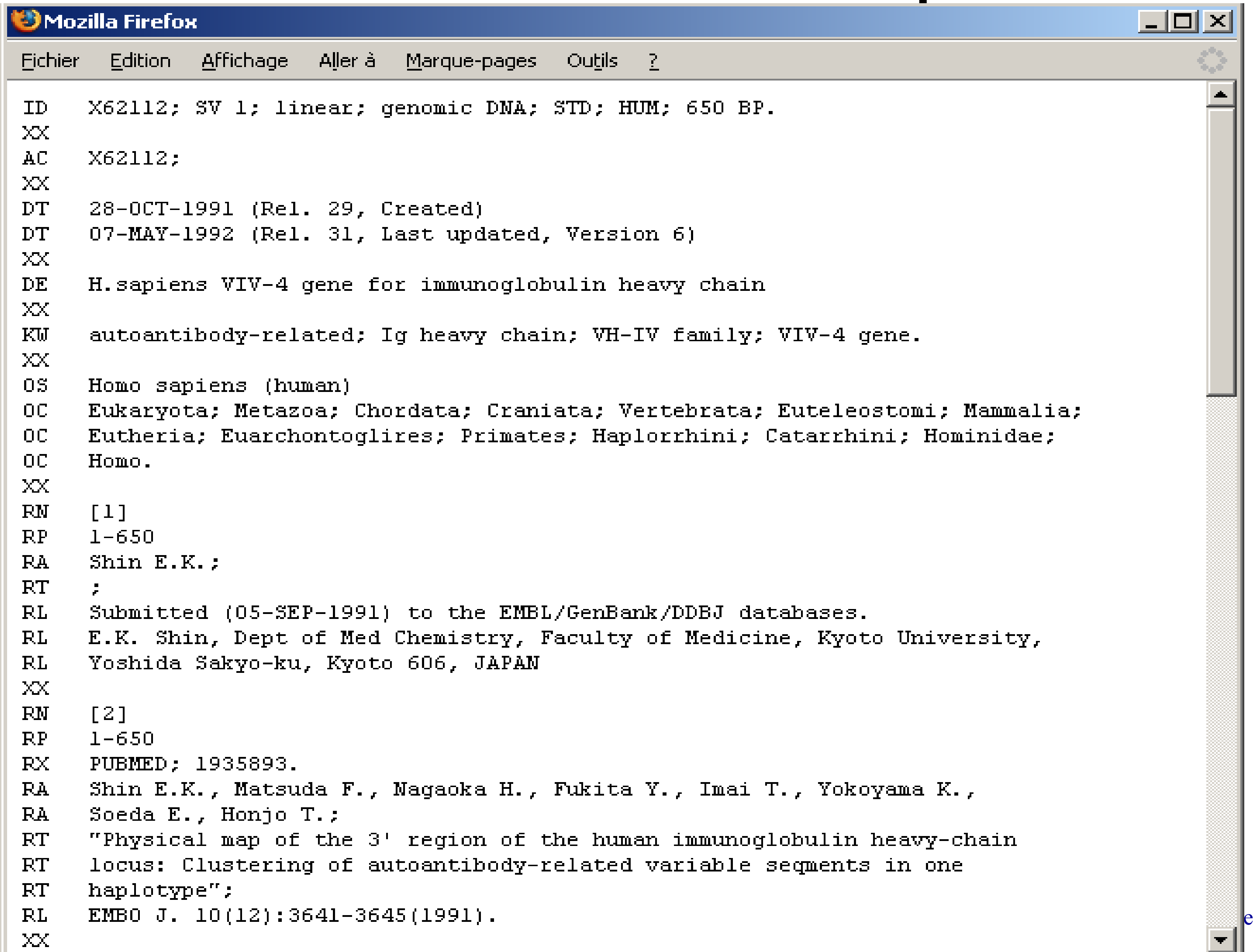


Link	Explanation
Access	Database queries, Completed genomes webserver, FTP archives (EMBL release, alignments etc), EMBL sequence version archive (SVA), Browse by geography.
Submission	Primary sequence submissions, third party annotation, updates and alignment submissions.
Documentation	Release notes user manual, Information for Submitters, FAQ, Release information, Forthcoming Changes , EMBL database statistics, Feature table, XML documentation, Sample entry, Accession Number Prefix Codes, Examples of annotation, EMBL Features & Qualifiers, DE line standards, Database Policies
Publications	Group publications
People	Group members
Contact	How to contact the EMBL Nucleotide Sequence Database
News	List of recent changes on this site

Users can now submit re-annotations/ re-assemblies of sequences already present in EMBL and owned by other groups.

The Nucleotide Sequence Database is produced in collaboration with GenBank (USA).

Structure des informations distribuées par EMBL



```
ID   X62112; SV 1; linear; genomic DNA; STD; HUM; 650 BP.
XX
AC   X62112;
XX
DT   28-OCT-1991 (Rel. 29, Created)
DT   07-MAY-1992 (Rel. 31, Last updated, Version 6)
XX
DE   H.sapiens VIV-4 gene for immunoglobulin heavy chain
XX
KW   autoantibody-related; Ig heavy chain; VH-IV family; VIV-4 gene.
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   1-650
RA   Shin E.K.;
RT   ;
RL   Submitted (05-SEP-1991) to the EMBL/GenBank/DDBJ databases.
RL   E.K. Shin, Dept of Med Chemistry, Faculty of Medicine, Kyoto University,
RL   Yoshida Sakyo-ku, Kyoto 606, JAPAN
XX
RN   [2]
RP   1-650
RX   PUBMED; 1935893.
RA   Shin E.K., Matsuda F., Nagaoka H., Fukita Y., Imai T., Yokoyama K.,
RA   Soeda E., Honjo T.;
RT   "Physical map of the 3' region of the human immunoglobulin heavy-chain
RT   locus: Clustering of autoantibody-related variable segments in one
RT   haplotype";
RL   EMBO J. 10(12):3641-3645(1991).
XX
```

Les classes EMBL

La classe de données pour chaque entrée reflète l'approche méthodologique utilisée pour la génération de la séquence

CON	Entry constructed from segment entry sequences, drawing annotation from segment entries
ANN	Entry constructed from segment entry sequences with its own annotation
PAT	Patent EST Expressed Sequence Tag
GSS	Genome Survey Sequence, short single pass genomic sequences
HTC	High Thoughput CDNA sequencing
HTG	High Thoughput Genome sequencing
MGA	Mass Genome Annotation
WGS	Whole Genome Shotgun
TPA	Third Party Annotation
STS	Sequence Tagged Site
STD	Standard (all entries not classified as above)

Les divisions EMBL et GenBank

Création de divisions taxonomiques

EMBL

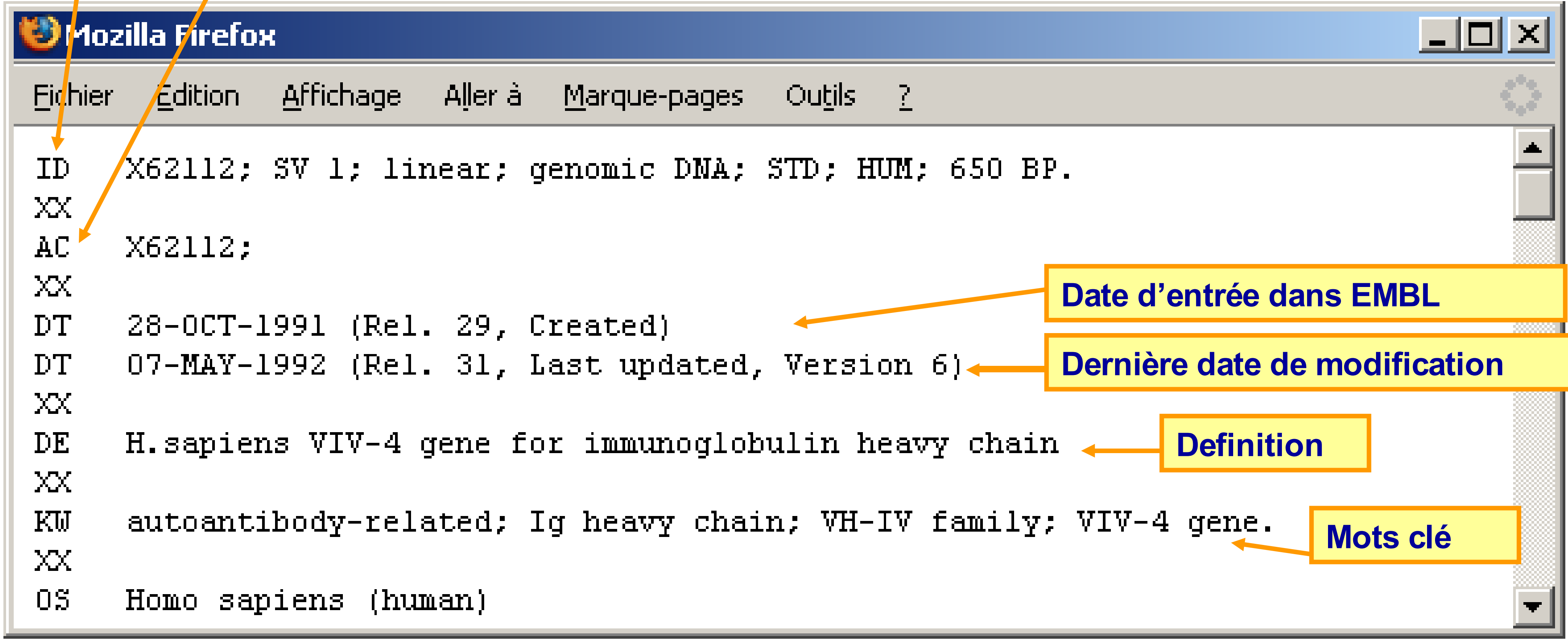
Bacteriophage	PHG
Environmental Sample	ENV
Fungal	FUN
Human	HUM
Invertebrate	INV
Other Mammal	MAM
Other Vertebrate	VRT
Mus musculus	MUS
Plant	PLN
Prokaryote	PRO
Other Rodent	ROD
Synthetic	SYN
Unclassified	UNC
Viral	VRL

GenBanK

PRI - primate sequences
ROD - rodent sequences
MAM - other mammalian sequences
VRT - other vertebrate sequences
INV - invertebrate sequences
PLN - plant, fungal, and algal sequences
BCT - bacterial sequences
VRL - viral sequences
PHG - bacteriophage sequences
SYN - synthetic sequences
UNA - unannotated sequences
EST - EST sequences (expressed sequence tags)
PAT - patent sequences
STS - STS sequences (sequence tagged sites)
GSS - GSS sequences (genome survey sequences)
HTG - HTGS sequences (high throughput genomic sequences)
HTC - HTC sequences (high throughput cDNA sequences)
ENV - Environmental sampling sequences

nom de l'entrée; version de la séquence; classe de la donnée; molécule; division; longueur.

Numéro d’accès



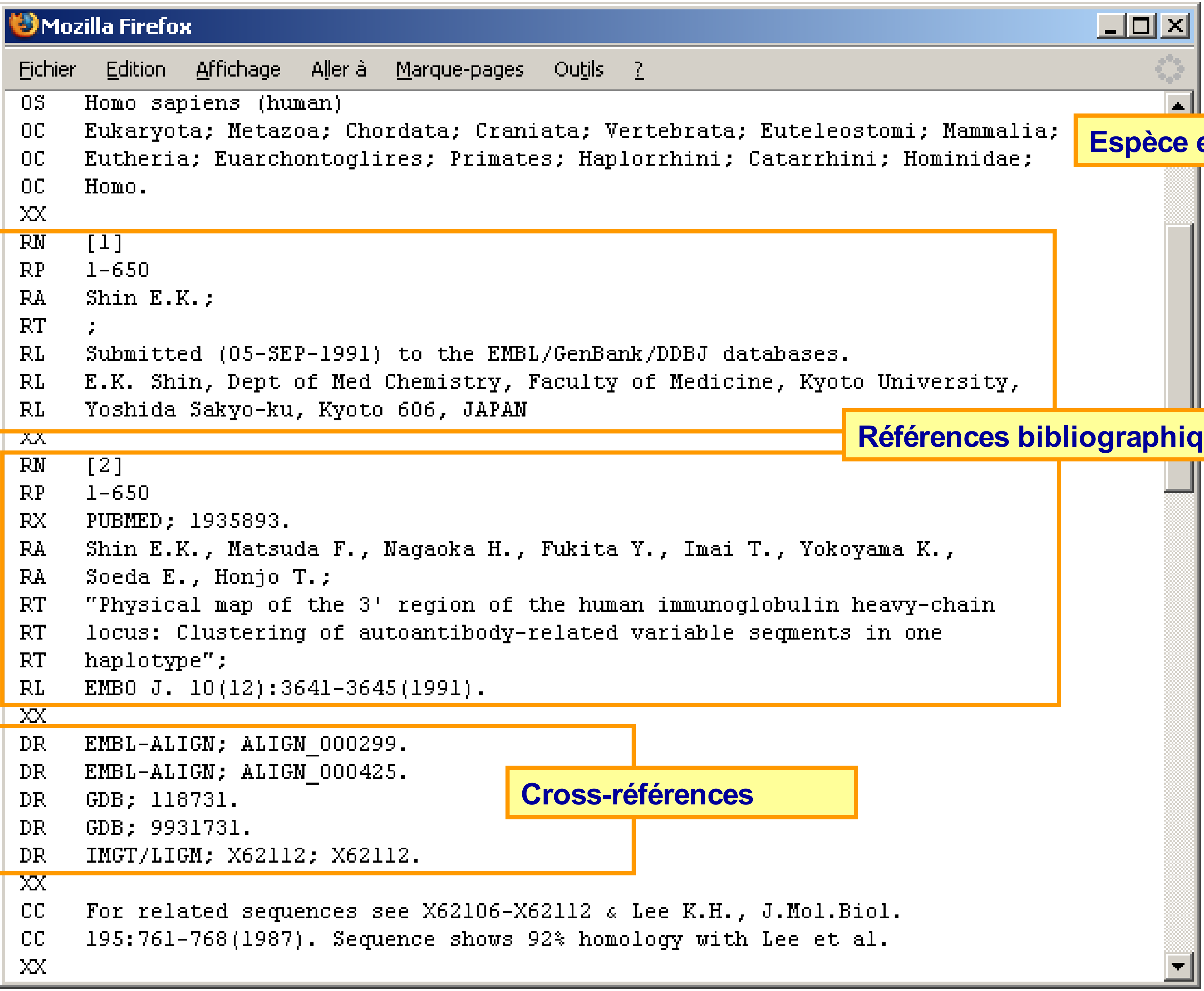
```
ID   X62112; SV 1; linear; genomic DNA; STD; HUM; 650 BP.
XX
AC   X62112;
XX
DT   28-OCT-1991 (Rel. 29, Created)
DT   07-MAY-1992 (Rel. 31, Last updated, Version 6)
XX
DE   H.sapiens VIV-4 gene for immunoglobulin heavy chain
XX
KW   autoantibody-related; Ig heavy chain; VH-IV family; VIV-4 gene.
XX
OS   Homo sapiens (human)
```

Date d’entrée dans EMBL

Dernière date de modification

Definition

Mots clé



```
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   1-650
RA   Shin E.K.;
RT   ;
RL   Submitted (05-SEP-1991) to the EMBL/GenBank/DDBJ databases.
RL   E.K. Shin, Dept of Med Chemistry, Faculty of Medicine, Kyoto University,
RL   Yoshida Sakyo-ku, Kyoto 606, JAPAN
XX
RN   [2]
RP   1-650
RX   PUBMED; 1935893.
RA   Shin E.K., Matsuda F., Nagaoka H., Fukita Y., Imai T., Yokoyama K.,
RA   Soeda E., Honjo T.;
RT   "Physical map of the 3' region of the human immunoglobulin heavy-chain
RT   locus: Clustering of autoantibody-related variable segments in one
RT   haplotype";
RL   EMBO J. 10(12):3641-3645(1991).
XX
DR   EMBL-ALIGN; ALIGN_000299.
DR   EMBL-ALIGN; ALIGN_000425.
DR   GDB; 118731.
DR   GDB; 9931731.
DR   IMGT/LIGM; X62112; X62112.
XX
CC   For related sequences see X62106-X62112 & Lee K.H., J.Mol.Biol.
CC   195:761-768(1987). Sequence shows 92% homology with Lee et al.
XX
```

Espèce et classification

Références bibliographiques

Cross-références

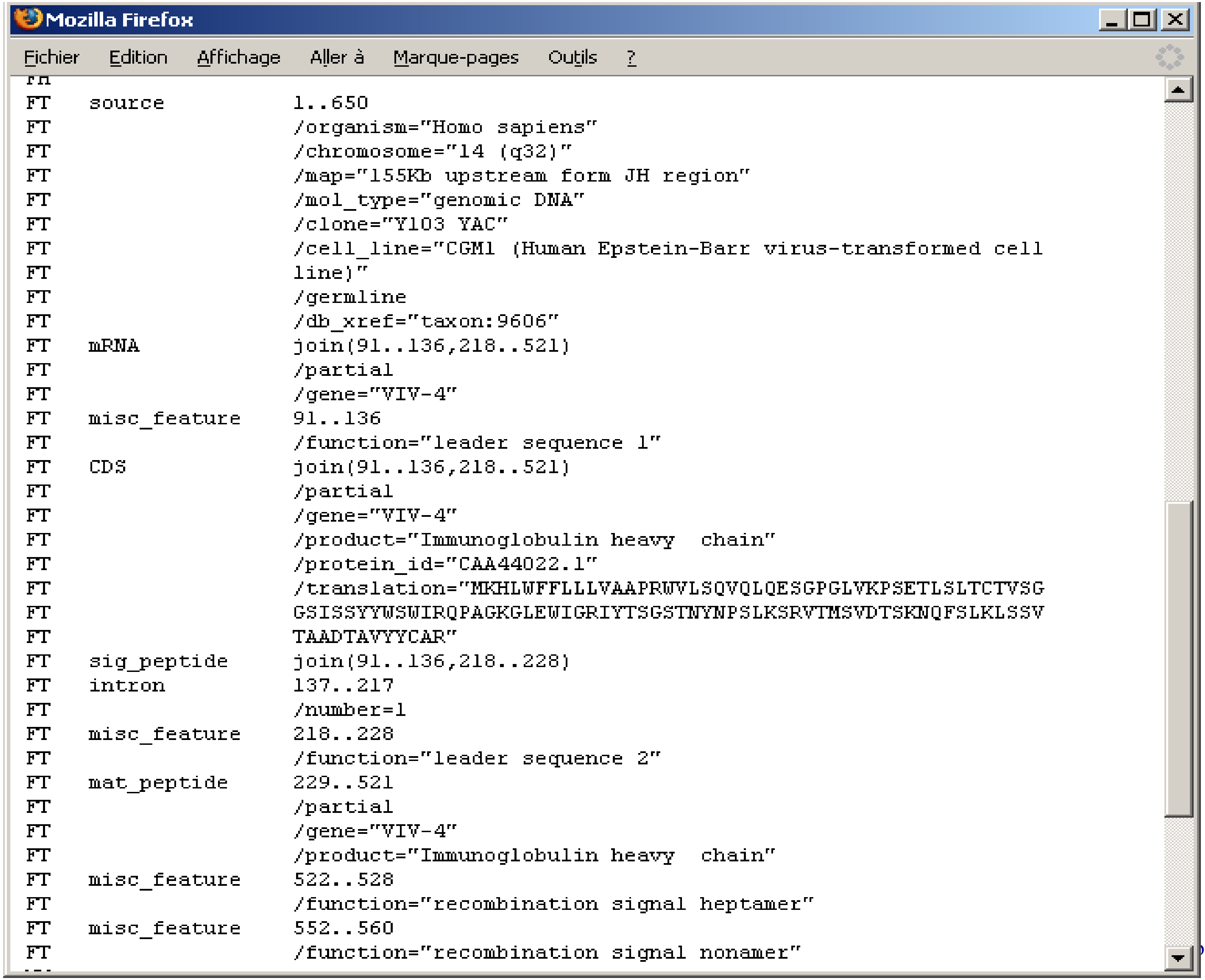
Les lignes et codes de EMBL

CC	Commentaires.
FH	En-tête du champ FT (feature)
FT	Motifs/caractéristiques de la séquence. **"The feature table"**.
SQ	Longueur de la séquence en paire de bases, composition Séquence avec 60 nucléotides par ligne dans le sens 5'--->3'.

Les lignes FT: les annotations

Features	Caractérisation des motifs de la séquence présentation dans un ordre hierarchique comprend le nom et la delimitation du motif dans la séquence
Qualifiers	ajoutent des informations concernant la feature comprend un nom et un texte associé (texte libre ou vocabulaire contrôlé)

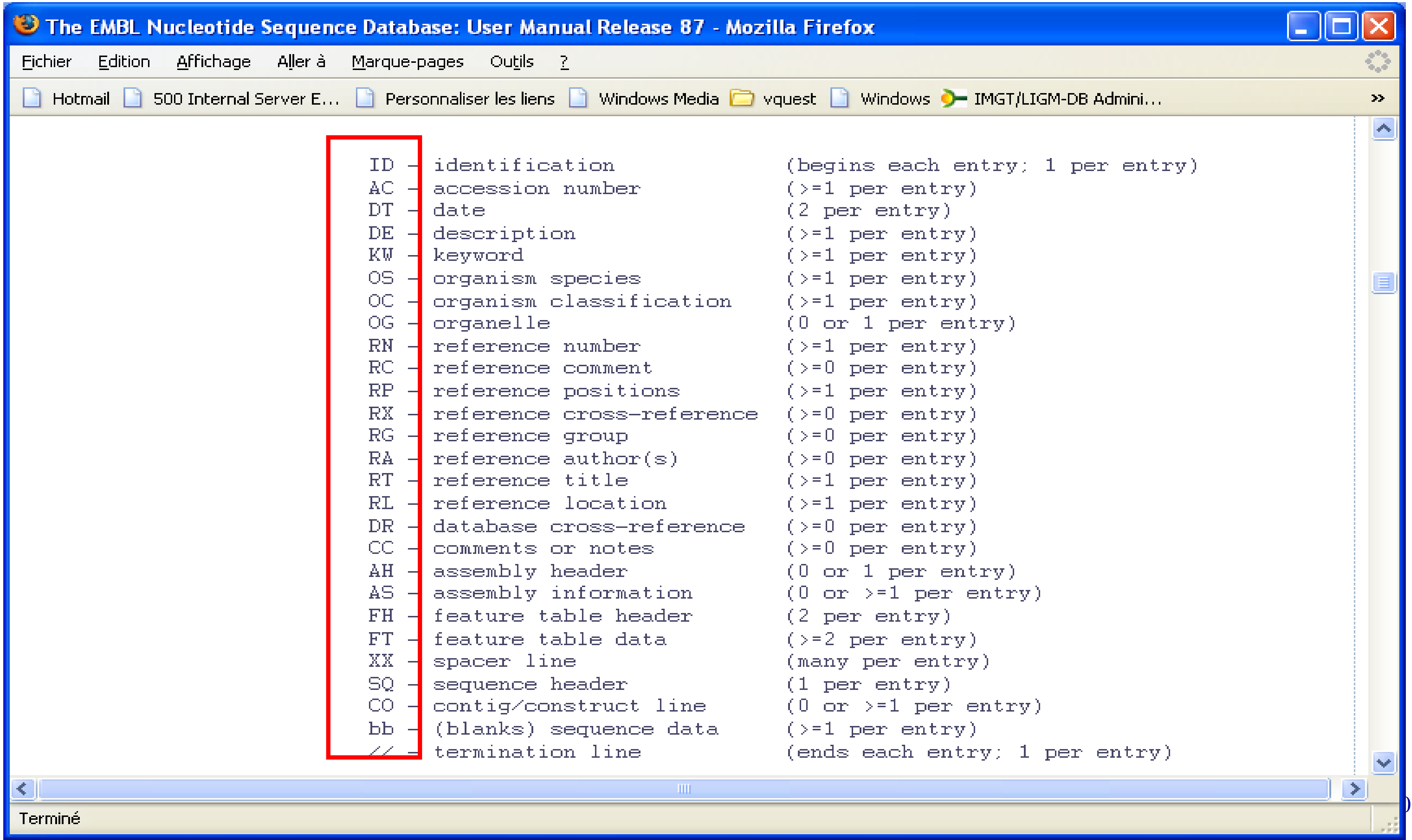
```
FT CDS              1..391
FT                  /codon_start=2
FT                  /db_xref="PID:e1358278"
FT                  /partial
FT                  /gene="IgH"
FT                  /product="immunoglobulin heavy chain"
FT                  /translation="MSWVFLVAILEGVQCEVQLVESGGGLVQPGGSLRLSCAASGFTFS"
```



```
FH
FT   source          1..650
FT                   /organism="Homo sapiens"
FT                   /chromosome="14 (q32)"
FT                   /map="155Kb upstream form JH region"
FT                   /mol_type="genomic DNA"
FT                   /clone="Y103 YAC"
FT                   /cell_line="CGM1 (Human Epstein-Barr virus-transformed cell
FT                   line)"
FT                   /germline
FT                   /db_xref="taxon:9606"
FT   mRNA            join(91..136,218..521)
FT                   /partial
FT                   /gene="VIV-4"
FT   misc_feature    91..136
FT                   /function="leader sequence 1"
FT   CDS             join(91..136,218..521)
FT                   /partial
FT                   /gene="VIV-4"
FT                   /product="Immunoglobulin heavy  chain"
FT                   /protein_id="CAA44022.1"
FT                   /translation="MKHLWFFLLLVAAPRWVLSQVQLQESGPGLVKPSETLSLTCTVSG
FT                   GSISSYYWSWIRQPAGKGLEWIGRIYTSGSTNYNPSLKSRVTMSVDTSKNQFSLKLSSV
FT                   TAADTAVYYCAR"
FT   sig_peptide     join(91..136,218..228)
FT   intron          137..217
FT                   /number=1
FT   misc_feature    218..228
FT                   /function="leader sequence 2"
FT   mat_peptide     229..521
FT                   /partial
FT                   /gene="VIV-4"
FT                   /product="Immunoglobulin heavy  chain"
FT   misc_feature    522..528
FT                   /function="recombination signal heptamer"
FT   misc_feature    552..560
FT                   /function="recombination signal nonamer"
```

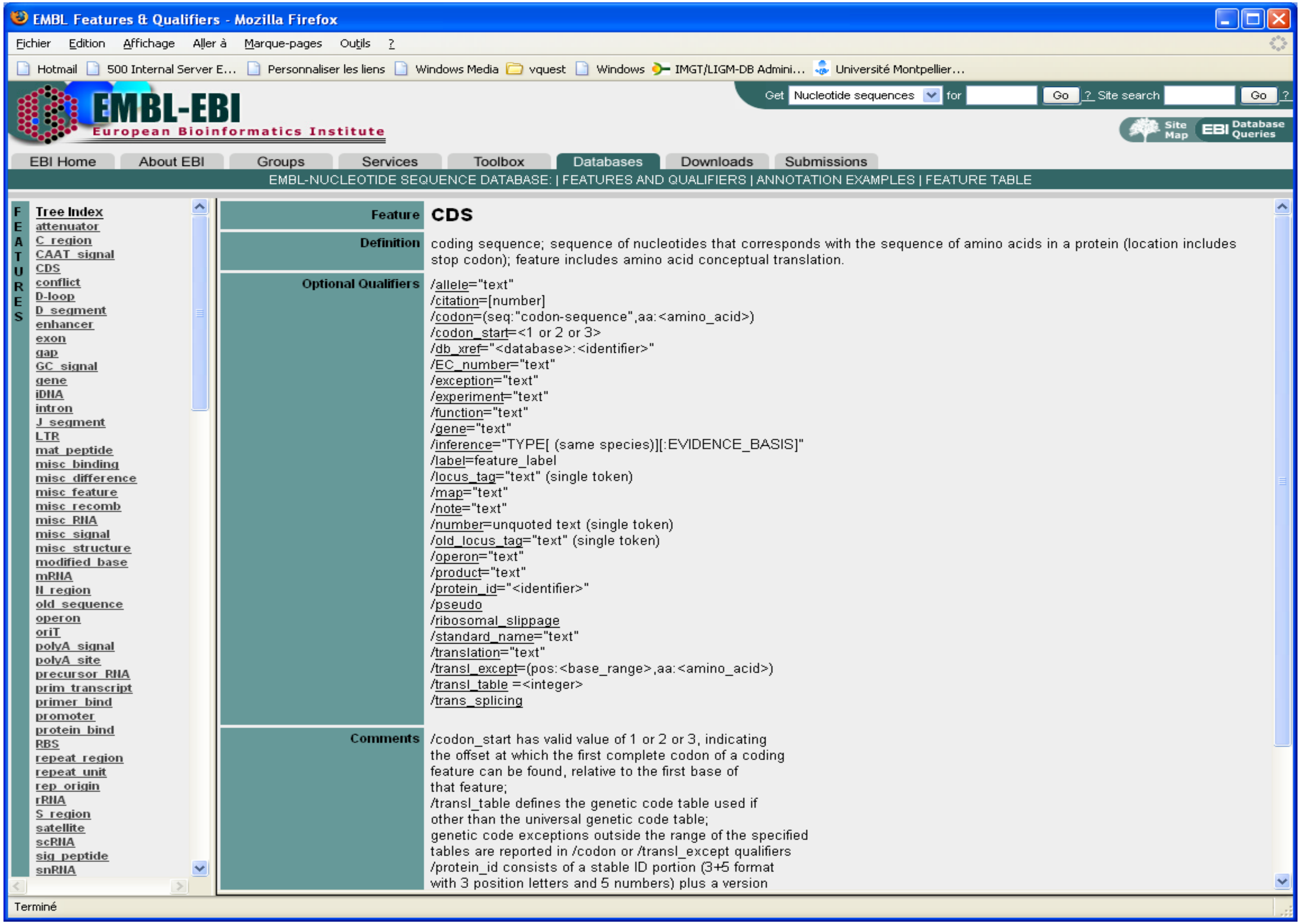
Structure de l'information dans les bases de données généralistes DDBJ/EMBL/GenBank

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)
```

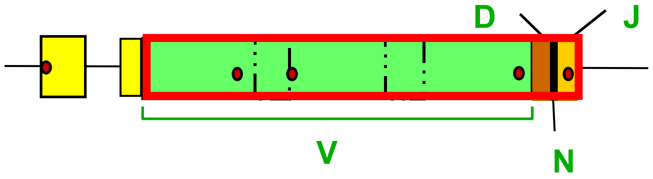
EMBL Feature labels

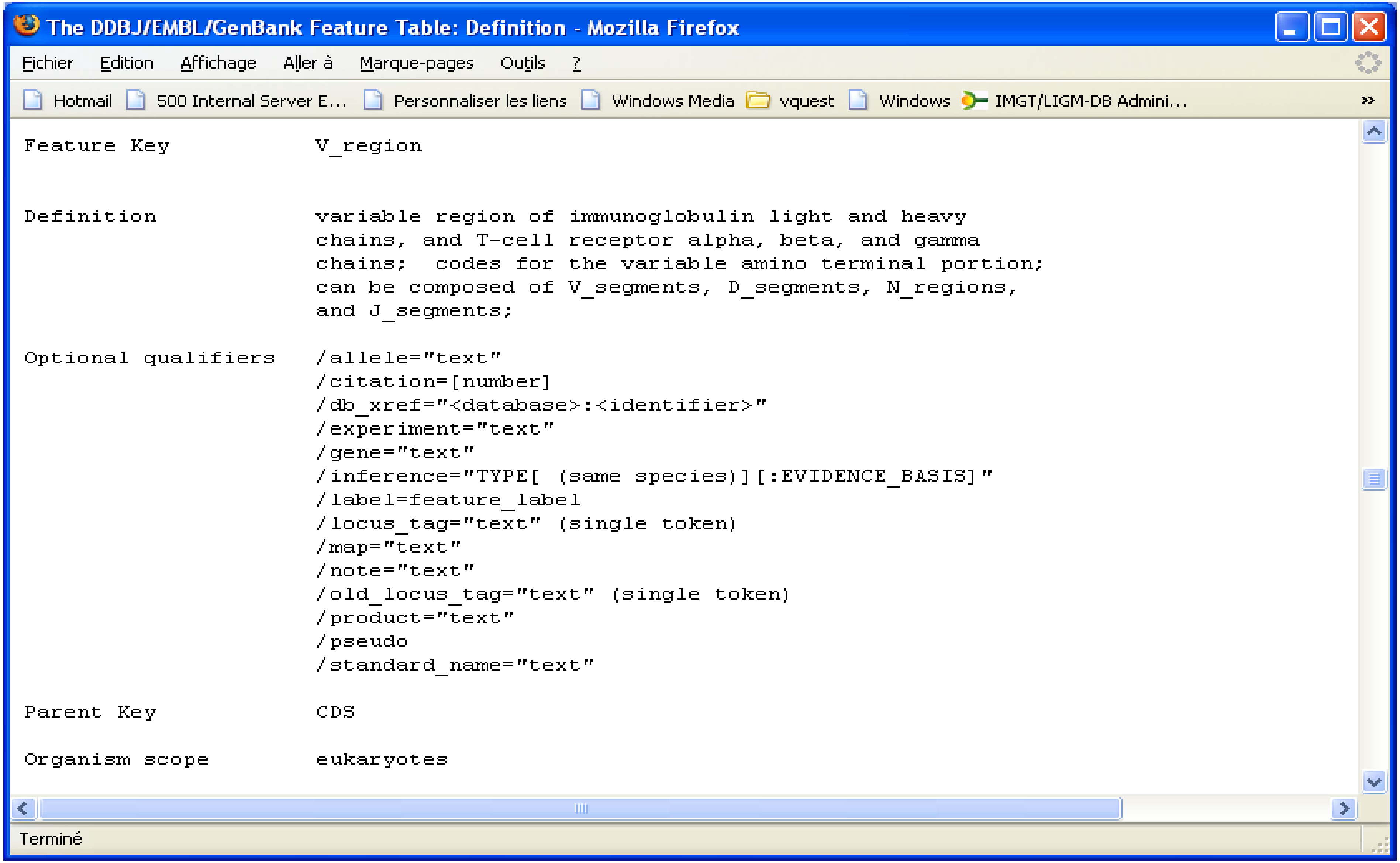


Feature	**CDS**
Definition	coding sequence; sequence of nucleotides that corresponds with the sequence of amino acids in a protein (location includes stop codon); feature includes amino acid conceptual translation.
Optional Qualifiers	/allele="text" /citation=[number] /codon=(seq:"codon-sequence",aa:<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" /translation="text" /transl_except=(pos:<base_range>,aa:<amino_acid>) /transl_table =<integer> /trans_splicing
Comments	/codon_start has valid value of 1 or 2 or 3, indicating the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature; /transl_table defines the genetic code table used if other than the universal genetic code table; genetic code exceptions outside the range of the specified tables are reported in /codon or /transl_except qualifiers /protein_id consists of a stable ID portion (3+5 format with 3 position letters and 5 numbers) plus a version

http://www3.ebi.ac.uk/Services/WebFeat/

EMBL Feature labels

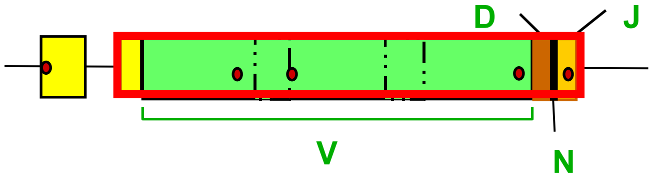


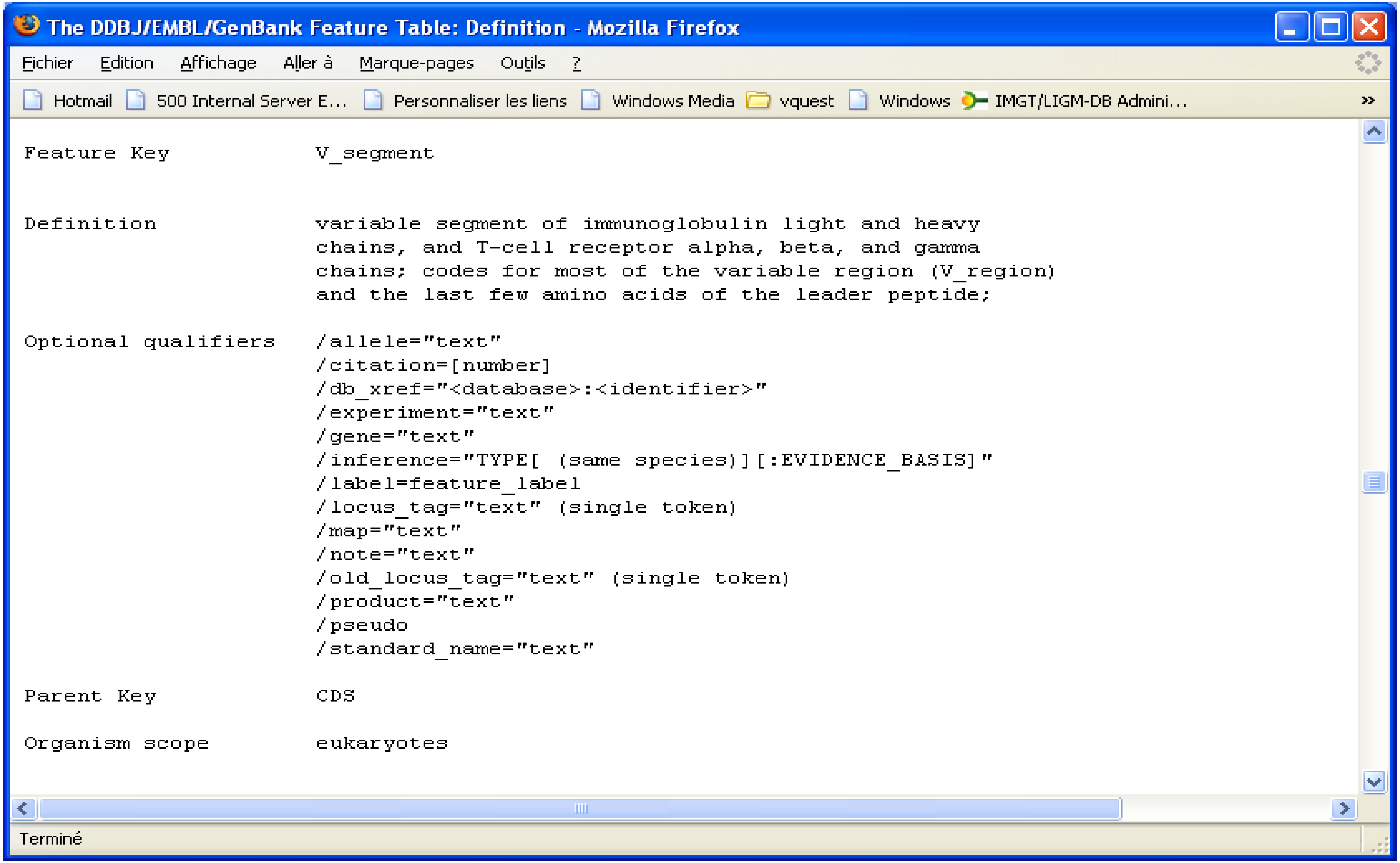


```
Feature Key           V_region


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"

Parent Key            CDS

Organism scope        eukaryotes
```

http://www.ebi.ac.uk/embl/Documentation/FT_definitions/feature_table.html

EMBL Feature labels





```
Feature Key           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="<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"

Parent Key            CDS

Organism scope        eukaryotes
```

http://www.ebi.ac.uk/embl/Documentation/FT_definitions/feature_table.html

Les bases de données généralistes pour les protéines

SWISS-PROT	Université de Genève, Suisse EBI, UK http://us.expasy.org/sprot/ http://www.ebi.ac.uk
PIR	NBRF, USA http://pir.georgetown.edu/
UNIPROT	http://www.expasy.uniprot.org/

Welcome to UniProt

UniProt (Universal Protein Resource) is the world's most comprehensive catalog of information on proteins. It is a central repository of protein sequence and function created by joining the information contained in Swiss-Prot, TrEMBL, and PIR.

UniProt is comprised of three components, each optimized for different uses. The **UniProt Knowledgebase (UniProtKB)** is the central access point for extensive curated protein information, including function, classification, and cross-reference. The **UniProt Reference Clusters (UniRef)** databases combine closely related sequences into a single record to speed searches. The **UniProt Archive (UniParc)** is a comprehensive repository, reflecting the history of all protein sequences.

The sequences and information in UniProt are accessible via text search, BLAST similarity search, and FTP.

SWISS-PROT
Banque de séquences protéiques

Créée en 1986 **par Amos Bairoth, Suisse**

Privilégie la qualité des données, n’accepte pas toutes les séquences
Un complément: TREMBL

Localisation Université de Genève, Suisse; EBI, UK

October 2006 **>**234.000 séquences

Origine des séquences

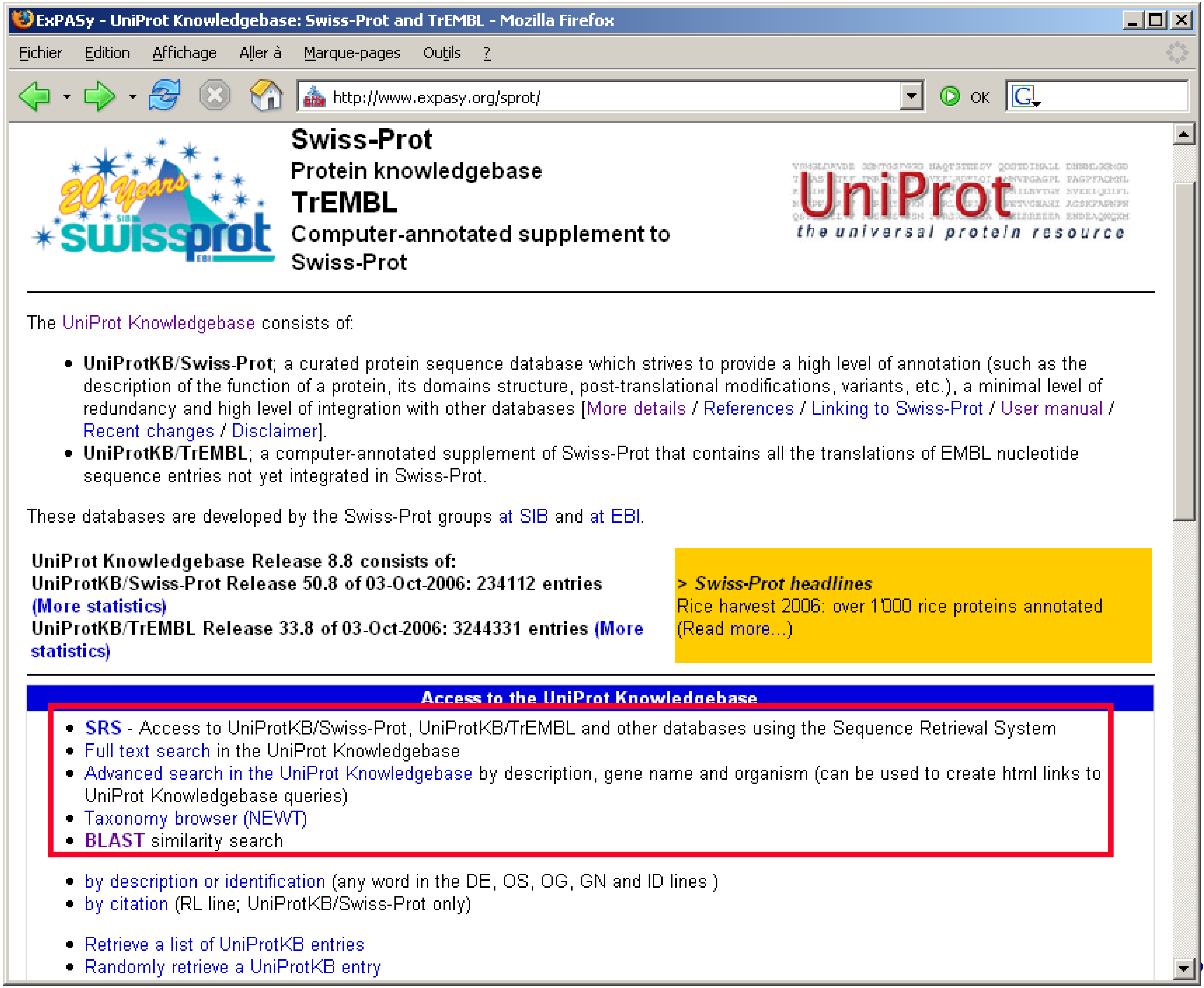
Analyse de la littérature, soumission par les auteurs,
traduction de séquences nucléotidiques

Distribution (fichiers à plat)

Web (http://www.expasy.org/sprot, http://www.ebi.ac.uk), FTP

Outils de soumission des séquences

SPIN http://www.ebi.ac.uk/swissprot/Submissions/spin/index.jsp



Swiss-Prot
Protein knowledgebase
TrEMBL
Computer-annotated supplement to Swiss-Prot

The UniProt Knowledgebase consists of:

- **UniProtKB/Swiss-Prot**; a curated protein sequence database which strives to provide a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases [More details / References / Linking to Swiss-Prot / User manual / Recent changes / Disclaimer].
- **UniProtKB/TrEMBL**; a computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot.

These databases are developed by the Swiss-Prot groups at SIB and at EBI.

UniProt Knowledgebase Release 8.8 consists of:
UniProtKB/Swiss-Prot Release 50.8 of 03-Oct-2006: 234112 entries (More statistics)
UniProtKB/TrEMBL Release 33.8 of 03-Oct-2006: 3244331 entries (More statistics)

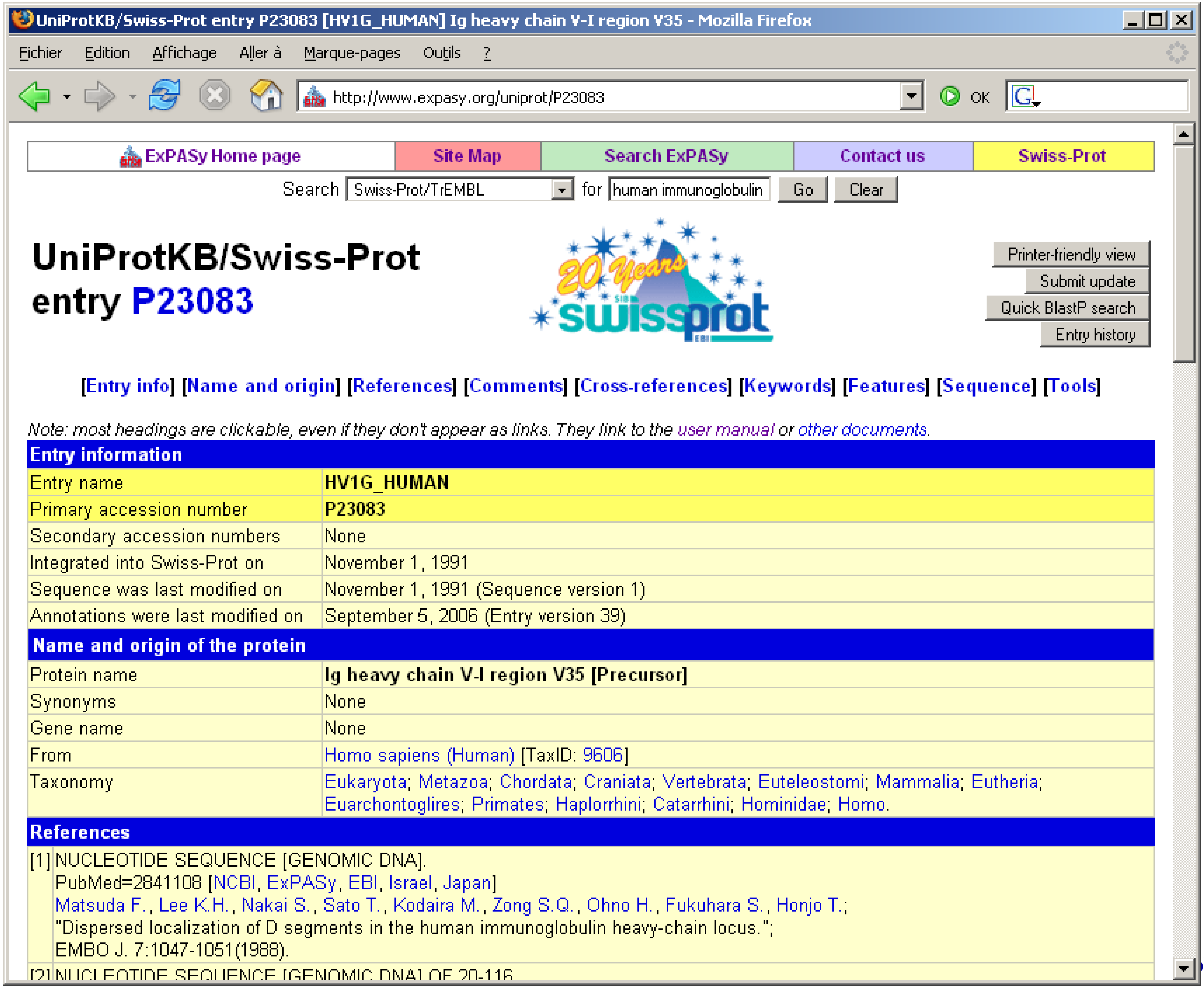
> Swiss-Prot headlines
Rice harvest 2006: over 1'000 rice proteins annotated (Read more...)

Access to the UniProt Knowledgebase

- **SRS** - Access to UniProtKB/Swiss-Prot, UniProtKB/TrEMBL and other databases using the Sequence Retrieval System
- Full text search in the UniProt Knowledgebase
- Advanced search in the UniProt Knowledgebase by description, gene name and organism (can be used to create html links to UniProt Knowledgebase queries)
- Taxonomy browser (NEWT)
- **BLAST** similarity search

- by description or identification (any word in the DE, OS, OG, GN and ID lines)
- by citation (RL line; UniProtKB/Swiss-Prot only)

- Retrieve a list of UniProtKB entries
- Randomly retrieve a UniProtKB entry



UniProtKB/Swiss-Prot entry P23083

[Entry info] [Name and origin] [References] [Comments] [Cross-references] [Keywords] [Features] [Sequence] [Tools]

Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.

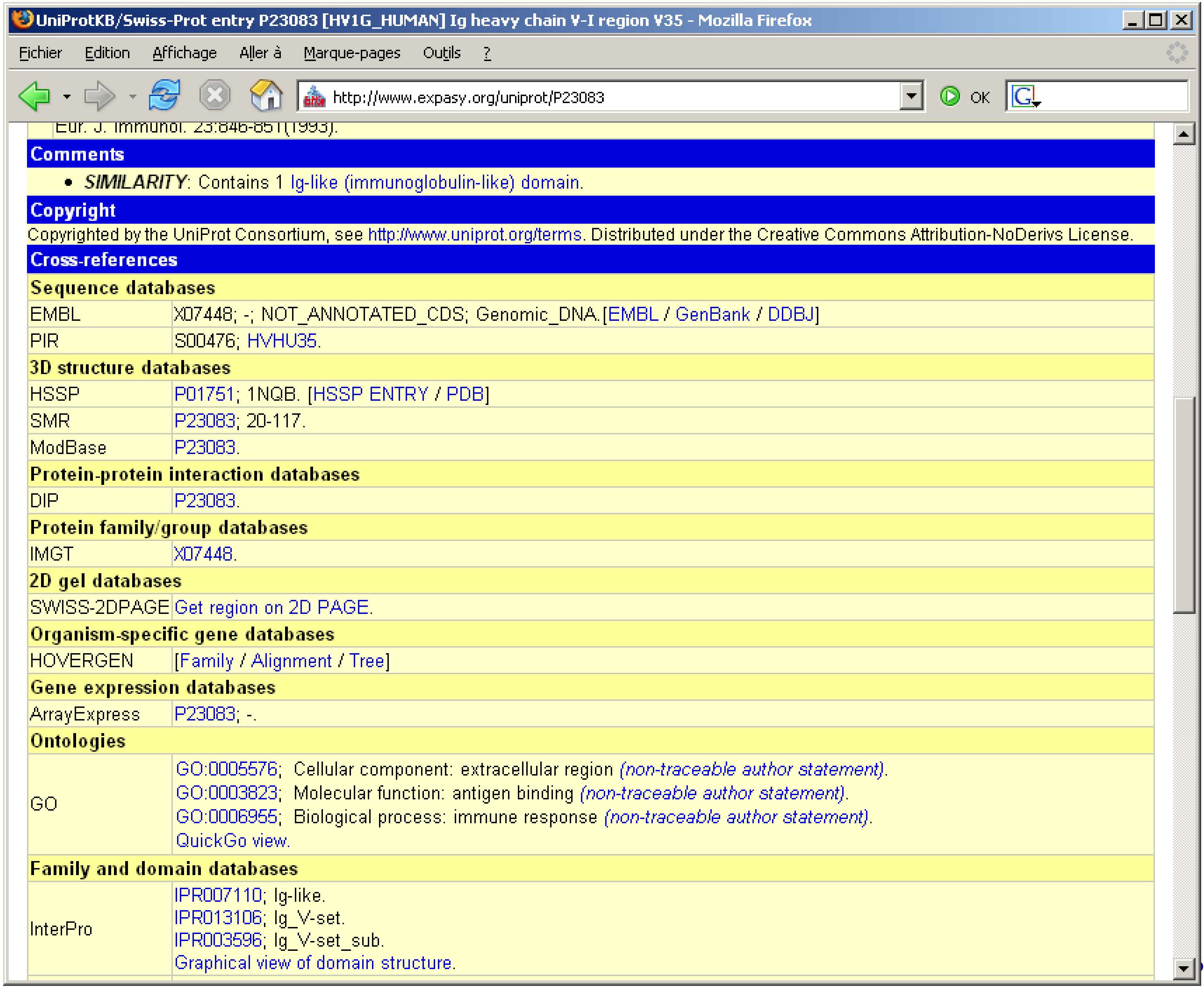
Entry information	
Entry name	**HV1G_HUMAN**
Primary accession number	**P23083**
Secondary accession numbers	None
Integrated into Swiss-Prot on	November 1, 1991
Sequence was last modified on	November 1, 1991 (Sequence version 1)
Annotations were last modified on	September 5, 2006 (Entry version 39)

Name and origin of the protein	
Protein name	**Ig heavy chain V-I region V35 [Precursor]**
Synonyms	None
Gene name	None
From	Homo sapiens (Human) [TaxID: 9606]
Taxonomy	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

References

[1] NUCLEOTIDE SEQUENCE [GENOMIC DNA].
PubMed=2841108 [NCBI, ExPASy, EBI, Israel, Japan]
Matsuda F., Lee K.H., Nakai S., Sato T., Kodaira M., Zong S.Q., Ohno H., Fukuhara S., Honjo T.;
"Dispersed localization of D segments in the human immunoglobulin heavy-chain locus.";
EMBO J. 7:1047-1051(1988).

[2] NUCLEOTIDE SEQUENCE [GENOMIC DNA] OF 20-116



Eur. J. Immunol. 23:846-851(1993).

Comments

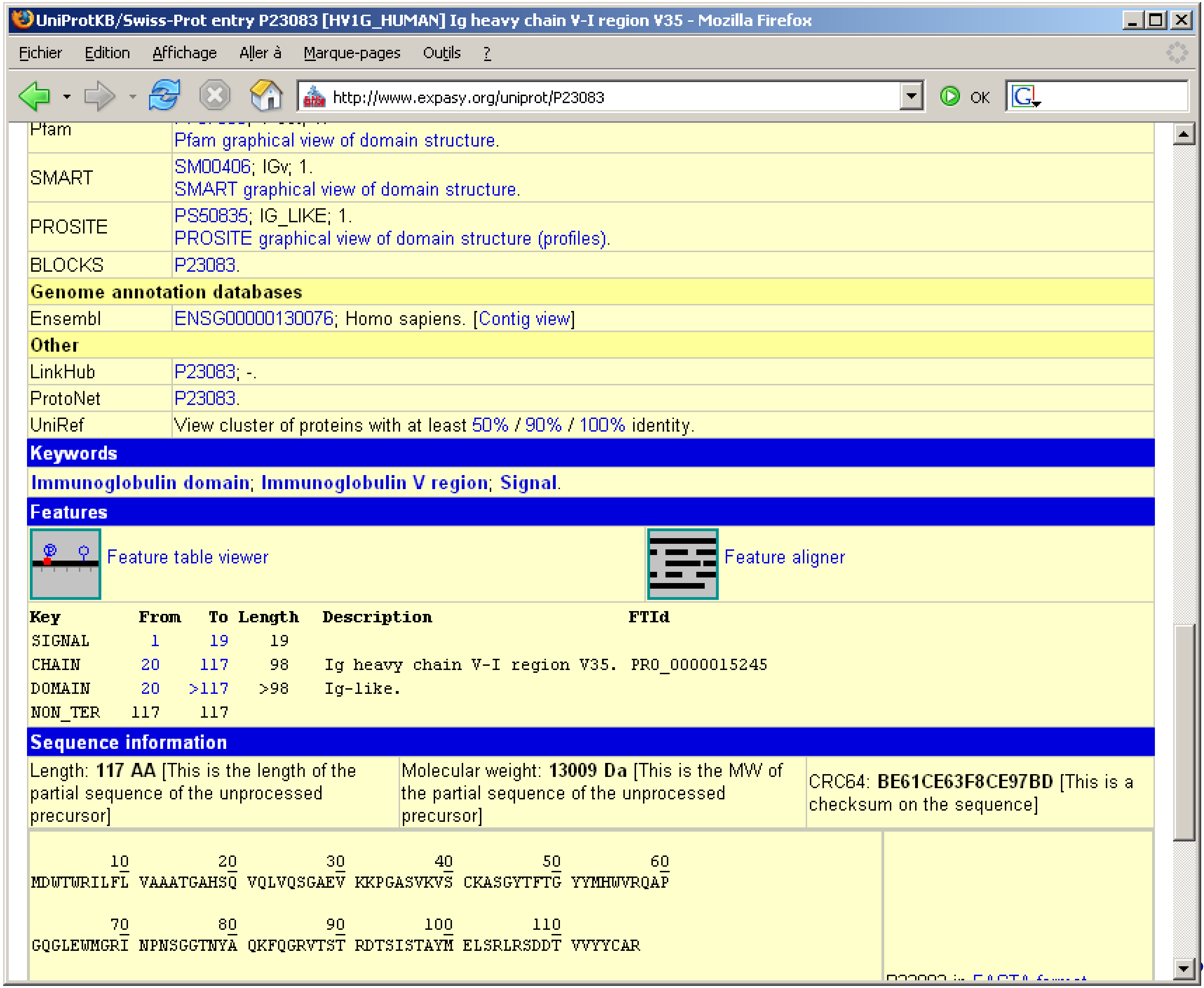
- *SIMILARITY*: Contains 1 Ig-like (immunoglobulin-like) domain.

Copyright

Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms. Distributed under the Creative Commons Attribution-NoDerivs License.

Cross-references

Sequence databases	
EMBL	X07448; -; NOT_ANNOTATED_CDS; Genomic_DNA. [EMBL / GenBank / DDBJ]
PIR	S00476; HVHU35.
3D structure databases	
HSSP	P01751; 1NQB. [HSSP ENTRY / PDB]
SMR	P23083; 20-117.
ModBase	P23083.
Protein-protein interaction databases	
DIP	P23083.
Protein family/group databases	
IMGT	X07448.
2D gel databases	
SWISS-2DPAGE	Get region on 2D PAGE.
Organism-specific gene databases	
HOVERGEN	[Family / Alignment / Tree]
Gene expression databases	
ArrayExpress	P23083; -.
Ontologies	
GO	GO:0005576; Cellular component: extracellular region *(non-traceable author statement)*. GO:0003823; Molecular function: antigen binding *(non-traceable author statement)*. GO:0006955; Biological process: immune response *(non-traceable author statement)*. QuickGo view.
Family and domain databases	
InterPro	IPR007110; Ig-like. IPR013106; Ig_V-set. IPR003596; Ig_V-set_sub. Graphical view of domain structure.



UniProtKB/Swiss-Prot entry P23083 [HV1G_HUMAN] Ig heavy chain V-I region V35 - Mozilla Firefox

http://www.expasy.org/uniprot/P23083

Pfam	Pfam graphical view of domain structure.
SMART	SM00406; IGv; 1. SMART graphical view of domain structure.
PROSITE	PS50835; IG_LIKE; 1. PROSITE graphical view of domain structure (profiles).
BLOCKS	P23083.
Genome annotation databases	
Ensembl	ENSG00000130076; Homo sapiens. [Contig view]
Other	
LinkHub	P23083; -.
ProtoNet	P23083.
UniRef	View cluster of proteins with at least 50% / 90% / 100% identity.

Keywords

Immunoglobulin domain; **Immunoglobulin V region**; **Signal**.

Features

Feature table viewer

Feature aligner

Key	From	To	Length	Description	FTId
SIGNAL	1	19	19		
CHAIN	20	117	98	Ig heavy chain V-I region V35.	PRO_0000015245
DOMAIN	20	>117	>98	Ig-like.	
NON_TER	117	117			

Sequence information

Length: **117 AA** [This is the length of the partial sequence of the unprocessed precursor]	Molecular weight: **13009 Da** [This is the MW of the partial sequence of the unprocessed precursor]	CRC64: **BE61CE63F8CE97BD** [This is a checksum on the sequence]

```
        10         20         30         40         50         60
MDWTWRILFL VAAATGAHSQ VQLVQSGAEV KKPGASVKVS CKASGYTFTG YYMHWVRQAP

        70         80         90        100        110
GQGLEWMGRI NPNSGGTNYA QKFQGRVTST RDTSISTAYM ELSRLRSDDT VVYYCAR
```

PDB PROTEIN DATA BANK
Banque de structures tridimentionnelles

Localisation Créée par Brookhaven National Laboratory,
NIH, Bethesda, MD, USA
Maintenue depuis juillet 1999 par le RCSB
(Research Collaboratory for Structural Bioinformatics)

Contient les coordonnées 3D de 39.223 structures
(protéines ou acides nucléiques) (Octobre 2006)

Outils de visualisation 3D Rasmol, Cn3D, Web Labviewer, WPDB

Distribution (fichiers à plat) Web (http://www.rcsb.org/pdb/Welcome.do), FTP

Systèmes d'information et bases de données spécialisés

Ils couvrent un secteur défini de la biologie

Avantages

- Ils fournissent des informations détaillées, spécifiques du domaine biologique qui n'existent pas dans les systèmes généralistes
- Les données sont en général contrôlées, donc plus fiables et de meilleure qualité que dans les bases généralistes
- Ils évoluent en fonction des progrès scientifiques dans le domaine plus facilement

Systèmes d'information et bases de données spécialisés

Ils couvrent un secteur défini de la biologie

Avantages

- Ils fournissent des informations détaillées, spécifiques du domaine biologique qui n'existent pas dans les systèmes généralistes
- Les données sont en général contrôlées, donc plus fiables et de meilleure qualité que dans les bases généralistes
- Ils évoluent en fonction des progrès scientifiques dans le domaine plus facilement

Le système d'information IMGT

17 Octobre 2006

Im Muno Gene Tics Information system®
http://imgt.cines.fr

IMGT, the international ImMunoGeneTics information system® http://imgt.cines.fr

Bases de données

Bases de données de séquences

IMGT/LIGM-DB
+ de 100.000 séquences nucléotidiques IG et TR de 170 espèces de vertébrés (sur le Web depuis 1995)

- **IMGT/PRIMER-DB**
- **IMGT/MHC-DB**

Base de données de gènes

- **IMGT/GENE-DB**

Base de données de structures 3D

- **IMGT/3Dstructure-DB**

Outils en ligne

Analyse de séquences

- **IMGT/V-QUEST**
- **IMGT/JunctionAnalysis**
- **IMGT/PhyloGene**
- **IMGT/Allele-Align**

Analyse de génomes

- **IMGT/GeneSearch**
- **IMGT/GeneView**
- **IMGT/LocusView**

Analyse de structures 3D

- **IMGT/StructuralQuery**

Ressources Web

Séquences

IMGT Repertoire:

- **Alignments of alleles**
- **Tables of alleles**
- **Protein displays, etc.**

Génomes

- **Chromosomal localizations**
- **Locus representations**
- **Gene tables, etc.**

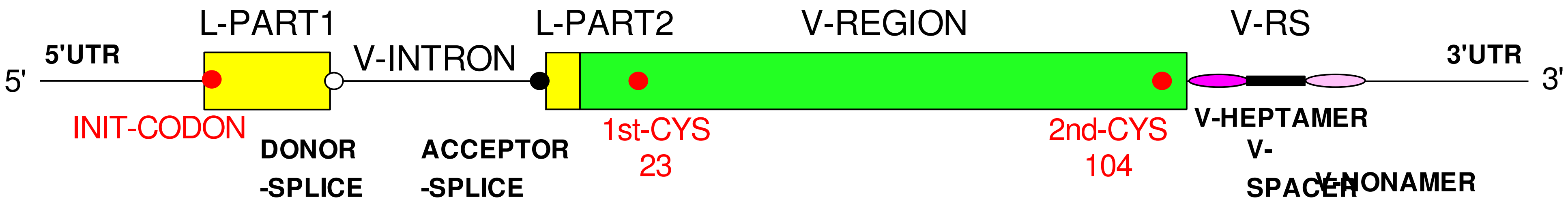
Structures 2D et 3D

- **2D Colliers de Perles**
- **3D representations, etc.**

Séquences d’ADN génomique en configuration germline

1 - Exemple d’un V-GENE

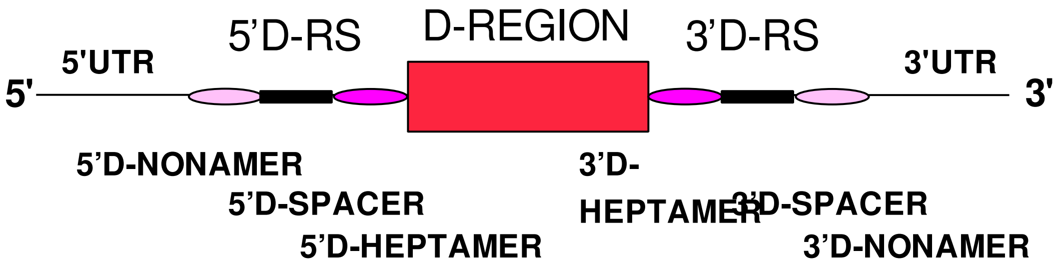
```
>X62106.0|HSVI2|Homo sapiens VI-2 gene for immunoglobulin heavy chain
tgagagctcc gttcctcacc atggactgga cctggaggat cctcttcttg gtggcagcag       60
ccacaggtaa gaggctccct agtcccagtg atgagaaaga gattgagtcc agtccaggga      120
gatctcatcc acttctgtgt tctctccaca ggagcccact cccaggtgca gctggtgcag      180
tctggggctg aggtgaagaa gcctggggcc tcagtgaagg tctcctgcaa ggcttctgga      240
tacaccttca ccggctacta tatgcactgg gtgcgacagg cccctggaca agggcttgag      300
tggatgggat ggatcaaccc taacagtggt ggcacaaact atgcacagaa gtttcagggc      360
agggtcacca tgaccaggga cacgtccatc agcacagcct acatg                g      420
                                                      tgaaa
agatctgacg acacggccgt gtattactgt gcgagagaca cagtg                c      480
                                                      acccacatcc
tgagggtgtc agaaacccaa gggaggaggc ag
```



Séquences d’ADN génomique en configuration germline

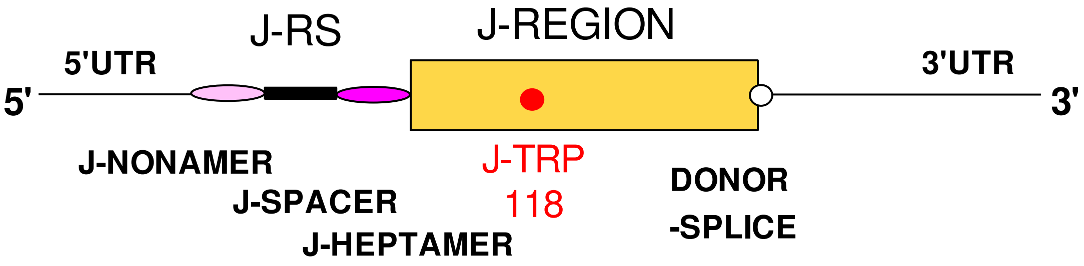
2 - Exemple d’un D-GENE

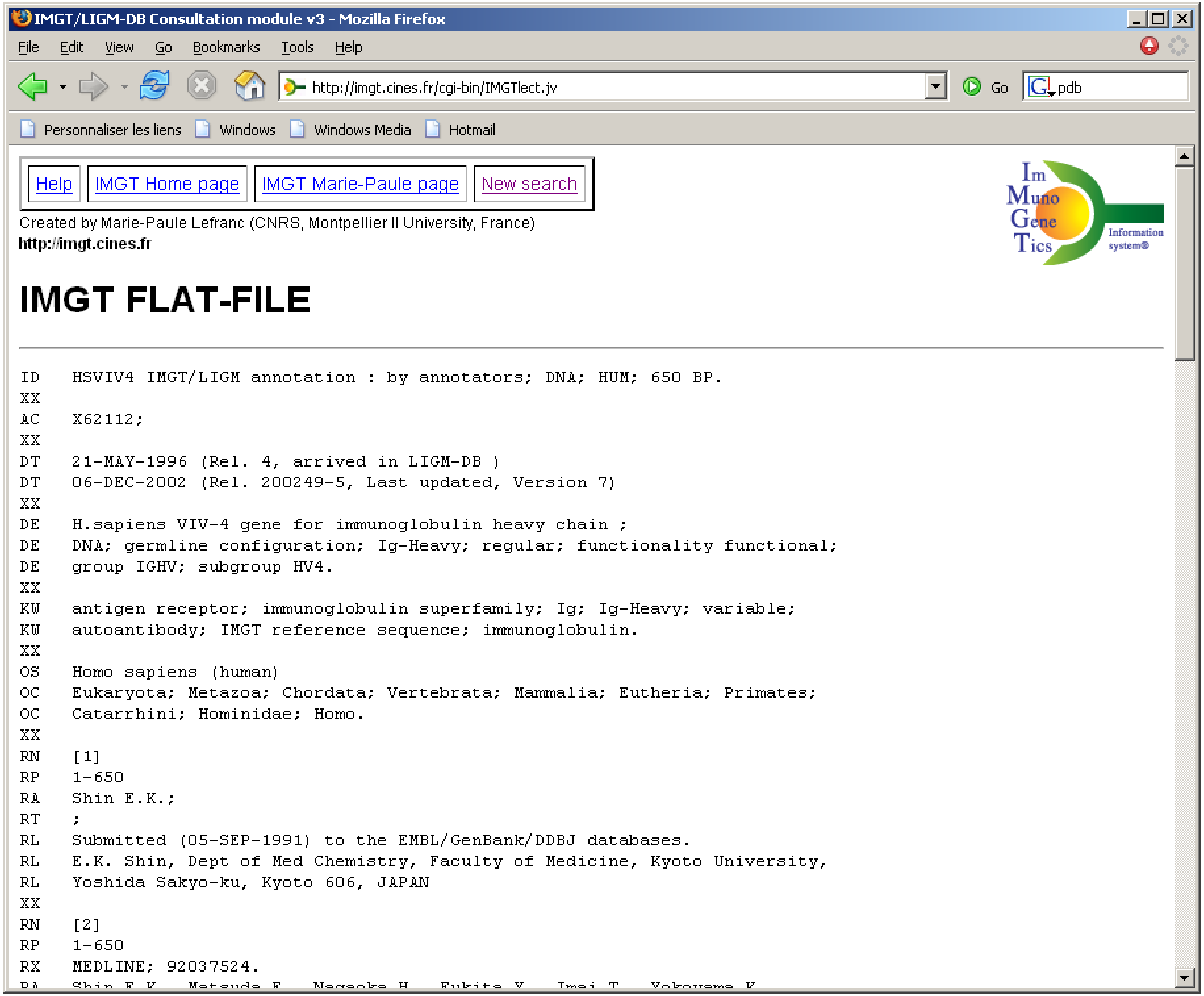
ggtttttggc tgagctgaga accactgtgc taactgggga cacagtcatt
ggcagctcta caaaaacc



3 - Exemple d’un J-GENE

ggtttct gtgcccctgg ctcagggctg actcaccgtg gctgaatact
tccagcactg gggccagggc accctggtca ccgtctcctc ag

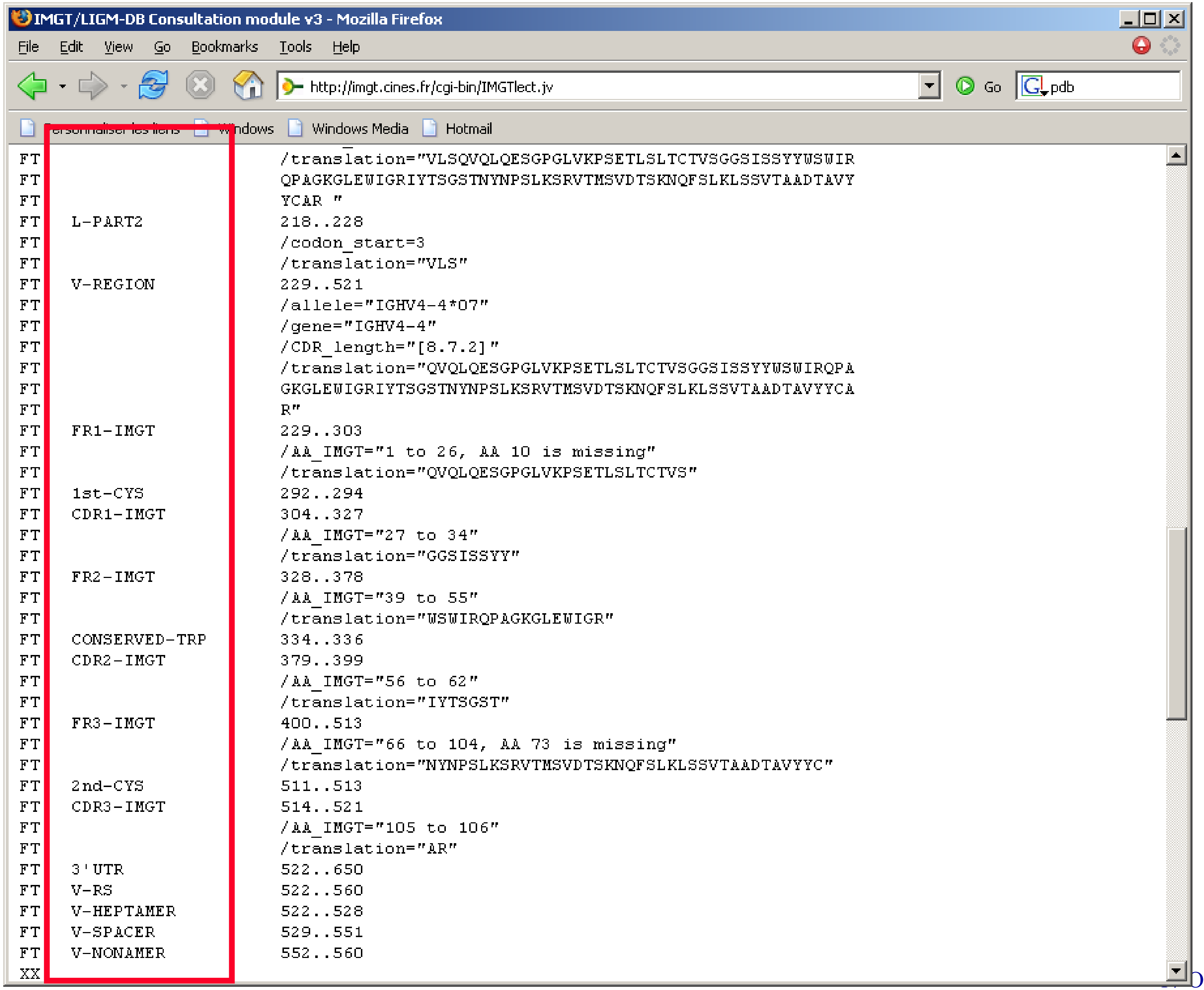




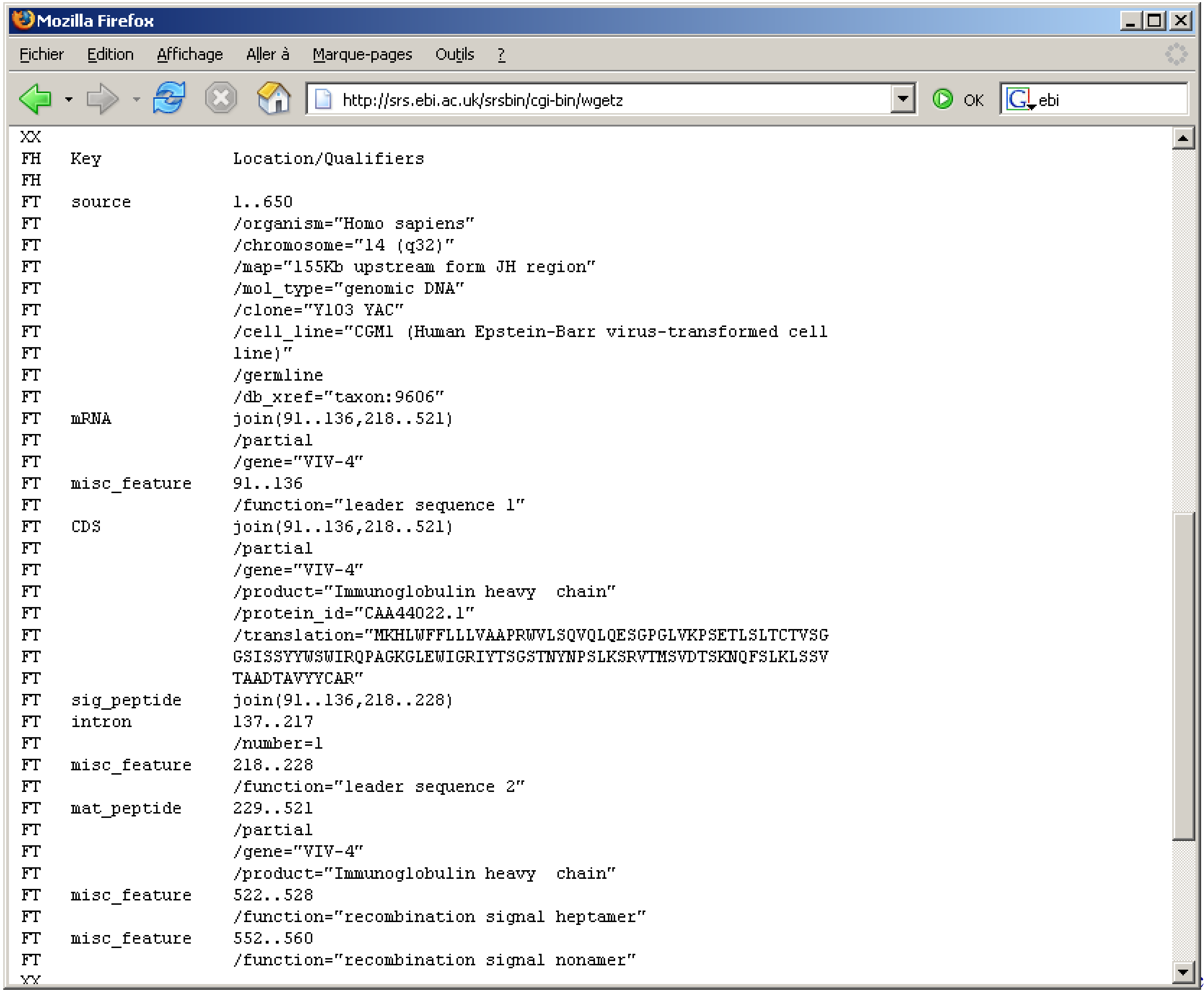
Created by Marie-Paule Lefranc (CNRS, Montpellier II University, France)
http://imgt.cines.fr

IMGT FLAT-FILE

```
ID   HSVIV4 IMGT/LIGM annotation : by annotators; DNA; HUM; 650 BP.
XX
AC   X62112;
XX
DT   21-MAY-1996 (Rel. 4, arrived in LIGM-DB )
DT   06-DEC-2002 (Rel. 200249-5, Last updated, Version 7)
XX
DE   H.sapiens VIV-4 gene for immunoglobulin heavy chain ;
DE   DNA; germline configuration; Ig-Heavy; regular; functionality functional;
DE   group IGHV; subgroup HV4.
XX
KW   antigen receptor; immunoglobulin superfamily; Ig; Ig-Heavy; variable;
KW   autoantibody; IMGT reference sequence; immunoglobulin.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates;
OC   Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-650
RA   Shin E.K.;
RT   ;
RL   Submitted (05-SEP-1991) to the EMBL/GenBank/DDBJ databases.
RL   E.K. Shin, Dept of Med Chemistry, Faculty of Medicine, Kyoto University,
RL   Yoshida Sakyo-ku, Kyoto 606, JAPAN
XX
RN   [2]
RP   1-650
RX   MEDLINE; 92037524.
RA   Shin E.K., Matsuda F., Nagaoka H., Fukita Y., Imai T., Yokoyama K.,
```



```
FT                                /translation="VLSQVQLQESGPGLVKPSETLSLTCTVSGGSISSYYWSWIR
FT                                QPAGKGLEWIGRIYTSGSTNYNPSLKSRVTMSVDTSKNQFSLKLSSVTAADTAVY
FT                                YCAR "
FT   L-PART2                      218..228
FT                                /codon_start=3
FT                                /translation="VLS"
FT   V-REGION                     229..521
FT                                /allele="IGHV4-4*07"
FT                                /gene="IGHV4-4"
FT                                /CDR_length="[8.7.2]"
FT                                /translation="QVQLQESGPGLVKPSETLSLTCTVSGGSISSYYWSWIRQPA
FT                                GKGLEWIGRIYTSGSTNYNPSLKSRVTMSVDTSKNQFSLKLSSVTAADTAVYYCA
FT                                R"
FT   FR1-IMGT                     229..303
FT                                /AA_IMGT="1 to 26, AA 10 is missing"
FT                                /translation="QVQLQESGPGLVKPSETLSLTCTVS"
FT   1st-CYS                      292..294
FT   CDR1-IMGT                    304..327
FT                                /AA_IMGT="27 to 34"
FT                                /translation="GGSISSYY"
FT   FR2-IMGT                     328..378
FT                                /AA_IMGT="39 to 55"
FT                                /translation="WSWIRQPAGKGLEWIGR"
FT   CONSERVED-TRP                334..336
FT   CDR2-IMGT                    379..399
FT                                /AA_IMGT="56 to 62"
FT                                /translation="IYTSGST"
FT   FR3-IMGT                     400..513
FT                                /AA_IMGT="66 to 104, AA 73 is missing"
FT                                /translation="NYNPSLKSRVTMSVDTSKNQFSLKLSSVTAADTAVYYC"
FT   2nd-CYS                      511..513
FT   CDR3-IMGT                    514..521
FT                                /AA_IMGT="105 to 106"
FT                                /translation="AR"
FT   3'UTR                        522..650
FT   V-RS                         522..560
FT   V-HEPTAMER                   522..528
FT   V-SPACER                     529..551
FT   V-NONAMER                    552..560
XX
```



```
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..650
FT                   /organism="Homo sapiens"
FT                   /chromosome="14 (q32)"
FT                   /map="155Kb upstream form JH region"
FT                   /mol_type="genomic DNA"
FT                   /clone="Y103 YAC"
FT                   /cell_line="CGM1 (Human Epstein-Barr virus-transformed cell
FT                   line)"
FT                   /germline
FT                   /db_xref="taxon:9606"
FT   mRNA            join(91..136,218..521)
FT                   /partial
FT                   /gene="VIV-4"
FT   misc_feature    91..136
FT                   /function="leader sequence 1"
FT   CDS             join(91..136,218..521)
FT                   /partial
FT                   /gene="VIV-4"
FT                   /product="Immunoglobulin heavy  chain"
FT                   /protein_id="CAA44022.1"
FT                   /translation="MKHLWFFLLLVAAPRWVLSQVQLQESGPGLVKPSETLSLTCTVSG
FT                   GSISSYYWSWIRQPAGKGLEWIGRIYTSGSTNYNPSLKSRVTMSVDTSKNQFSLKLSSV
FT                   TAADTAVYYCAR"
FT   sig_peptide     join(91..136,218..228)
FT   intron          137..217
FT                   /number=1
FT   misc_feature    218..228
FT                   /function="leader sequence 2"
FT   mat_peptide     229..521
FT                   /partial
FT                   /gene="VIV-4"
FT                   /product="Immunoglobulin heavy  chain"
FT   misc_feature    522..528
FT                   /function="recombination signal heptamer"
FT   misc_feature    552..560
FT                   /function="recombination signal nonamer"
XX
```

Sequence Database Retrieval

1- SRS

(sequence retrieval system)

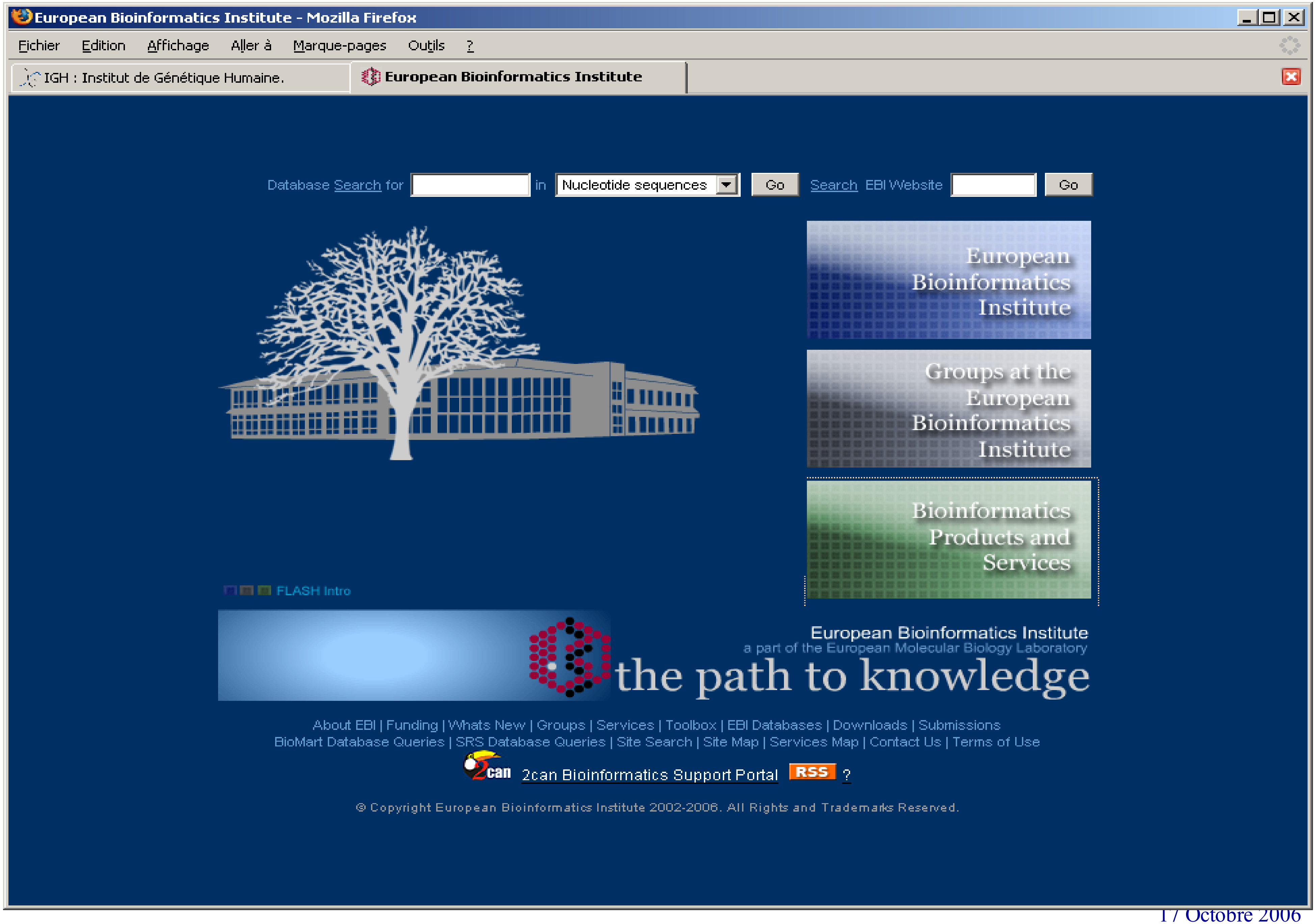
2- ENTREZ

caractéristiques

- Système d'interrogation multibase

- Accès à des bases de données dont les objectifs peuvent être très différents (BD generalistes, specialiées, ontologies, references bibliographiques)

- interfaces puissantes pour répondre à des requêtes complexes



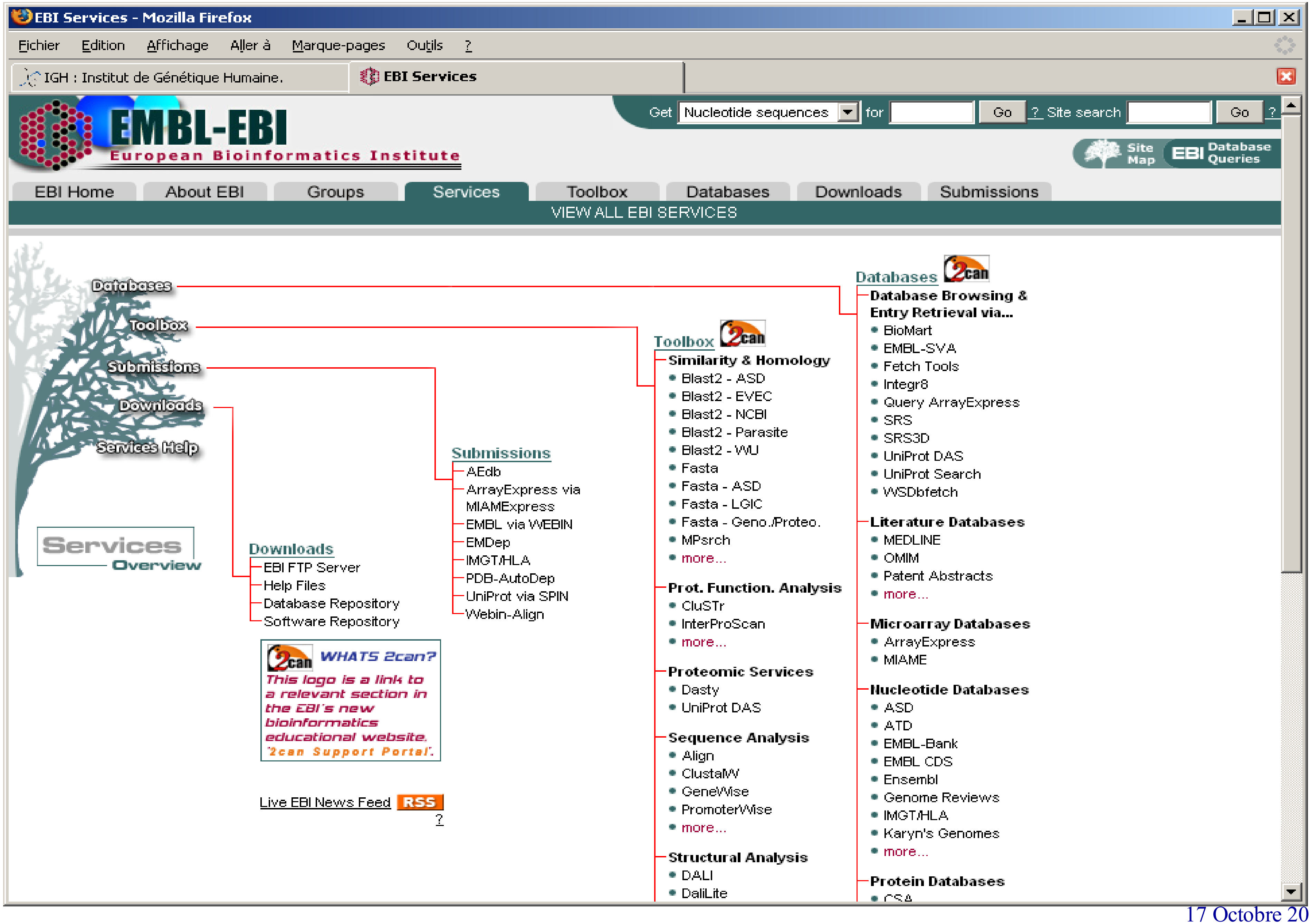
European Bioinformatics Institute

Groups at the European Bioinformatics Institute

Bioinformatics Products and Services

European Bioinformatics Institute
a part of the European Molecular Biology Laboratory
the path to knowledge

© Copyright European Bioinformatics Institute 2002-2006. All Rights and Trademarks Reserved.



Submissions

- AEdb
- ArrayExpress via MIAMExpress
- EMBL via WEBIN
- EMDep
- IMGT/HLA
- PDB-AutoDep
- UniProt via SPIN
- Webin-Align

Downloads

- EBI FTP Server
- Help Files
- Database Repository
- Software Repository

WHATS 2can?

This logo is a link to a relevant section in the EBI's new bioinformatics educational website. '2can Support Portal'.

Live EBI News Feed RSS
?

Toolbox 2can

Similarity & Homology

- Blast2 - ASD
- Blast2 - EVEC
- Blast2 - NCBI
- Blast2 - Parasite
- Blast2 - WU
- Fasta
- Fasta - ASD
- Fasta - LGIC
- Fasta - Geno./Proteo.
- MPsrch
- more...

Prot. Function. Analysis

- CluSTr
- InterProScan
- more...

Proteomic Services

- Dasty
- UniProt DAS

Sequence Analysis

- Align
- ClustalW
- GeneWise
- PromoterWise
- more...

Structural Analysis

- DALI
- DaliLite

Databases 2can

Database Browsing & Entry Retrieval via...

- BioMart
- EMBL-SVA
- Fetch Tools
- Integr8
- Query ArrayExpress
- SRS
- SRS3D
- UniProt DAS
- UniProt Search
- WSDbfetch

Literature Databases

- MEDLINE
- OMIM
- Patent Abstracts
- more...

Microarray Databases

- ArrayExpress
- MIAME

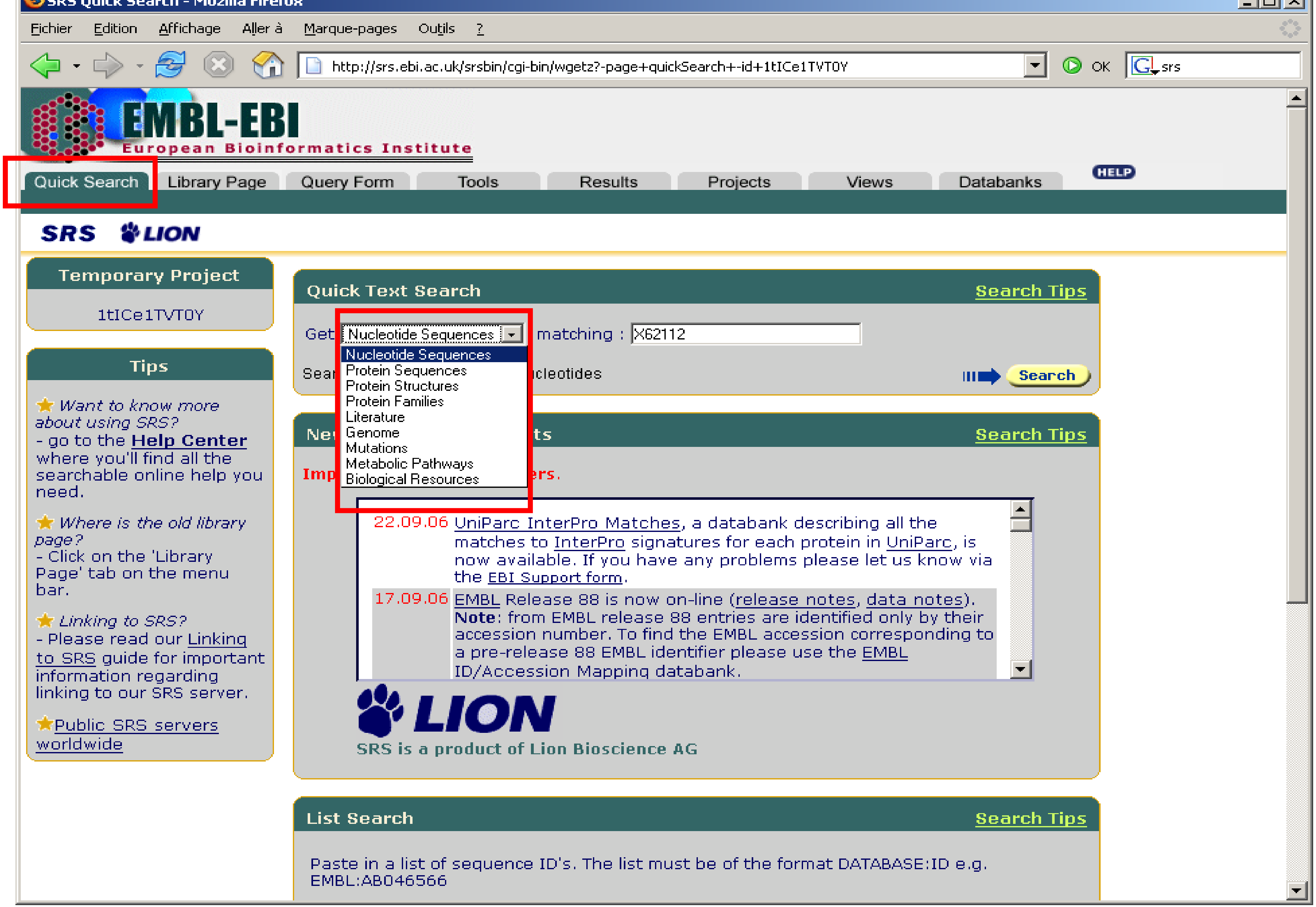
Nucleotide Databases

- ASD
- ATD
- EMBL-Bank
- EMBL CDS
- Ensembl
- Genome Reviews
- IMGT/HLA
- Karyn's Genomes
- more...

Protein Databases

SRS

http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page+srsq2+-noSession



Temporary Project

1tICe1TVT0Y

Tips

Want to know more about using SRS?
- go to the **Help Center** where you'll find all the searchable online help you need.

Where is the old library page?
- Click on the 'Library Page' tab on the menu bar.

Linking to SRS?
- Please read our Linking to SRS guide for important information regarding linking to our SRS server.

Public SRS servers worldwide

Quick Text Search

Get Nucleotide Sequences matching : X62112

Sear ... cleotides

Ne ... ts

Imp ... ers.

22.09.06 UniParc InterPro Matches, a databank describing all the matches to InterPro signatures for each protein in UniParc, is now available. If you have any problems please let us know via the EBI Support form.

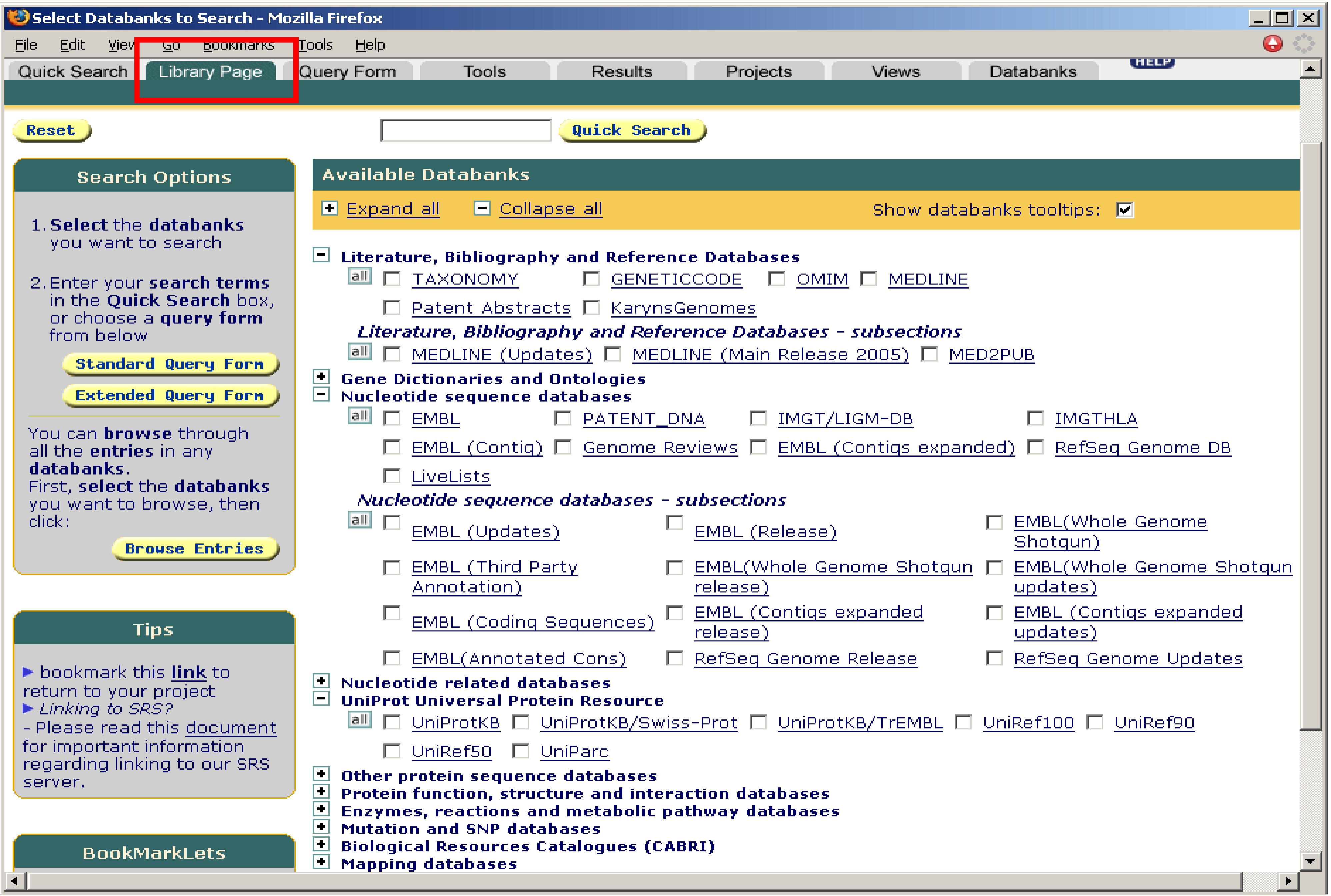
17.09.06 EMBL Release 88 is now on-line (release notes, data notes). **Note**: from EMBL release 88 entries are identified only by their accession number. To find the EMBL accession corresponding to a pre-release 88 EMBL identifier please use the EMBL ID/Accession Mapping databank.

SRS is a product of Lion Bioscience AG

List Search

Paste in a list of sequence ID's. The list must be of the format DATABASE:ID e.g. EMBL:AB046566

Selection des bases de données



Search Options

1. **Select** the **databanks** you want to search
2. Enter your **search terms** in the **Quick Search** box, or choose a **query form** from below

You can **browse** through all the **entries** in any **databanks**.
First, **select** the **databanks** you want to browse, then click:

Tips

- bookmark this link to return to your project
- *Linking to SRS?*

- Please read this document for important information regarding linking to our SRS server.

BookMarkLets

Available Databanks

⊞ Expand all ⊟ Collapse all Show databanks tooltips: ☑

⊟ **Literature, Bibliography and Reference Databases**

all ☐ TAXONOMY ☐ GENETICCODE ☐ OMIM ☐ MEDLINE
☐ Patent Abstracts ☐ KarynsGenomes

Literature, Bibliography and Reference Databases - subsections

all ☐ MEDLINE (Updates) ☐ MEDLINE (Main Release 2005) ☐ MED2PUB

⊞ **Gene Dictionaries and Ontologies**

⊟ **Nucleotide sequence databases**

all ☐ EMBL ☐ PATENT_DNA ☐ IMGT/LIGM-DB ☐ IMGTHLA
☐ EMBL (Contig) ☐ Genome Reviews ☐ EMBL (Contigs expanded) ☐ RefSeq Genome DB
☐ LiveLists

Nucleotide sequence databases - subsections

all ☐ EMBL (Updates) ☐ EMBL (Release) ☐ EMBL(Whole Genome Shotgun)
☐ EMBL (Third Party Annotation) ☐ EMBL(Whole Genome Shotgun release) ☐ EMBL(Whole Genome Shotgun updates)
☐ EMBL (Coding Sequences) ☐ EMBL (Contigs expanded release) ☐ EMBL (Contigs expanded updates)
☐ EMBL(Annotated Cons) ☐ RefSeq Genome Release ☐ RefSeq Genome Updates

⊞ **Nucleotide related databases**

⊟ **UniProt Universal Protein Resource**

all ☐ UniProtKB ☐ UniProtKB/Swiss-Prot ☐ UniProtKB/TrEMBL ☐ UniRef100 ☐ UniRef90
☐ UniRef50 ☐ UniParc

⊞ **Other protein sequence databases**

⊞ **Protein function, structure and interaction databases**

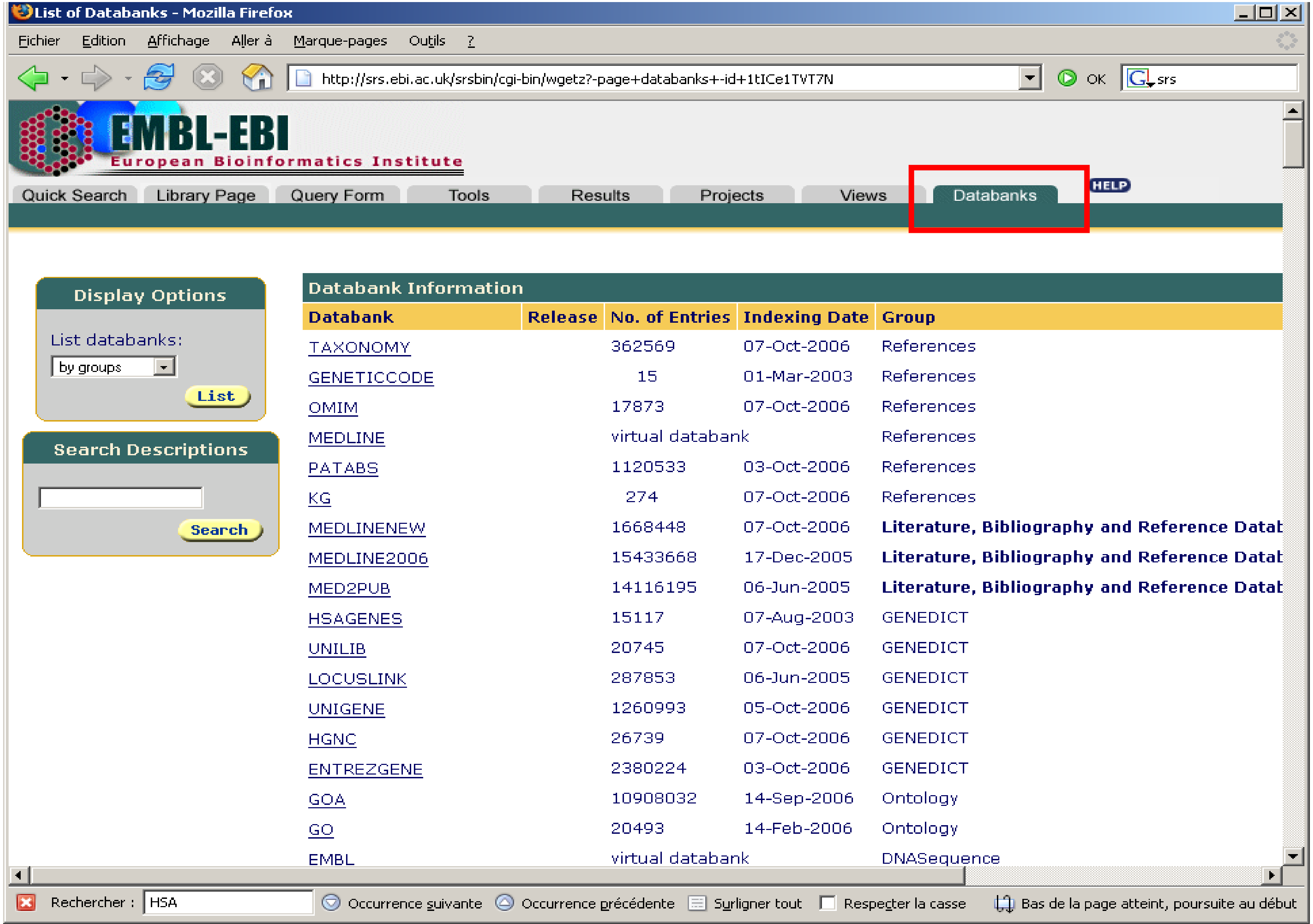
⊞ **Enzymes, reactions and metabolic pathway databases**

⊞ **Mutation and SNP databases**

⊞ **Biological Resources Catalogues (CABRI)**

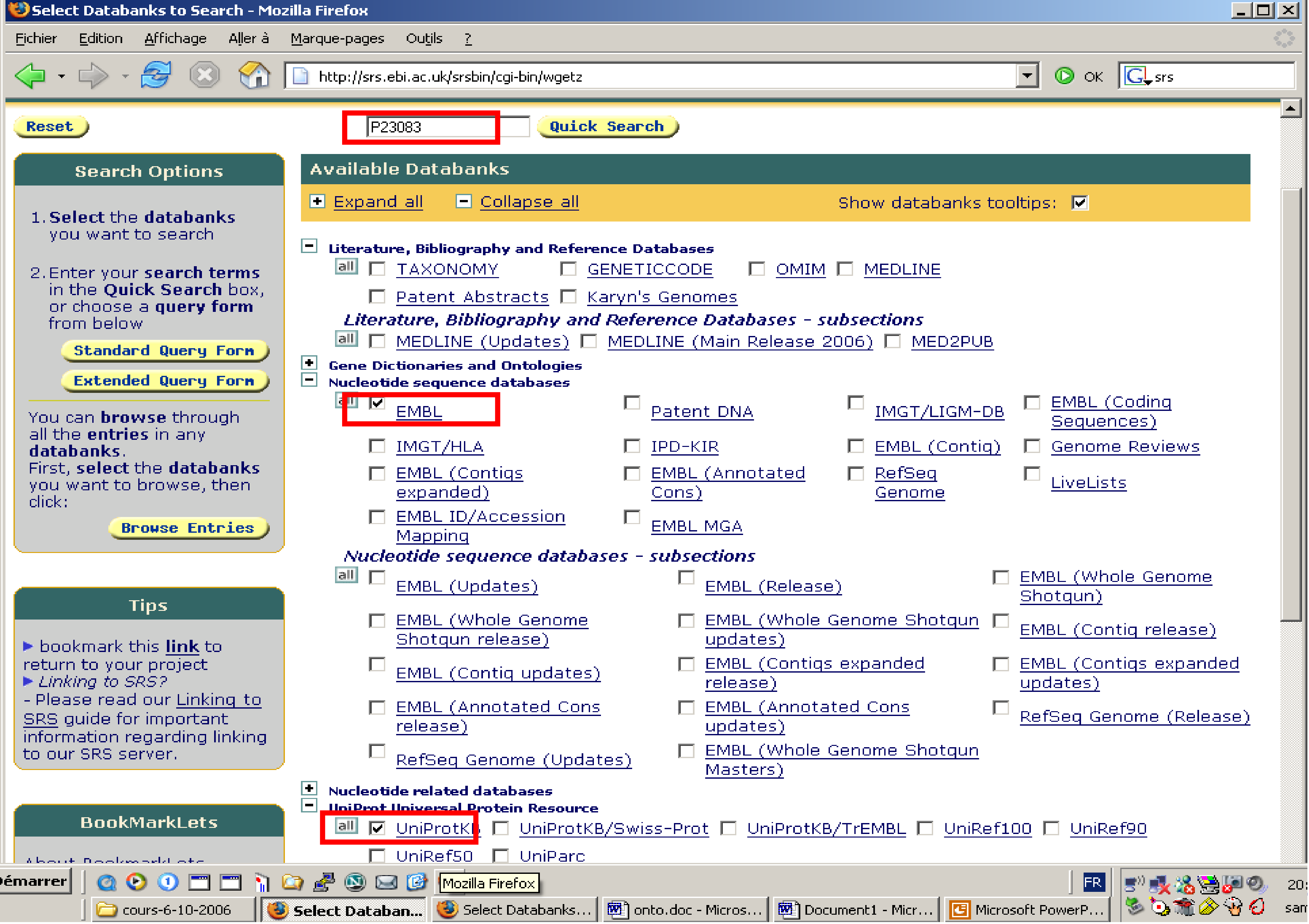
⊞ **Mapping databases**

Pour en savoir plus sur les bases de données diponibles



Databank Information

Databank	Release	No. of Entries	Indexing Date	Group
TAXONOMY		362569	07-Oct-2006	References
GENETICCODE		15	01-Mar-2003	References
OMIM		17873	07-Oct-2006	References
MEDLINE		virtual databank		References
PATABS		1120533	03-Oct-2006	References
KG		274	07-Oct-2006	References
MEDLINENEW		1668448	07-Oct-2006	**Literature, Bibliography and Reference Data**
MEDLINE2006		15433668	17-Dec-2005	**Literature, Bibliography and Reference Data**
MED2PUB		14116195	06-Jun-2005	**Literature, Bibliography and Reference Data**
HSAGENES		15117	07-Aug-2003	GENEDICT
UNILIB		20745	07-Oct-2006	GENEDICT
LOCUSLINK		287853	06-Jun-2005	GENEDICT
UNIGENE		1260993	05-Oct-2006	GENEDICT
HGNC		26739	07-Oct-2006	GENEDICT
ENTREZGENE		2380224	03-Oct-2006	GENEDICT
GOA		10908032	14-Sep-2006	Ontology
GO		20493	14-Feb-2006	Ontology
EMBL		virtual databank		DNASequence



P23083

Search Options

1. **Select** the **databanks** you want to search
2. Enter your **search terms** in the **Quick Search** box, or choose a **query form** from below

You can **browse** through all the **entries** in any **databanks**.
First, **select** the **databanks** you want to browse, then click:

Tips

- bookmark this **link** to return to your project
- *Linking to SRS?*

- Please read our Linking to SRS guide for important information regarding linking to our SRS server.

BookMarkLets

Available Databanks

Show databanks tooltips: ☑

Literature, Bibliography and Reference Databases

☐ TAXONOMY ☐ GENETICCODE ☐ OMIM ☐ MEDLINE
☐ Patent Abstracts ☐ Karyn's Genomes

Literature, Bibliography and Reference Databases - subsections

☐ MEDLINE (Updates) ☐ MEDLINE (Main Release 2006) ☐ MED2PUB

Gene Dictionaries and Ontologies

Nucleotide sequence databases

☑ EMBL ☐ Patent DNA ☐ IMGT/LIGM-DB ☐ EMBL (Coding Sequences)
☐ IMGT/HLA ☐ IPD-KIR ☐ EMBL (Contig) ☐ Genome Reviews
☐ EMBL (Contigs expanded) ☐ EMBL (Annotated Cons) ☐ RefSeq Genome ☐ LiveLists
☐ EMBL ID/Accession Mapping ☐ EMBL MGA

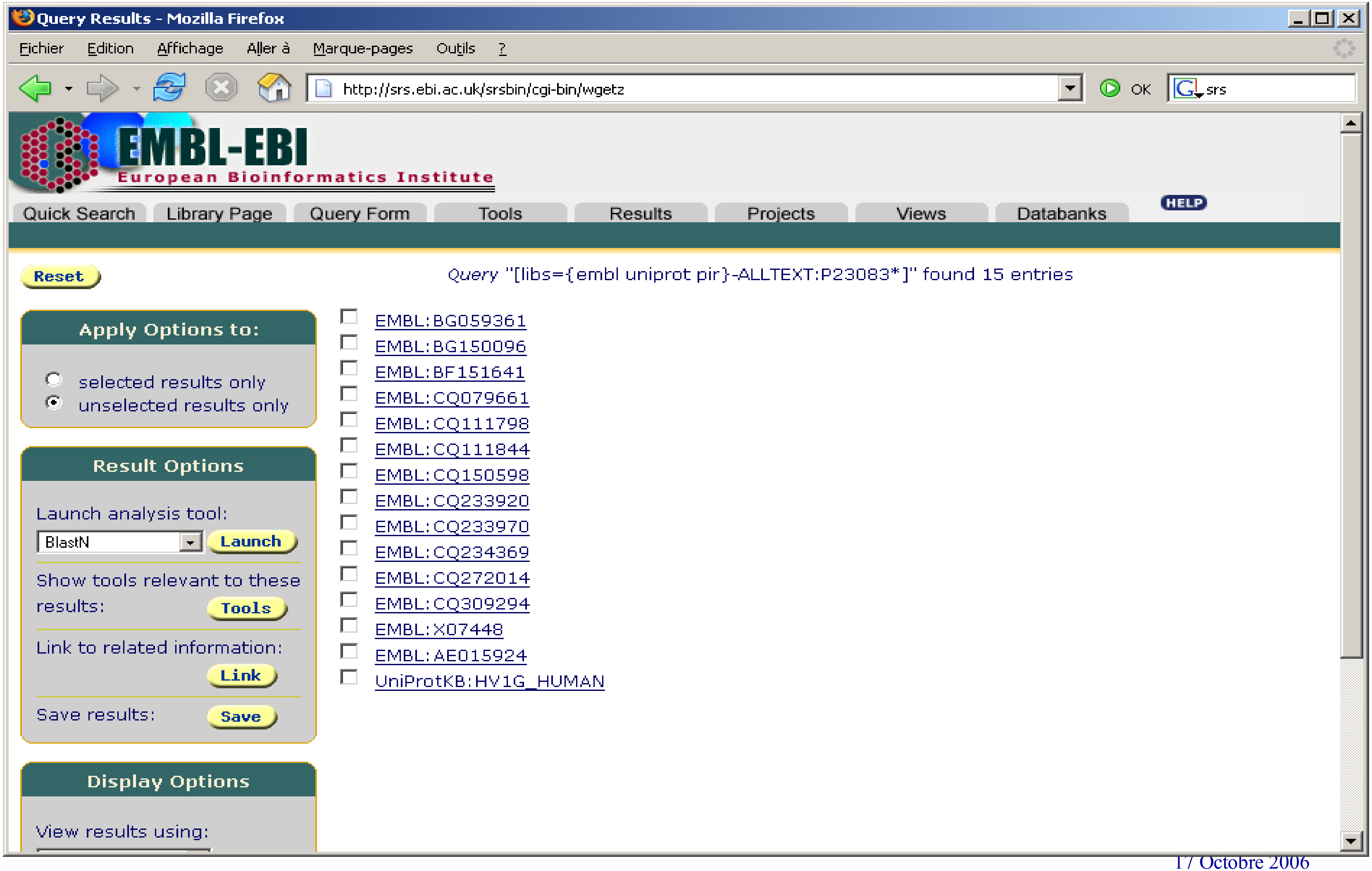
Nucleotide sequence databases - subsections

☐ EMBL (Updates) ☐ EMBL (Release) ☐ EMBL (Whole Genome Shotgun)
☐ EMBL (Whole Genome Shotgun release) ☐ EMBL (Whole Genome Shotgun updates) ☐ EMBL (Contig release)
☐ EMBL (Contig updates) ☐ EMBL (Contigs expanded release) ☐ EMBL (Contigs expanded updates)
☐ EMBL (Annotated Cons release) ☐ EMBL (Annotated Cons updates) ☐ RefSeq Genome (Release)
☐ RefSeq Genome (Updates) ☐ EMBL (Whole Genome Shotgun Masters)

Nucleotide related databases

UniProt Universal Protein Resource

☑ UniProtKB ☐ UniProtKB/Swiss-Prot ☐ UniProtKB/TrEMBL ☐ UniRef100 ☐ UniRef90
☐ UniRef50 ☐ UniParc



http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz

Query "[libs={embl uniprot pir}-ALLTEXT:P23083*]" found 15 entries

Apply Options to:

○ selected results only
● unselected results only

Result Options

Launch analysis tool: BlastN

Show tools relevant to these results:

Link to related information:

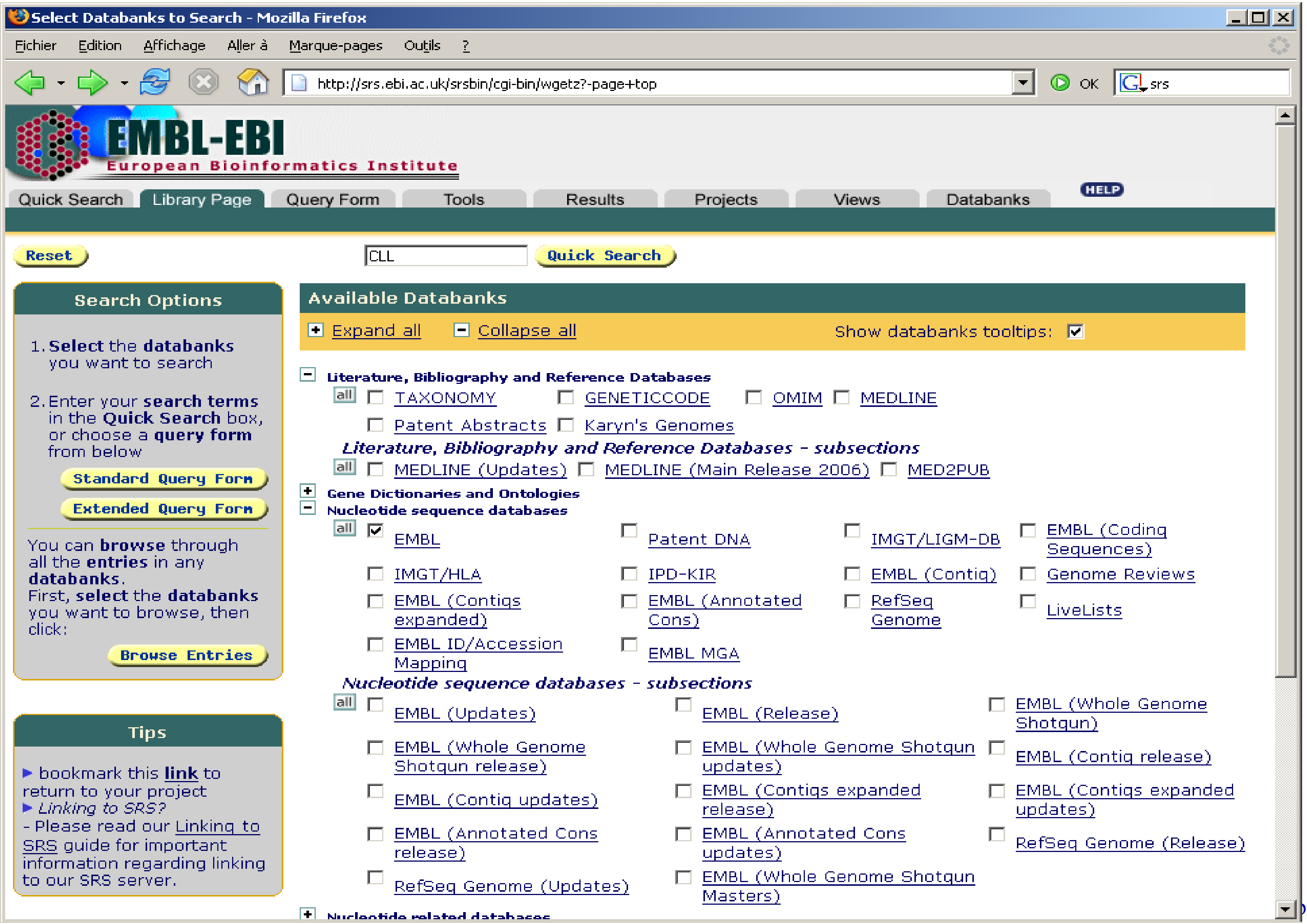
Save results:

Display Options

View results using:

- ☐ EMBL:BG059361
- ☐ EMBL:BG150096
- ☐ EMBL:BF151641
- ☐ EMBL:CQ079661
- ☐ EMBL:CQ111798
- ☐ EMBL:CQ111844
- ☐ EMBL:CQ150598
- ☐ EMBL:CQ233920
- ☐ EMBL:CQ233970
- ☐ EMBL:CQ234369
- ☐ EMBL:CQ272014
- ☐ EMBL:CQ309294
- ☐ EMBL:X07448
- ☐ EMBL:AE015924
- ☐ UniProtKB:HV1G_HUMAN

1 - Recherche rapide: par mots clés
chronic lymphocytic leukemia et/ou CLL



CLL

Search Options

1. **Select** the **databanks** you want to search
2. Enter your **search terms** in the **Quick Search** box, or choose a **query form** from below

You can **browse** through all the **entries** in any **databanks**.
First, **select** the **databanks** you want to browse, then click:

Tips

- bookmark this **link** to return to your project
- *Linking to SRS?*
- Please read our Linking to SRS guide for important information regarding linking to our SRS server.

Available Databanks

⊞ Expand all ⊟ Collapse all Show databanks tooltips: ☑

⊟ **Literature, Bibliography and Reference Databases**

all ☐ TAXONOMY ☐ GENETICCODE ☐ OMIM ☐ MEDLINE
☐ Patent Abstracts ☐ Karyn's Genomes

Literature, Bibliography and Reference Databases - subsections

all ☐ MEDLINE (Updates) ☐ MEDLINE (Main Release 2006) ☐ MED2PUB

⊞ **Gene Dictionaries and Ontologies**

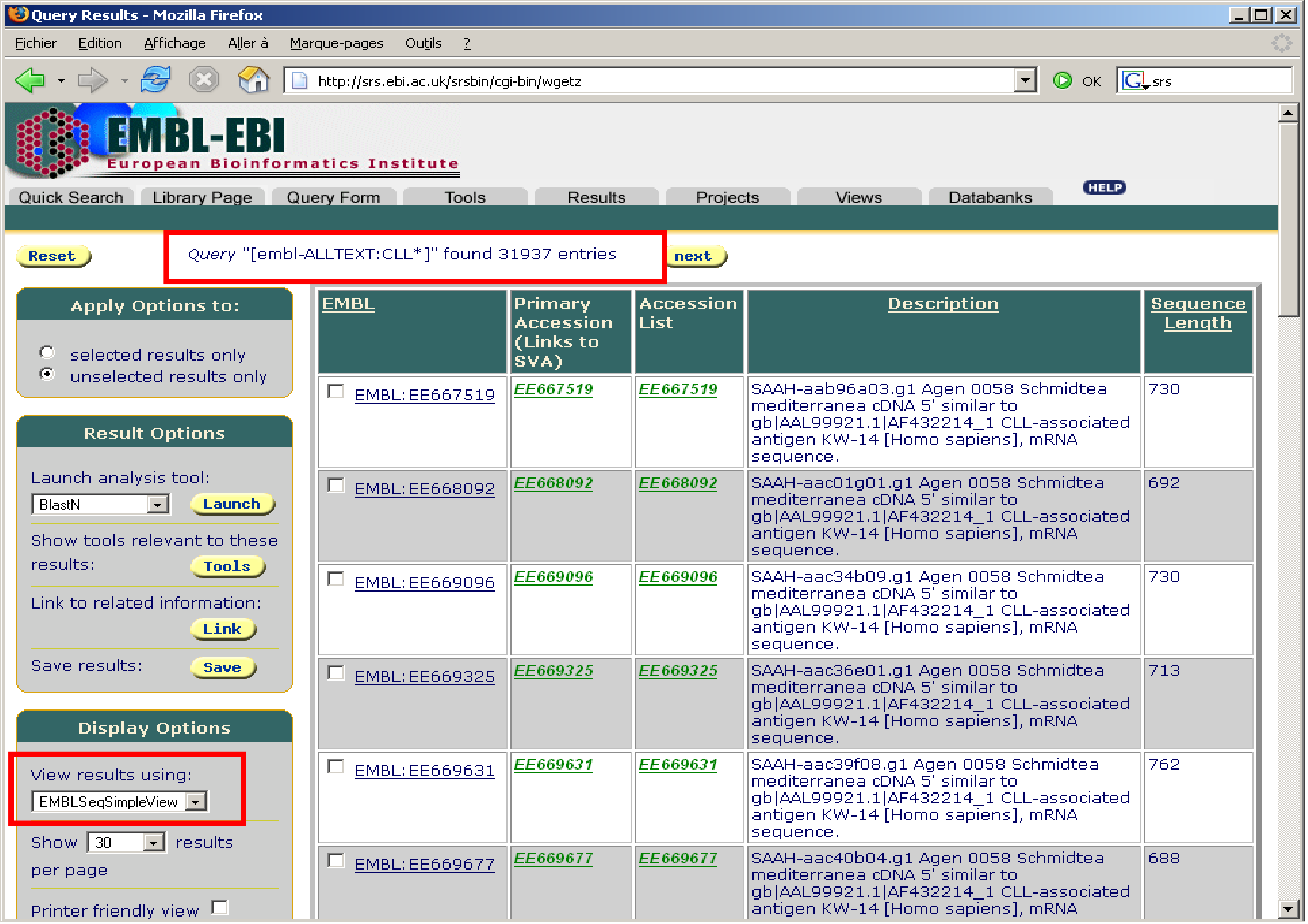
⊟ **Nucleotide sequence databases**

all ☑ EMBL ☐ Patent DNA ☐ IMGT/LIGM-DB ☐ EMBL (Coding Sequences)
☐ IMGT/HLA ☐ IPD-KIR ☐ EMBL (Contig) ☐ Genome Reviews
☐ EMBL (Contigs expanded) ☐ EMBL (Annotated Cons) ☐ RefSeq Genome ☐ LiveLists
☐ EMBL ID/Accession Mapping ☐ EMBL MGA

Nucleotide sequence databases - subsections

all ☐ EMBL (Updates) ☐ EMBL (Release) ☐ EMBL (Whole Genome Shotgun)
☐ EMBL (Whole Genome Shotgun release) ☐ EMBL (Whole Genome Shotgun updates) ☐ EMBL (Contig release)
☐ EMBL (Contig updates) ☐ EMBL (Contigs expanded release) ☐ EMBL (Contigs expanded updates)
☐ EMBL (Annotated Cons release) ☐ EMBL (Annotated Cons updates) ☐ RefSeq Genome (Release)
☐ RefSeq Genome (Updates) ☐ EMBL (Whole Genome Shotgun Masters)

⊞ **Nucleotide related databases**



Query "[embl-ALLTEXT:CLL*]" found 31937 entries

Apply Options to:

○ selected results only
● unselected results only

Result Options

Launch analysis tool: BlastN

Show tools relevant to these results:

Link to related information:

Save results:

Display Options

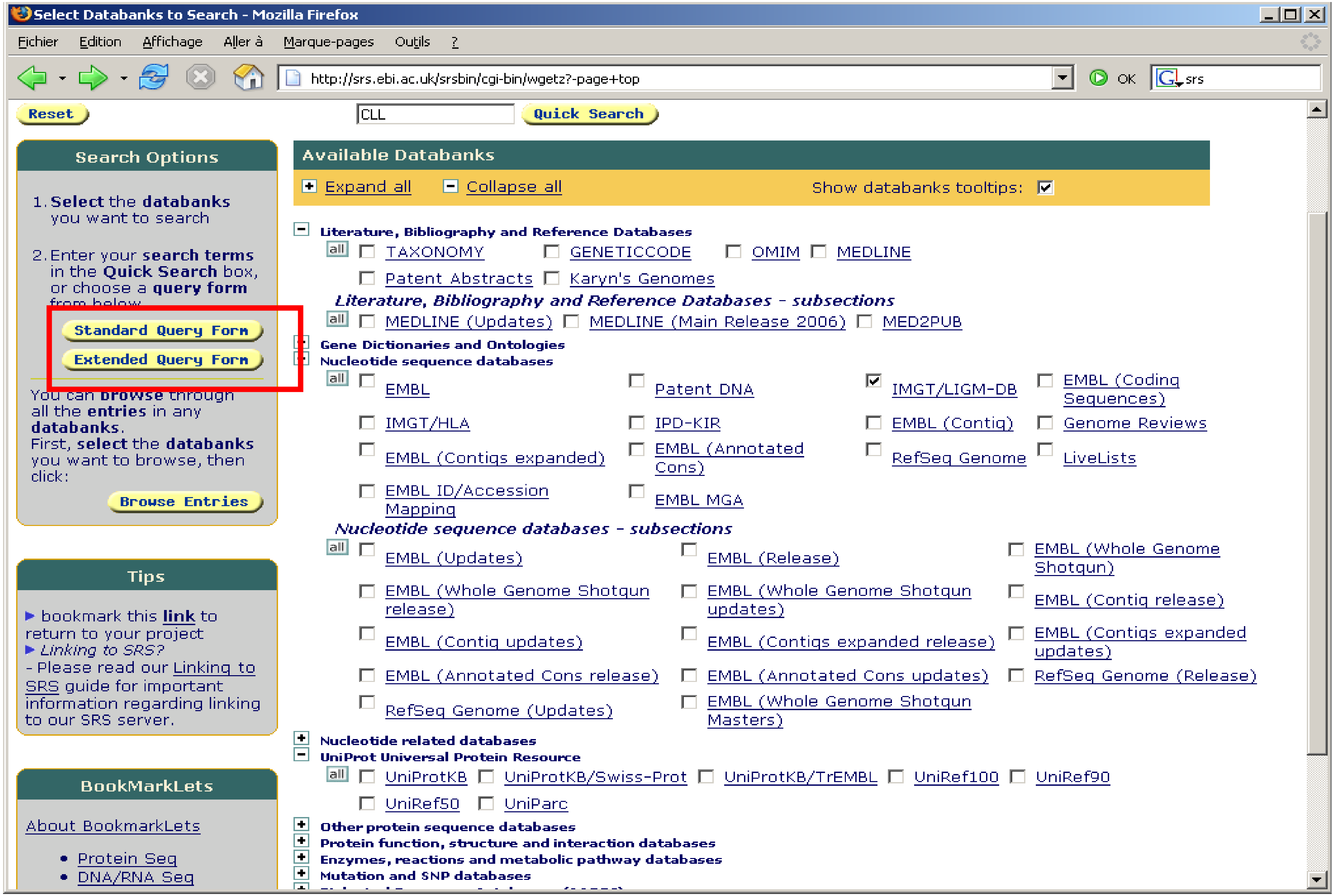
View results using: EMBLSeqSimpleView

Show 30 results per page

Printer friendly view ☐

EMBL	Primary Accession (Links to SVA)	Accession List	Description	Sequence Length
☐ EMBL:EE667519	EE667519	EE667519	SAAH-aab96a03.g1 Agen 0058 Schmidtea mediterranea cDNA 5' similar to gb|AAL99921.1|AF432214_1 CLL-associated antigen KW-14 [Homo sapiens], mRNA sequence.	730
☐ EMBL:EE668092	EE668092	EE668092	SAAH-aac01g01.g1 Agen 0058 Schmidtea mediterranea cDNA 5' similar to gb|AAL99921.1|AF432214_1 CLL-associated antigen KW-14 [Homo sapiens], mRNA sequence.	692
☐ EMBL:EE669096	EE669096	EE669096	SAAH-aac34b09.g1 Agen 0058 Schmidtea mediterranea cDNA 5' similar to gb|AAL99921.1|AF432214_1 CLL-associated antigen KW-14 [Homo sapiens], mRNA sequence.	730
☐ EMBL:EE669325	EE669325	EE669325	SAAH-aac36e01.g1 Agen 0058 Schmidtea mediterranea cDNA 5' similar to gb|AAL99921.1|AF432214_1 CLL-associated antigen KW-14 [Homo sapiens], mRNA sequence.	713
☐ EMBL:EE669631	EE669631	EE669631	SAAH-aac39f08.g1 Agen 0058 Schmidtea mediterranea cDNA 5' similar to gb|AAL99921.1|AF432214_1 CLL-associated antigen KW-14 [Homo sapiens], mRNA sequence.	762
☐ EMBL:EE669677	EE669677	EE669677	SAAH-aac40b04.g1 Agen 0058 Schmidtea mediterranea cDNA 5' similar to gb|AAL99921.1|AF432214_1 CLL-associated antigen KW-14 [Homo sapiens], mRNA	688

2 - Recherche standard



Search Options

1. **Select** the **databanks** you want to search
2. Enter your **search terms** in the **Quick Search** box, or choose a **query form** from below

You can **browse** through all the **entries** in any **databanks**.
First, **select** the **databanks** you want to browse, then click:

Tips

- bookmark this **link** to return to your project
- *Linking to SRS?*

- Please read our Linking to SRS guide for important information regarding linking to our SRS server.

BookMarkLets

About BookmarkLets

- Protein Seq
- DNA/RNA Seq

Available Databanks

⊞ Expand all ⊟ Collapse all Show databanks tooltips: ☑

⊟ **Literature, Bibliography and Reference Databases**

all ☐ TAXONOMY ☐ GENETICCODE ☐ OMIM ☐ MEDLINE ☐ Patent Abstracts ☐ Karyn's Genomes

Literature, Bibliography and Reference Databases - subsections

all ☐ MEDLINE (Updates) ☐ MEDLINE (Main Release 2006) ☐ MED2PUB

⊞ **Gene Dictionaries and Ontologies**

⊟ **Nucleotide sequence databases**

all ☐ EMBL ☐ Patent DNA ☑ IMGT/LIGM-DB ☐ EMBL (Coding Sequences) ☐ IMGT/HLA ☐ IPD-KIR ☐ EMBL (Contig) ☐ Genome Reviews ☐ EMBL (Contigs expanded) ☐ EMBL (Annotated Cons) ☐ RefSeq Genome ☐ LiveLists ☐ EMBL ID/Accession Mapping ☐ EMBL MGA

Nucleotide sequence databases - subsections

all ☐ EMBL (Updates) ☐ EMBL (Release) ☐ EMBL (Whole Genome Shotgun) ☐ EMBL (Whole Genome Shotgun release) ☐ EMBL (Whole Genome Shotgun updates) ☐ EMBL (Contig release) ☐ EMBL (Contig updates) ☐ EMBL (Contigs expanded release) ☐ EMBL (Contigs expanded updates) ☐ EMBL (Annotated Cons release) ☐ EMBL (Annotated Cons updates) ☐ RefSeq Genome (Release) ☐ RefSeq Genome (Updates) ☐ EMBL (Whole Genome Shotgun Masters)

⊞ **Nucleotide related databases**

⊟ **UniProt Universal Protein Resource**

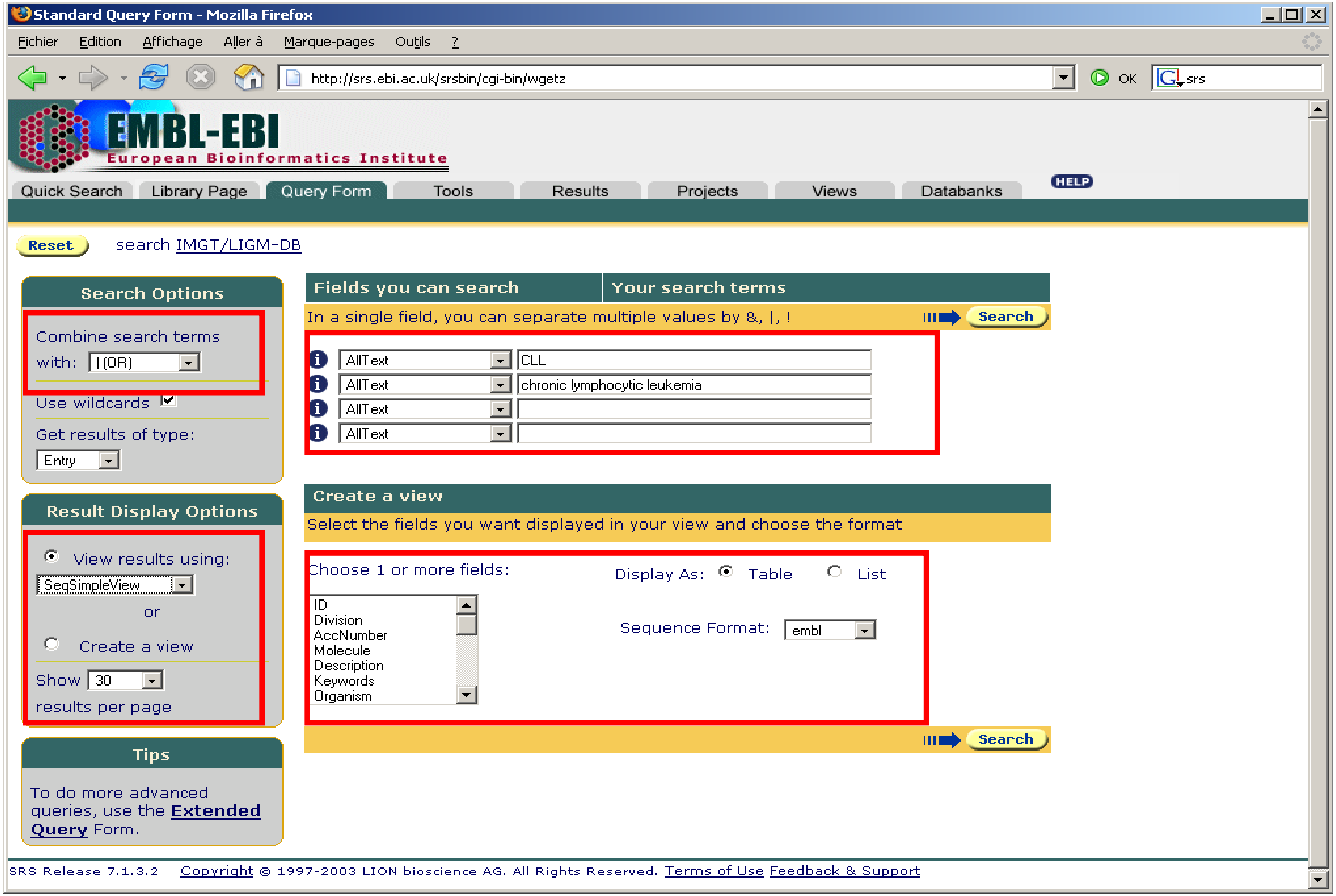
all ☐ UniProtKB ☐ UniProtKB/Swiss-Prot ☐ UniProtKB/TrEMBL ☐ UniRef100 ☐ UniRef90 ☐ UniRef50 ☐ UniParc

⊞ **Other protein sequence databases**

⊞ **Protein function, structure and interaction databases**

⊞ **Enzymes, reactions and metabolic pathway databases**

⊞ **Mutation and SNP databases**



search IMGT/LIGM-DB

Search Options

Combine search terms with: | (OR)

Use wildcards ☑

Get results of type: Entry

Result Display Options

● View results using: SeqSimpleView

or

○ Create a view

Show 30 results per page

Tips

To do more advanced queries, use the **Extended Query** Form.

Fields you can search	Your search terms
In a single field, you can separate multiple values by &, |, !	
AllText	CLL
AllText	chronic lymphocytic leukemia
AllText	
AllText	

Create a view

Select the fields you want displayed in your view and choose the format

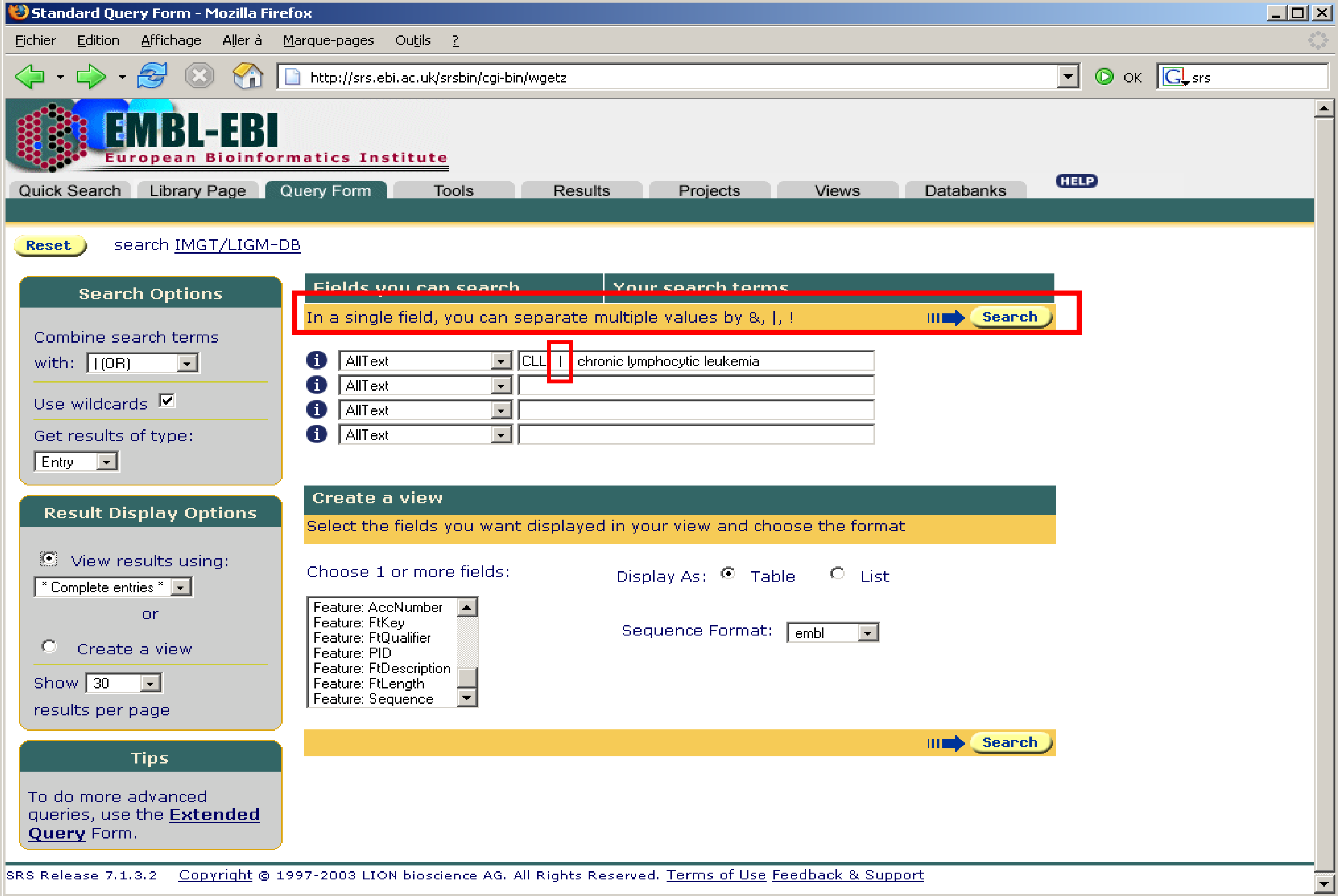
Choose 1 or more fields:

ID
Division
AccNumber
Molecule
Description
Keywords
Organism

Display As: ● Table ○ List

Sequence Format: embl

SRS Release 7.1.3.2 Copyright © 1997-2003 LION bioscience AG. All Rights Reserved. Terms of Use Feedback & Support



http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz

search IMGT/LIGM-DB

Search Options

Combine search terms with: | (OR)

Use wildcards ☑

Get results of type: Entry

Result Display Options

● View results using: * Complete entries *

or

○ Create a view

Show 30 results per page

Tips

To do more advanced queries, use the **Extended Query** Form.

Fields you can search | **Your search terms**

In a single field, you can separate multiple values by &, |, !

AllText: CLL | chronic lymphocytic leukemia

AllText:

AllText:

AllText:

Create a view

Select the fields you want displayed in your view and choose the format

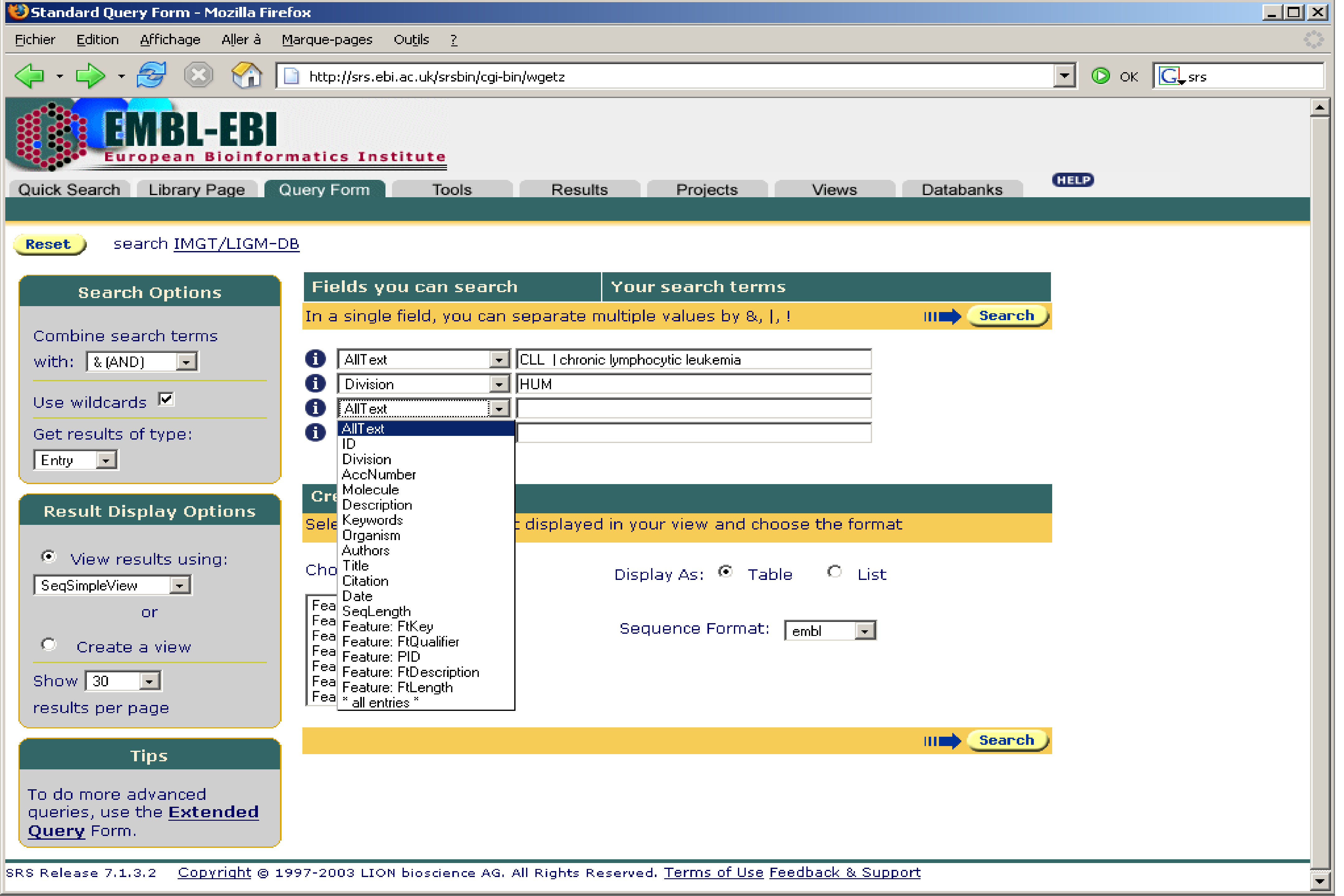
Choose 1 or more fields:

Feature: AccNumber
Feature: FtKey
Feature: FtQualifier
Feature: PID
Feature: FtDescription
Feature: FtLength
Feature: Sequence

Display As: ● Table ○ List

Sequence Format: embl

SRS Release 7.1.3.2 Copyright © 1997-2003 LION bioscience AG. All Rights Reserved. Terms of Use Feedback & Support



search IMGT/LIGM-DB

Search Options

Combine search terms with: & (AND)

Use wildcards ☑

Get results of type: Entry

Result Display Options

● View results using: SeqSimpleView

or

○ Create a view

Show 30 results per page

Tips

To do more advanced queries, use the **Extended Query** Form.

Fields you can search	Your search terms
AllText	CLL | chronic lymphocytic leukemia
Division	HUM
AllText	

In a single field, you can separate multiple values by &, |, !

Cre…

Sele… displayed in your view and choose the format

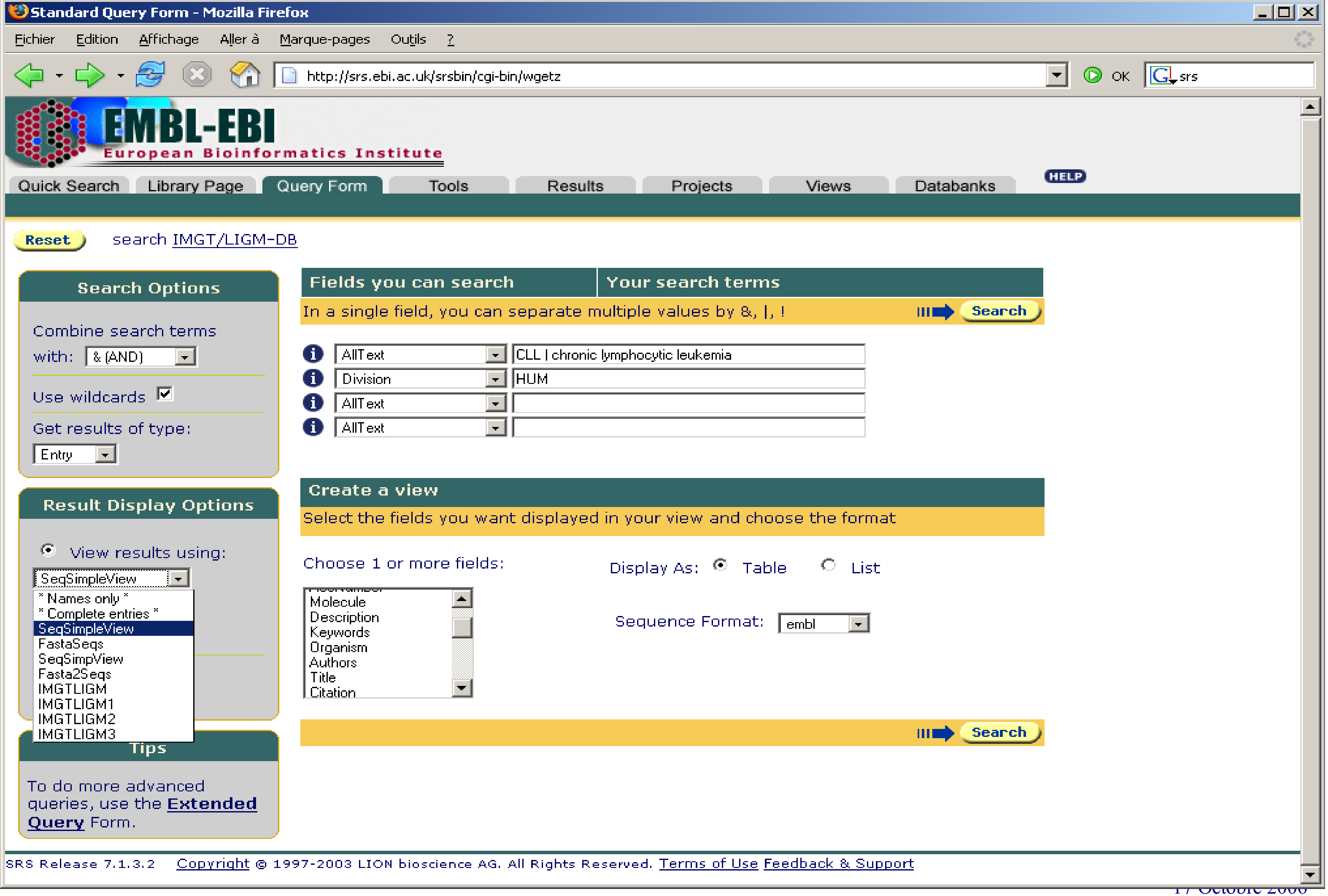
Cho…

Fea…

Display As: ● Table ○ List

Sequence Format: embl

SRS Release 7.1.3.2 Copyright © 1997-2003 LION bioscience AG. All Rights Reserved. Terms of Use Feedback & Support



http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz

search IMGT/LIGM-DB

Search Options

Combine search terms with: & (AND)

Use wildcards ☑

Get results of type: Entry

Result Display Options

● View results using: SeqSimpleView

- * Names only *
- * Complete entries *
- SeqSimpleView
- FastaSeqs
- SeqSimpView
- Fasta2Seqs
- IMGTLIGM
- IMGTLIGM1
- IMGTLIGM2
- IMGTLIGM3

Tips

To do more advanced queries, use the **Extended Query** Form.

Fields you can search	Your search terms
AllText	CLL | chronic lymphocytic leukemia
Division	HUM
AllText	
AllText	

In a single field, you can separate multiple values by &, |, !

Create a view

Select the fields you want displayed in your view and choose the format

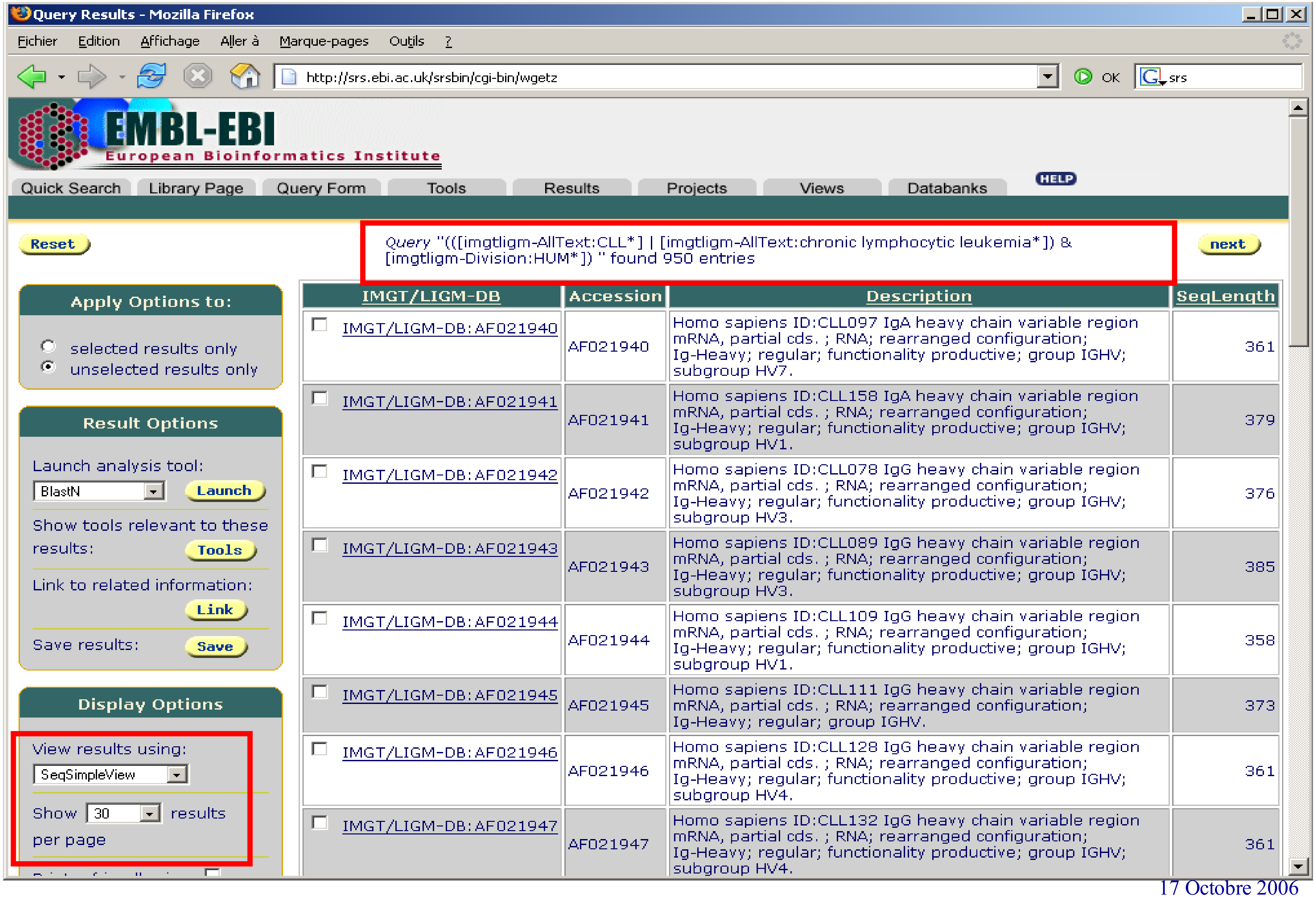
Choose 1 or more fields:

- Molecule
- Description
- Keywords
- Organism
- Authors
- Title
- Citation

Display As: ● Table ○ List

Sequence Format: embl

SRS Release 7.1.3.2 Copyright © 1997-2003 LION bioscience AG. All Rights Reserved. Terms of Use Feedback & Support



Query "(([imgtligm-AllText:CLL*] | [imgtligm-AllText:chronic lymphocytic leukemia*]) & [imgtligm-Division:HUM*]) " found 950 entries

Apply Options to:

○ selected results only
● unselected results only

Result Options

Launch analysis tool: BlastN

Show tools relevant to these results:

Link to related information:

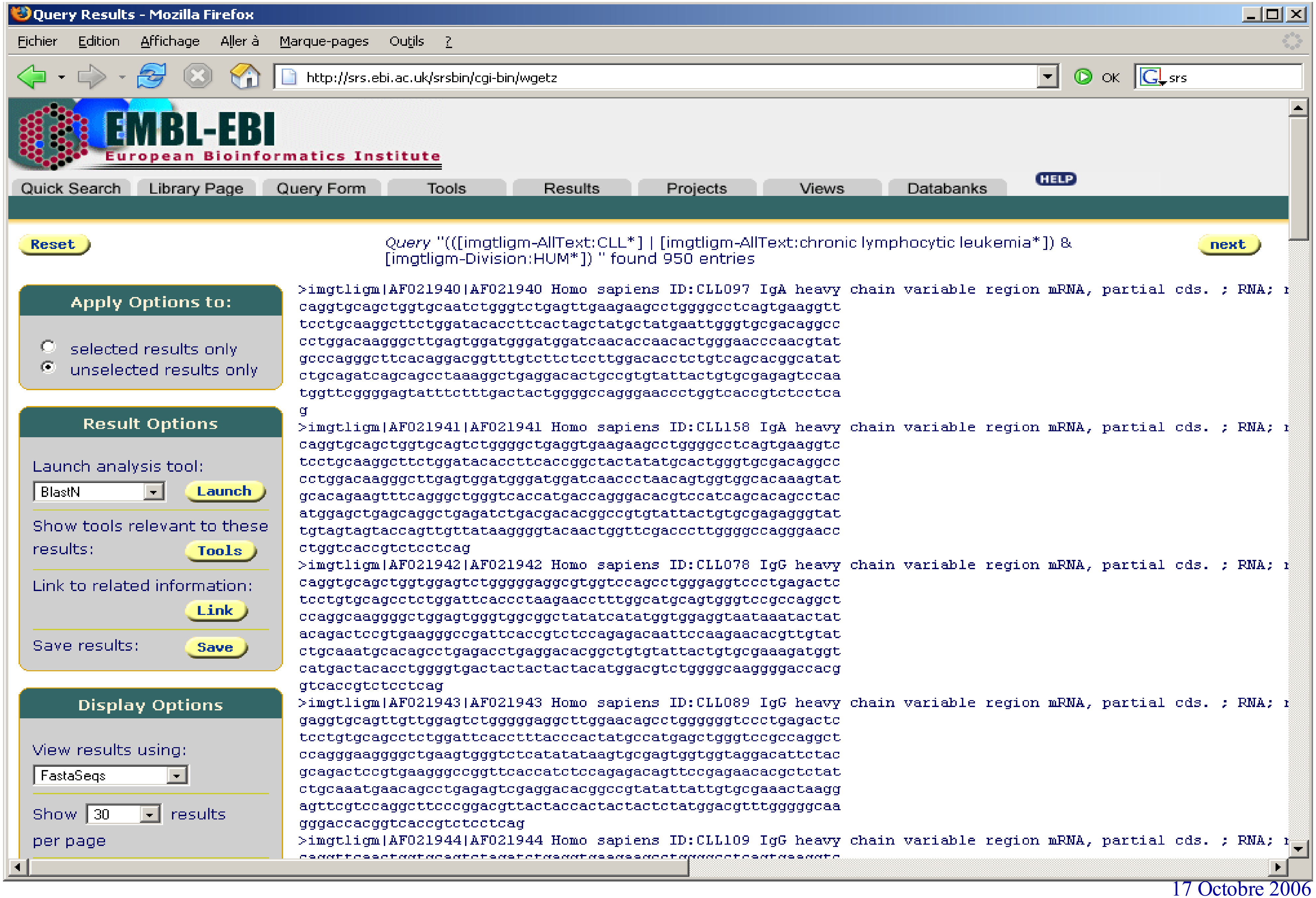
Save results:

Display Options

View results using: SeqSimpleView

Show 30 results per page

	IMGT/LIGM-DB	Accession	Description	SeqLength
☐	IMGT/LIGM-DB:AF021940	AF021940	Homo sapiens ID:CLL097 IgA heavy chain variable region mRNA, partial cds. ; RNA; rearranged configuration; Ig-Heavy; regular; functionality productive; group IGHV; subgroup HV7.	361
☐	IMGT/LIGM-DB:AF021941	AF021941	Homo sapiens ID:CLL158 IgA heavy chain variable region mRNA, partial cds. ; RNA; rearranged configuration; Ig-Heavy; regular; functionality productive; group IGHV; subgroup HV1.	379
☐	IMGT/LIGM-DB:AF021942	AF021942	Homo sapiens ID:CLL078 IgG heavy chain variable region mRNA, partial cds. ; RNA; rearranged configuration; Ig-Heavy; regular; functionality productive; group IGHV; subgroup HV3.	376
☐	IMGT/LIGM-DB:AF021943	AF021943	Homo sapiens ID:CLL089 IgG heavy chain variable region mRNA, partial cds. ; RNA; rearranged configuration; Ig-Heavy; regular; functionality productive; group IGHV; subgroup HV3.	385
☐	IMGT/LIGM-DB:AF021944	AF021944	Homo sapiens ID:CLL109 IgG heavy chain variable region mRNA, partial cds. ; RNA; rearranged configuration; Ig-Heavy; regular; functionality productive; group IGHV; subgroup HV1.	358
☐	IMGT/LIGM-DB:AF021945	AF021945	Homo sapiens ID:CLL111 IgG heavy chain variable region mRNA, partial cds. ; RNA; rearranged configuration; Ig-Heavy; regular; group IGHV.	373
☐	IMGT/LIGM-DB:AF021946	AF021946	Homo sapiens ID:CLL128 IgG heavy chain variable region mRNA, partial cds. ; RNA; rearranged configuration; Ig-Heavy; regular; functionality productive; group IGHV; subgroup HV4.	361
☐	IMGT/LIGM-DB:AF021947	AF021947	Homo sapiens ID:CLL132 IgG heavy chain variable region mRNA, partial cds. ; RNA; rearranged configuration; Ig-Heavy; regular; functionality productive; group IGHV; subgroup HV4.	361



http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz

Query "(([imgtligm-AllText:CLL*] | [imgtligm-AllText:chronic lymphocytic leukemia*]) & [imgtligm-Division:HUM*]) " found 950 entries

Apply Options to:

○ selected results only
● unselected results only

Result Options

Launch analysis tool: BlastN

Show tools relevant to these results:

Link to related information:

Save results:

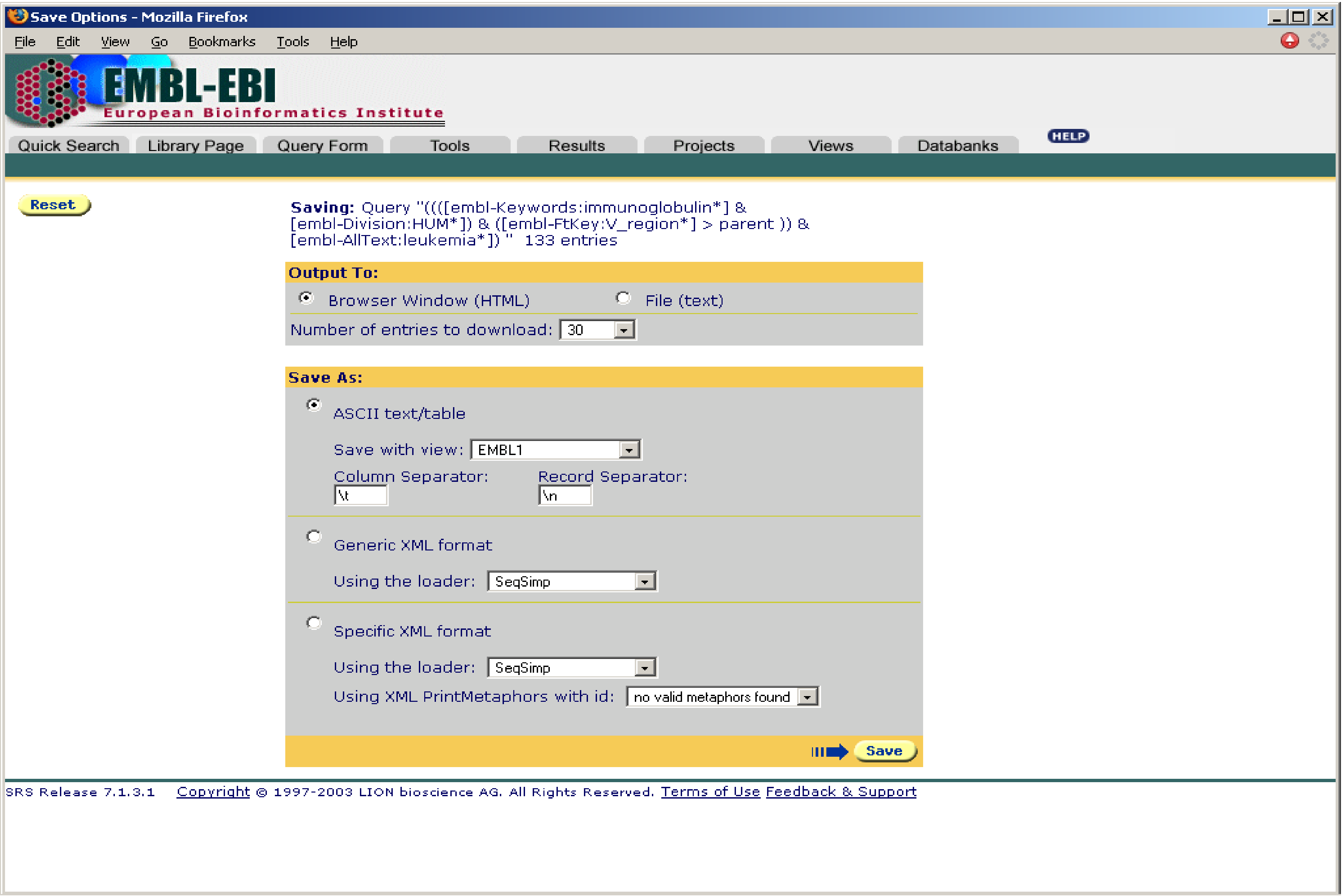
Display Options

View results using: FastaSeqs

Show 30 results per page

```
>imgtligm|AF021940|AF021940 Homo sapiens ID:CLL097 IgA heavy chain variable region mRNA, partial cds. ; RNA;
caggtgcagctggtgcaatctgggtctgagttgaagaagcctggggcctcagtgaaggtt
tcctgcaaggcttctggatacaccttcactagctatgctatgaattgggtgcgacaggcc
cctggacaagggcttgagtggatgggatggatcaacaccaacactgggaacccaacgtat
gcccagggcttcacaggacggtttgtcttctccttggacacctctgtcagcacggcatat
ctgcagatcagcagcctaaaggctgaggacactgccgtgtattactgtgcgagagtccaa
tggttcggggagtatttctttgactactggggccagggaaccctggtcaccgtctcctca
g
>imgtligm|AF021941|AF021941 Homo sapiens ID:CLL158 IgA heavy chain variable region mRNA, partial cds. ; RNA;
caggtgcagctggtgcagtctggggctgaggtgaagaagcctggggcctcagtgaaggtc
tcctgcaaggcttctggatacaccttcaccggctactatatgcactgggtgcgacaggcc
cctggacaagggcttgagtggatgggatggatcaaccctaacagtggtggcacaaagtat
gcacagaagtttcagggctgggtcaccatgaccagggacacgtccatcagcacagcctac
atggagctgagcaggctgagatctgacgacacggccgtgtattactgtgcgagagggtat
tgtagtagtaccagttgttataaggggtacaactggttcgacccttggggccagggaacc
ctggtcaccgtctcctcag
>imgtligm|AF021942|AF021942 Homo sapiens ID:CLL078 IgG heavy chain variable region mRNA, partial cds. ; RNA;
caggtgcagctggtggagtctgggggaggcgtggtccagcctgggaggtccctgagactc
tcctgtgcagcctctggattcaccctaagaacctttggcatgcagtgggtccgccaggct
ccaggcaaggggctggagtgggtggcggctatatcatatggtggaggtaataaatactat
acagactccgtgaagggccgattcaccgtctccagagacaattccaagaacacgttgtat
ctgcaaatgcacagcctgagacctgaggacacggctgtgtattactgtgcgaaagatggt
catgactacacctggggtgactactactactacatggacgtctggggcaaggggaccacg
gtcaccgtctcctcag
>imgtligm|AF021943|AF021943 Homo sapiens ID:CLL089 IgG heavy chain variable region mRNA, partial cds. ; RNA;
gaggtgcagttgttggagtctggggaggcttggaacagcctgggggtccctgagactc
tcctgtgcagcctctggattcacctttacccactatgccatgagctgggtccgccaggct
ccagggaaggggctgaagtgggtctcatatataagtgcgagtggtggtaggacattctac
gcagactccgtgaagggccggttcaccatctccagagacagttccgagaacacgctctat
ctgcaaatgaacagcctgagagtcgaggacacggccgtatattattgtgcgaaactaagg
agttcgtccaggcttcccggacgttactaccactactactctatggacgtttgggggcaa
gggaccacggtcaccgtctcctcag
>imgtligm|AF021944|AF021944 Homo sapiens ID:CLL109 IgG heavy chain variable region mRNA, partial cds. ; RNA;
```

Sauvegarde des resultats



Saving: Query "((([embl-Keywords:immunoglobulin*] & [embl-Division:HUM*]) & ([embl-FtKey:V_region*] > parent)) & [embl-AllText:leukemia*]) " 133 entries

Output To:

● Browser Window (HTML) ○ File (text)

Number of entries to download: 30

Save As:

● ASCII text/table

Save with view: EMBL1

Column Separator: \t Record Separator: \n

○ Generic XML format

Using the loader: SeqSimp

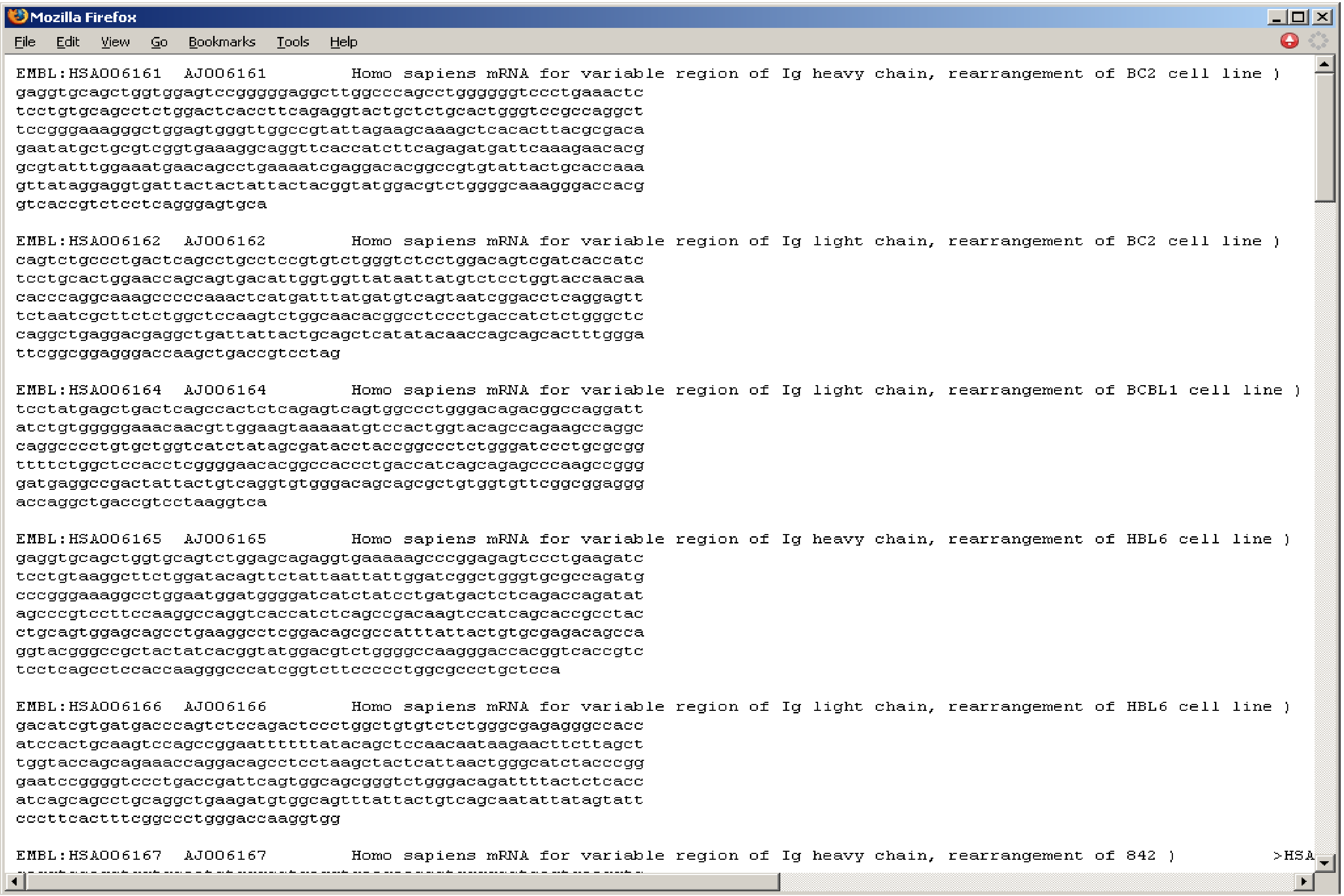
○ Specific XML format

Using the loader: SeqSimp

Using XML PrintMetaphors with id: no valid metaphors found

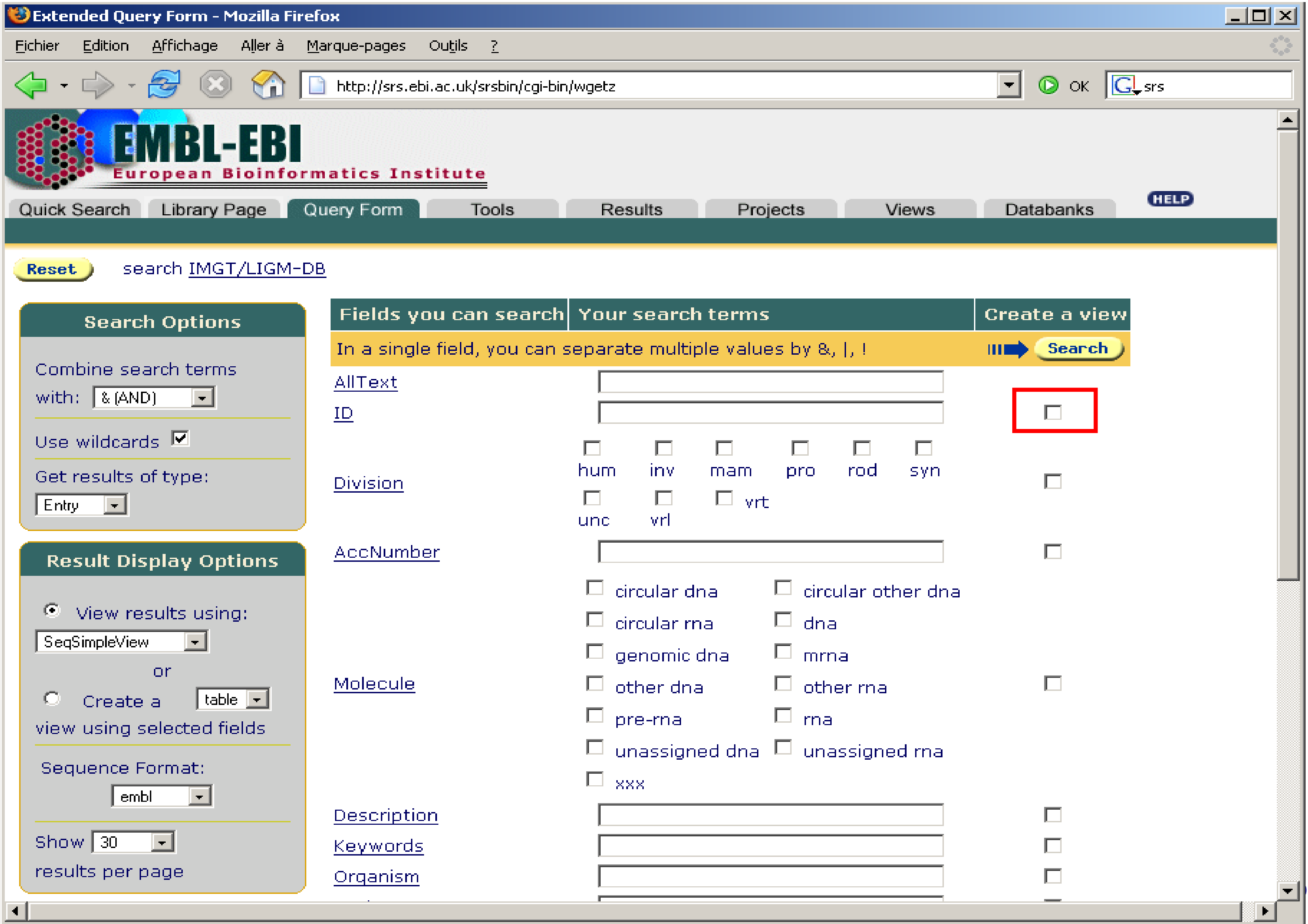
SRS Release 7.1.3.1 Copyright © 1997-2003 LION bioscience AG. All Rights Reserved. Terms of Use Feedback & Support

Sauvegarde des resultats



```
EMBL:HSA006161  AJ006161        Homo sapiens mRNA for variable region of Ig heavy chain, rearrangement of BC2 cell line )
gaggtgcagctggtggagtccgggggaggcttggcccagcctggggggtccctgaaactc
tcctgtgcagcctctggactcaccttcagaggtactgctctgcactgggtccgccaggct
tccgggaaagggctggagtgggttggccgtattagaagcaaagctcacacttacgcgaca
gaatatgctgcgtcggtgaaaggcaggttcaccatcttcagagatgattcaaagaacacg
gcgtatttggaaatgaacagcctgaaaatcgaggacacggccgtgtattactgcaccaaa
gttataggaggtgattactactattactacggtatggacgtctggggcaaagggaccacg
gtcaccgtctcctcagggagtgca

EMBL:HSA006162  AJ006162        Homo sapiens mRNA for variable region of Ig light chain, rearrangement of BC2 cell line )
cagtctgccctgactcagcctgcctccgtgtctgggtctcctggacagtcgatcaccatc
tcctgcactggaaccagcagtgacattggtggttataattatgtctcctggtaccaacaa
cacccaggcaaagcccccaaactcatgatttatgatgtcagtaatcggacctcaggagtt
tctaatcgcttctctggctccaagtctggcaacacggcctccctgaccatctctgggctc
caggctgaggacgaggctgattattactgcagctcatatacaaccagcagcactttggga
ttcggcggagggaccaagctgaccgtcctag

EMBL:HSA006164  AJ006164        Homo sapiens mRNA for variable region of Ig light chain, rearrangement of BCBL1 cell line )
tcctatgagctgactcagccactctcagagtcagtggccctgggacagacggccaggatt
atctgtgggggaaacaacgttggaagtaaaaatgtccactggtacagccagaagccaggc
caggcccctgtgctggtcatctatagcgatacctaccggccctctgggatccctgcgcgg
ttttctggctccacctcggggaacacggccaccctgaccatcagcagagcccaagccggg
gatgaggccgactattactgtcaggtgtgggacagcagcgctgtggtgttcggcggaggg
accaggctgaccgtcctaaggtca

EMBL:HSA006165  AJ006165        Homo sapiens mRNA for variable region of Ig heavy chain, rearrangement of HBL6 cell line )
gaggtgcagctggtgcagtctggagcagaggtgaaaaagcccggagagtccctgaagatc
tcctgtaaggcttctggatacagttctattaattattggatcggctgggtgcgccagatg
cccgggaaaggcctggaatggatggggatcatctatcctgatgactctcagaccagatat
agcccgtccttccaaggccaggtcaccatctcagccgacaagtccatcagcaccgcctac
ctgcagtggagcagcctgaaggcctcggacagcgccatttattactgtgcgagacagcca
ggtacgggccgctactatcacggtatggacgtctggggccaagggaccacggtcaccgtc
tcctcagcctccaccaagggcccatcggtcttccccctggcgccctgctcca

EMBL:HSA006166  AJ006166        Homo sapiens mRNA for variable region of Ig light chain, rearrangement of HBL6 cell line )
gacatcgtgatgacccagtctccagactccctggctgtgtctctgggcgagagggccacc
atccactgcaagtccagccggaatttttatacagctccaacaataagaacttcttagct
tggtaccagcagaaaccaggacagcctcctaagctactcattaactgggcatctacccgg
gaatccggggtccctgaccgattcagtggcagcgggtctgggacagatttactctcacc
atcagcagcctgcaggctgaagatgtggcagtttattactgtcagcaatattatagtatt
cccttcactttcggccctgggaccaaggtgg

EMBL:HSA006167  AJ006167        Homo sapiens mRNA for variable region of Ig heavy chain, rearrangement of 842 )        >HSA
```

3 - Recherche étendue:



search IMGT/LIGM-DB

Search Options

Combine search terms with: & (AND)

Use wildcards ☑

Get results of type: Entry

Result Display Options

● View results using: SeqSimpleView

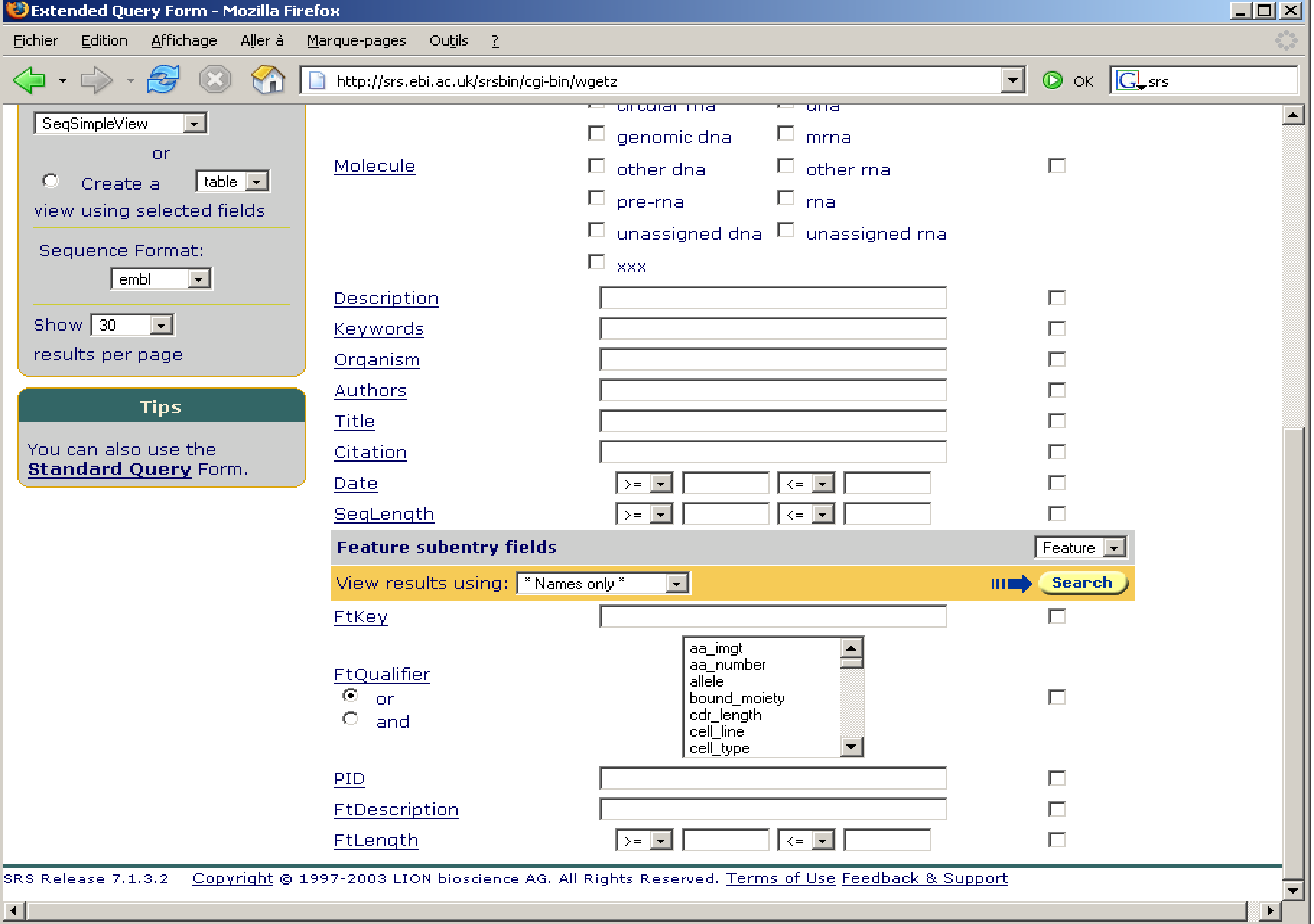
or

○ Create a table view using selected fields

Sequence Format: embl

Show 30 results per page

Fields you can search	Your search terms	Create a view
In a single field, you can separate multiple values by &, |, !		
AllText		
ID		☐
Division	☐ hum ☐ inv ☐ mam ☐ pro ☐ rod ☐ syn ☐ unc ☐ vrl ☐ vrt	☐
AccNumber		☐
Molecule	☐ circular dna ☐ circular other dna ☐ circular rna ☐ dna ☐ genomic dna ☐ mrna ☐ other dna ☐ other rna ☐ pre-rna ☐ rna ☐ unassigned dna ☐ unassigned rna ☐ xxx	☐
Description		☐
Keywords		☐
Organism		☐



SeqSimpleView

or

○ Create a table view using selected fields

Sequence Format: embl

Show 30 results per page

Tips

You can also use the **Standard Query** Form.

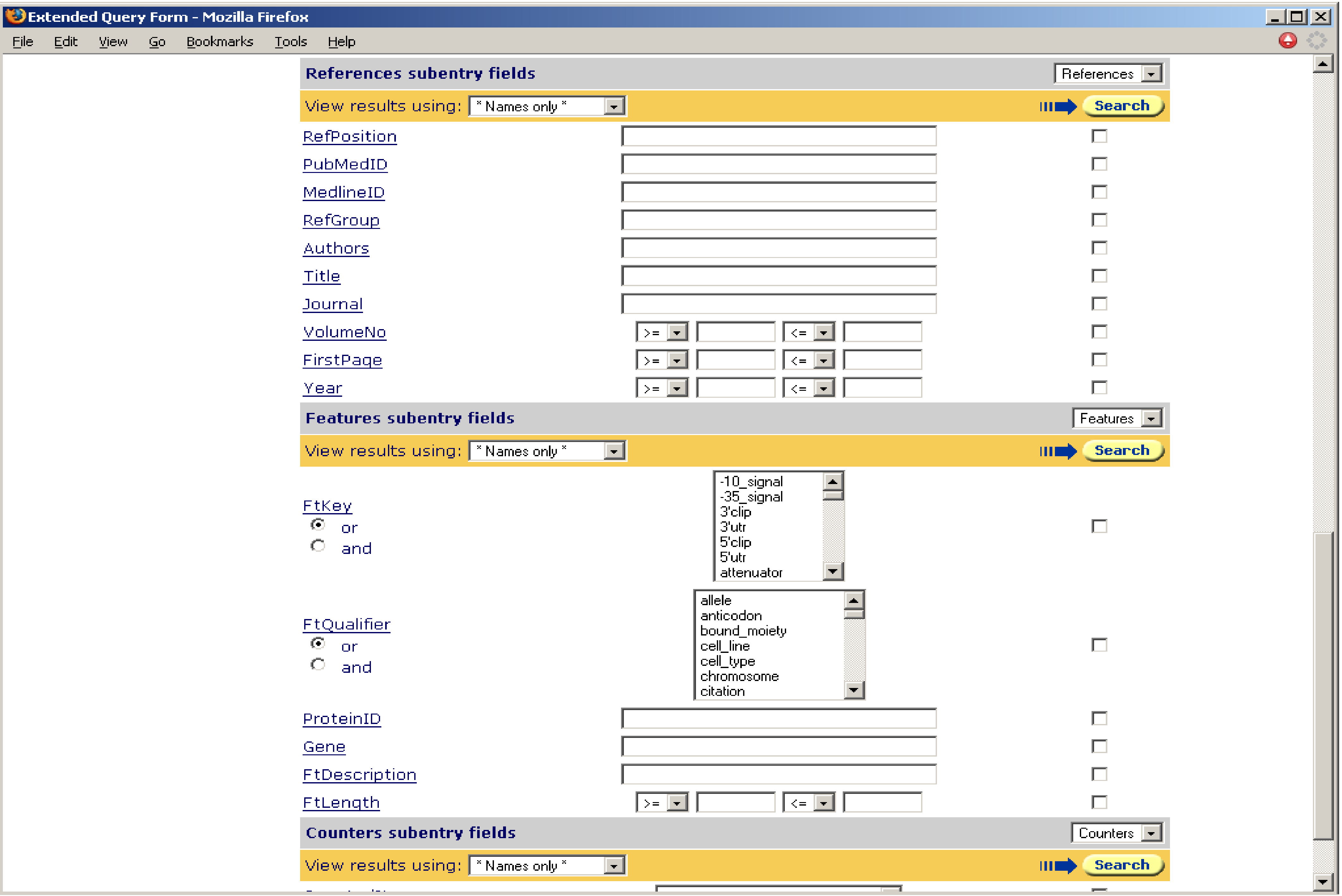
Molecule	☐ genomic dna ☐ other dna ☐ pre-rna ☐ unassigned dna ☐ xxx	☐ mrna ☐ other rna ☐ rna ☐ unassigned rna	☐
Description			☐
Keywords			☐
Organism			☐
Authors			☐
Title			☐
Citation			☐
Date	>=	<=	☐
SeqLength	>=	<=	☐

Feature subentry fields Feature

View results using: * Names only *

FtKey			☐
FtQualifier ● or ○ and	aa_imgt aa_number allele bound_moiety cdr_length cell_line cell_type		☐
PID			☐
FtDescription			☐
FtLength	>=	<=	☐

SRS Release 7.1.3.2 Copyright © 1997-2003 LION bioscience AG. All Rights Reserved. Terms of Use Feedback & Support



References subentry fields References

View results using: * Names only * Search

- RefPosition ☐
- PubMedID ☐
- MedlineID ☐
- RefGroup ☐
- Authors ☐
- Title ☐
- Journal ☐
- VolumeNo >= <= ☐
- FirstPage >= <= ☐
- Year >= <= ☐

Features subentry fields Features

View results using: * Names only * Search

FtKey ● or ○ and ☐
- -10_signal
- -35_signal
- 3'clip
- 3'utr
- 5'clip
- 5'utr
- attenuator

FtQualifier ● or ○ and ☐
- allele
- anticodon
- bound_moiety
- cell_line
- cell_type
- chromosome
- citation

- ProteinID ☐
- Gene ☐
- FtDescription ☐
- FtLength >= <= ☐

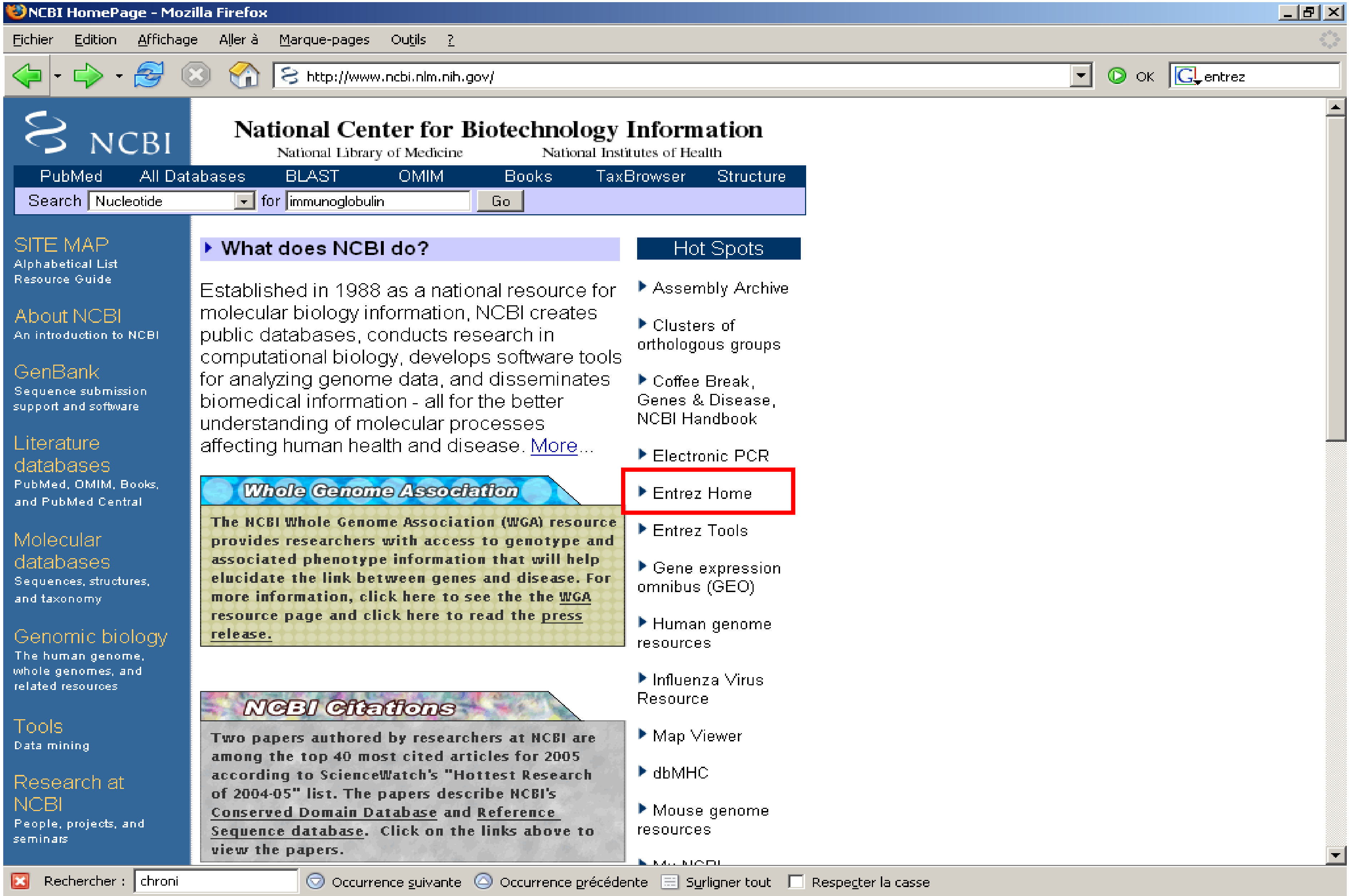
Counters subentry fields Counters

View results using: * Names only * Search

ENTREZ

http://www.ncbi.nlm.nih.gov/

La page d'accueil du NCBI (http://www.ncbi.nlm.nih.gov/)



National Center for Biotechnology Information

National Library of Medicine National Institutes of Health

Search: Nucleotide for: immunoglobulin

Rechercher : chroni

SITE MAP
Alphabetical List
Resource Guide

About NCBI
An introduction to NCBI

GenBank
Sequence submission support and software

Literature databases
PubMed, OMIM, Books, and PubMed Central

Molecular databases
Sequences, structures, and taxonomy

Genomic biology
The human genome, whole genomes, and related resources

Tools
Data mining

Research at NCBI
People, projects, and seminars

What does NCBI do?

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...

Whole Genome Association

The NCBI Whole Genome Association (WGA) resource provides researchers with access to genotype and associated phenotype information that will help elucidate the link between genes and disease. For more information, click here to see the the WGA resource page and click here to read the press release.

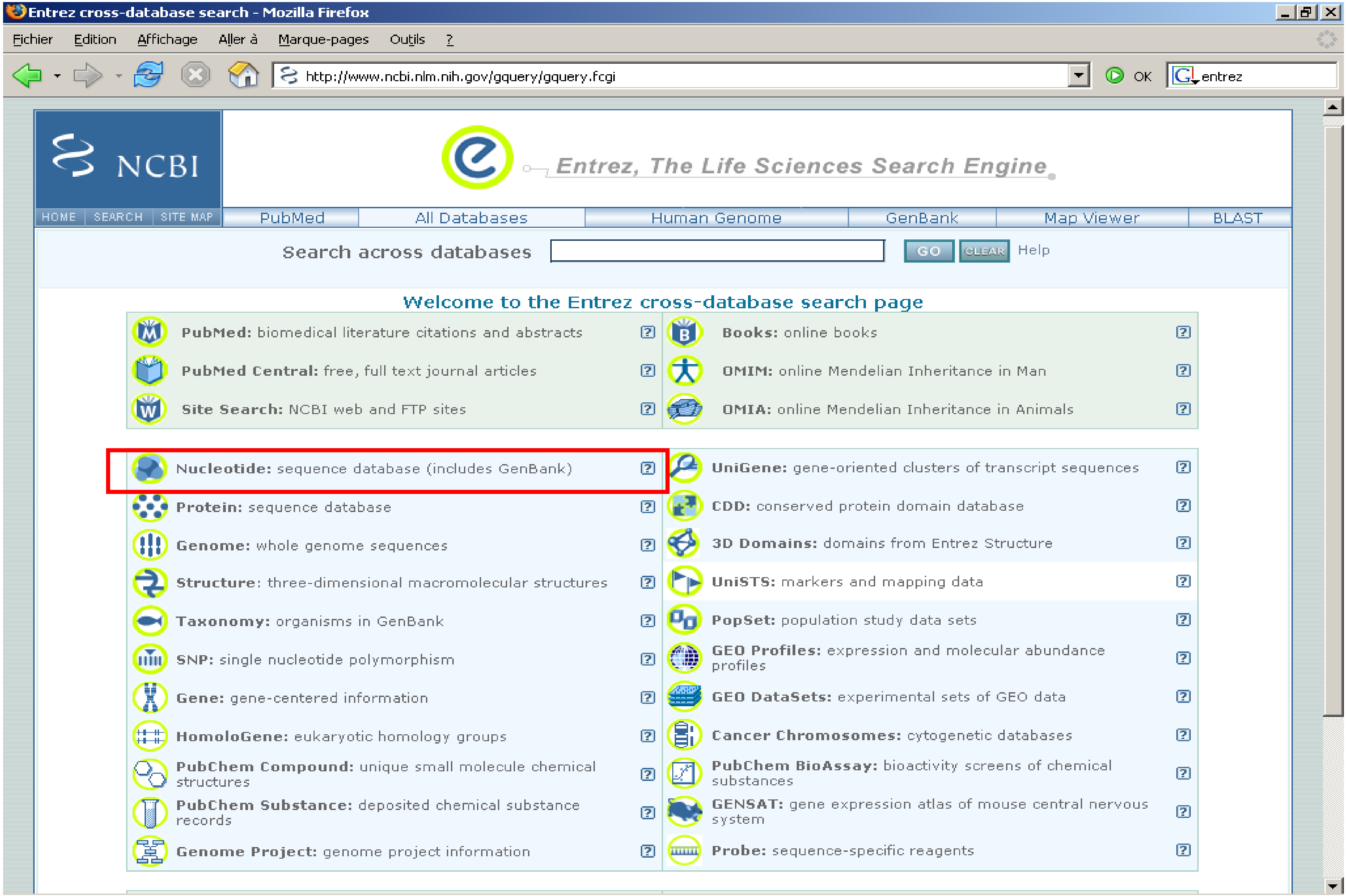
NCBI Citations

Two papers authored by researchers at NCBI are among the top 40 most cited articles for 2005 according to ScienceWatch's "Hottest Research of 2004-05" list. The papers describe NCBI's Conserved Domain Database and Reference Sequence database. Click on the links above to view the papers.

Hot Spots

- Assembly Archive
- Clusters of orthologous groups
- Coffee Break, Genes & Disease, NCBI Handbook
- Electronic PCR
- Entrez Home
- Entrez Tools
- Gene expression omnibus (GEO)
- Human genome resources
- Influenza Virus Resource
- Map Viewer
- dbMHC
- Mouse genome resources

Entrez at http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi



Search across databases

Welcome to the Entrez cross-database search page

PubMed: biomedical literature citations and abstracts

PubMed Central: free, full text journal articles

Site Search: NCBI web and FTP sites

Books: online books

OMIM: online Mendelian Inheritance in Man

OMIA: online Mendelian Inheritance in Animals

Nucleotide: sequence database (includes GenBank)

Protein: sequence database

Genome: whole genome sequences

Structure: three-dimensional macromolecular structures

Taxonomy: organisms in GenBank

SNP: single nucleotide polymorphism

Gene: gene-centered information

HomoloGene: eukaryotic homology groups

PubChem Compound: unique small molecule chemical structures

PubChem Substance: deposited chemical substance records

Genome Project: genome project information

UniGene: gene-oriented clusters of transcript sequences

CDD: conserved protein domain database

3D Domains: domains from Entrez Structure

UniSTS: markers and mapping data

PopSet: population study data sets

GEO Profiles: expression and molecular abundance profiles

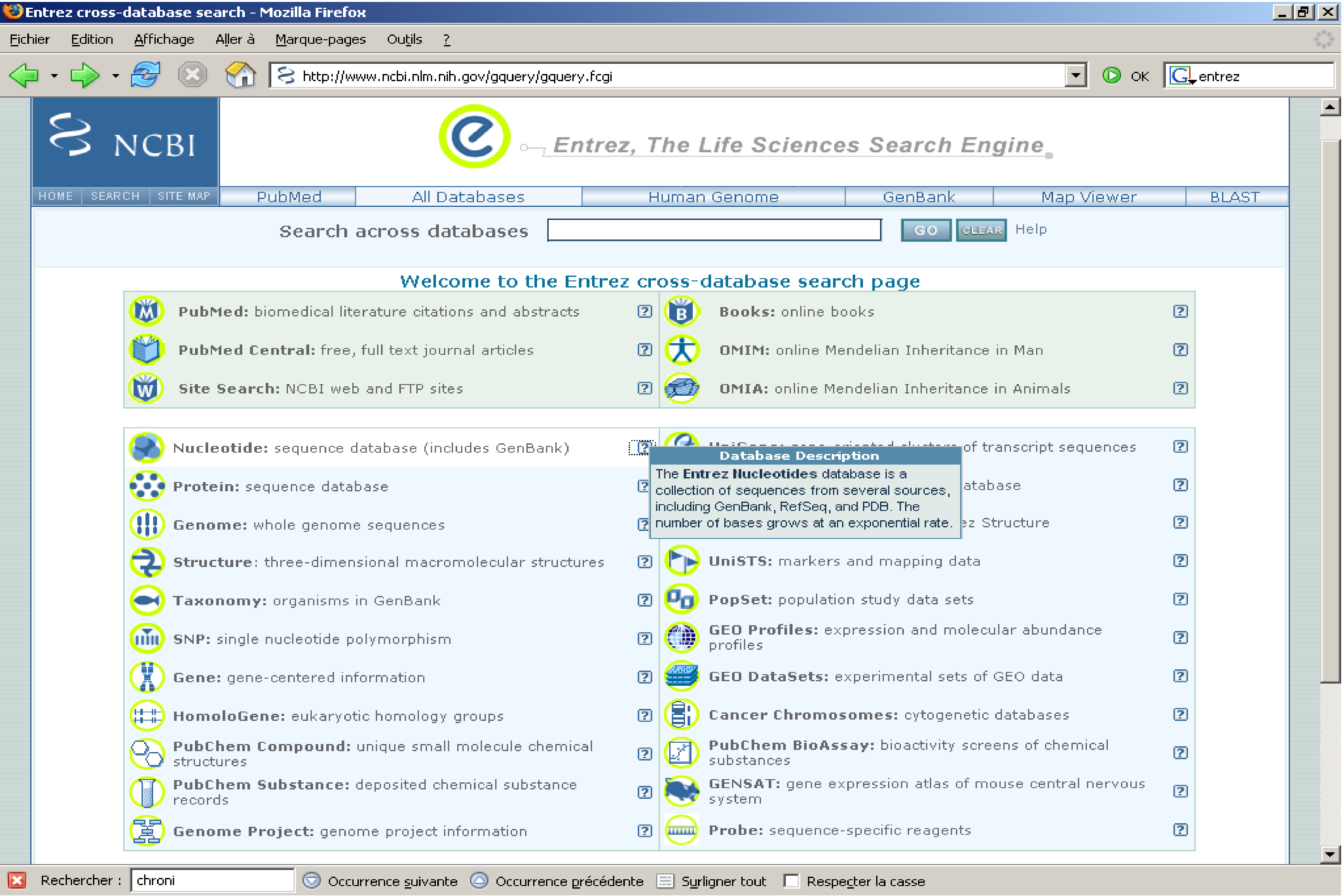
GEO DataSets: experimental sets of GEO data

Cancer Chromosomes: cytogenetic databases

PubChem BioAssay: bioactivity screens of chemical substances

GENSAT: gene expression atlas of mouse central nervous system

Probe: sequence-specific reagents



Rechercher : chroni

☐ Respecter la casse

Search across databases

Welcome to the Entrez cross-database search page

PubMed: biomedical literature citations and abstracts [?]

PubMed Central: free, full text journal articles [?]

Site Search: NCBI web and FTP sites [?]

Books: online books [?]

OMIM: online Mendelian Inheritance in Man [?]

OMIA: online Mendelian Inheritance in Animals [?]

Nucleotide: sequence database (includes GenBank) [?]

Protein: sequence database [?]

Genome: whole genome sequences [?]

Structure: three-dimensional macromolecular structures [?]

Taxonomy: organisms in GenBank [?]

SNP: single nucleotide polymorphism [?]

Gene: gene-centered information [?]

HomoloGene: eukaryotic homology groups [?]

PubChem Compound: unique small molecule chemical structures [?]

PubChem Substance: deposited chemical substance records [?]

Genome Project: genome project information [?]

of transcript sequences [?]

atabase [?]

ez Structure [?]

UniSTS: markers and mapping data [?]

PopSet: population study data sets [?]

GEO Profiles: expression and molecular abundance profiles [?]

GEO DataSets: experimental sets of GEO data [?]

Cancer Chromosomes: cytogenetic databases [?]

PubChem BioAssay: bioactivity screens of chemical substances [?]

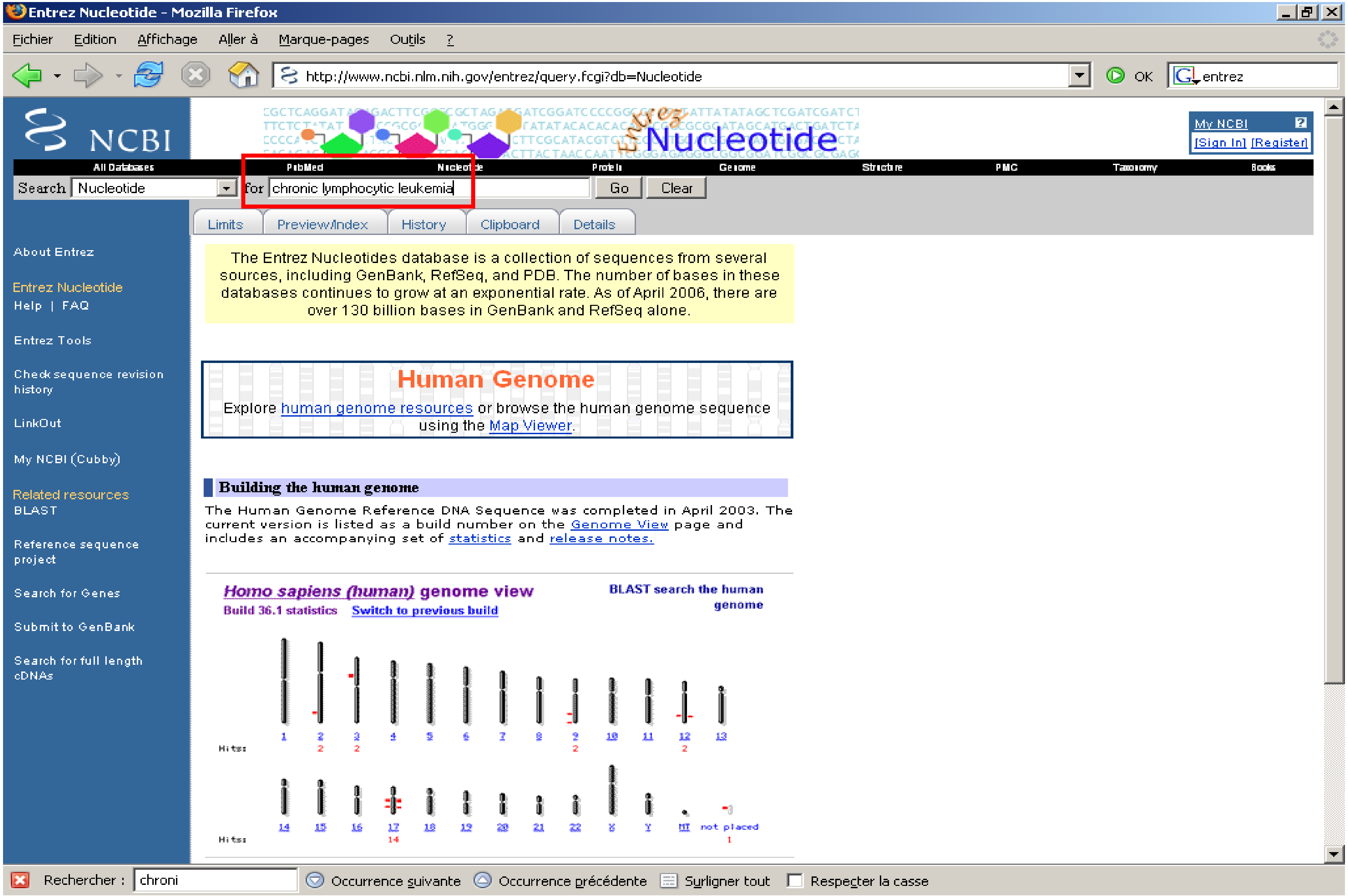
GENSAT: gene expression atlas of mouse central nervous system [?]

Probe: sequence-specific reagents [?]

Database Description

The **Entrez Nucleotides** database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. The number of bases grows at an exponential rate.

Recherche générale



http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide

entrez

Search Nucleotide for chronic lymphocytic leukemia

The Entrez Nucleotides database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. The number of bases in these databases continues to grow at an exponential rate. As of April 2006, there are over 130 billion bases in GenBank and RefSeq alone.

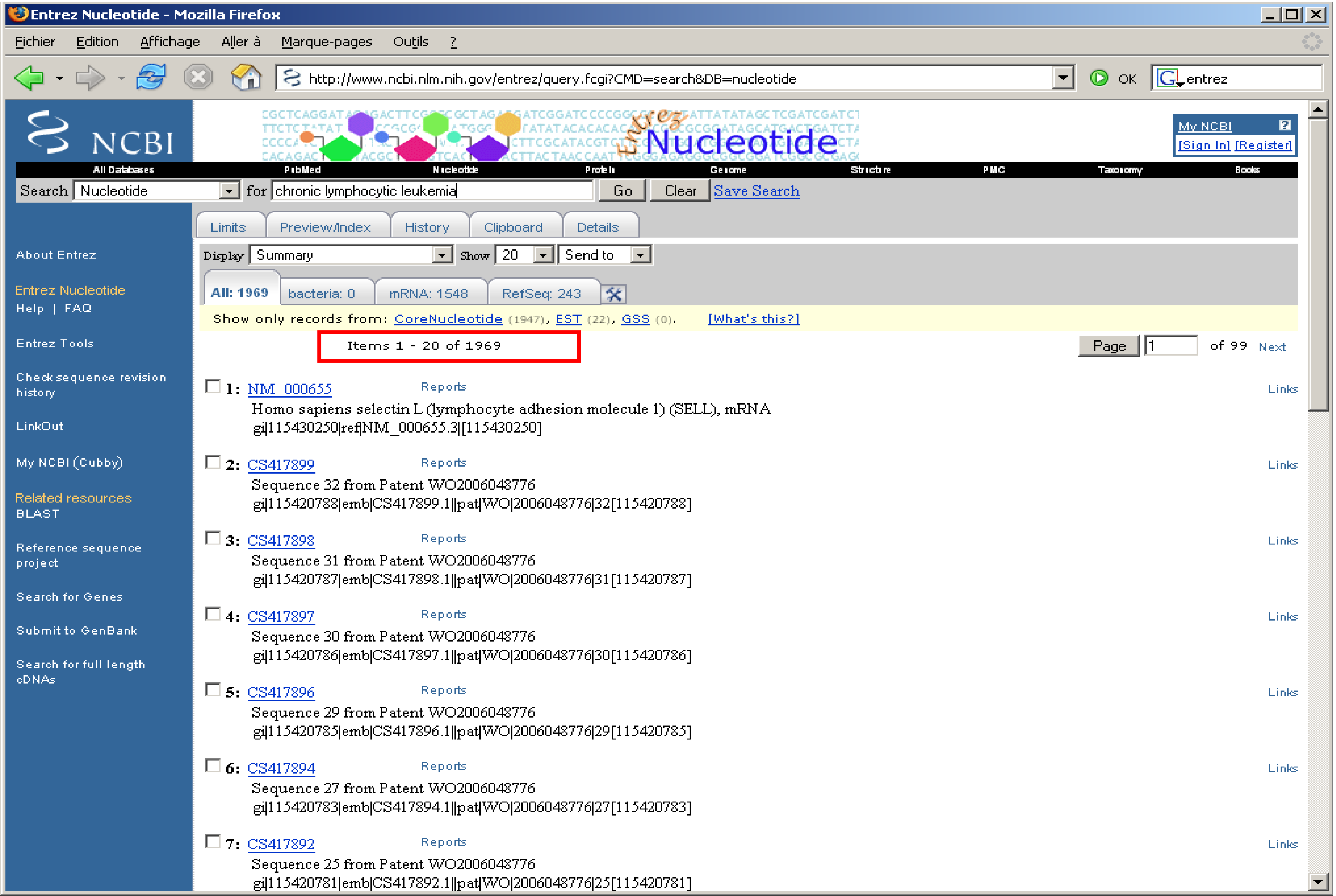
Human Genome

Explore human genome resources or browse the human genome sequence using the Map Viewer.

Building the human genome

The Human Genome Reference DNA Sequence was completed in April 2003. The current version is listed as a build number on the Genome View page and includes an accompanying set of statistics and release notes.

Rechercher : chroni



http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=nucleotide

entrez

Search Nucleotide for chronic lymphocytic leukemia

Display Summary Show 20

All: 1969 | bacteria: 0 | mRNA: 1548 | RefSeq: 243

Show only records from: CoreNucleotide (1947), EST (22), GSS (0). [What's this?]

Items 1 - 20 of 1969

Page 1 of 99

☐ 1: NM_000655
Homo sapiens selectin L (lymphocyte adhesion molecule 1) (SELL), mRNA
gi|115430250|ref|NM_000655.3|[115430250]

☐ 2: CS417899
Sequence 32 from Patent WO2006048776
gi|115420788|emb|CS417899.1||pat|WO|2006048776|32[115420788]

☐ 3: CS417898
Sequence 31 from Patent WO2006048776
gi|115420787|emb|CS417898.1||pat|WO|2006048776|31[115420787]

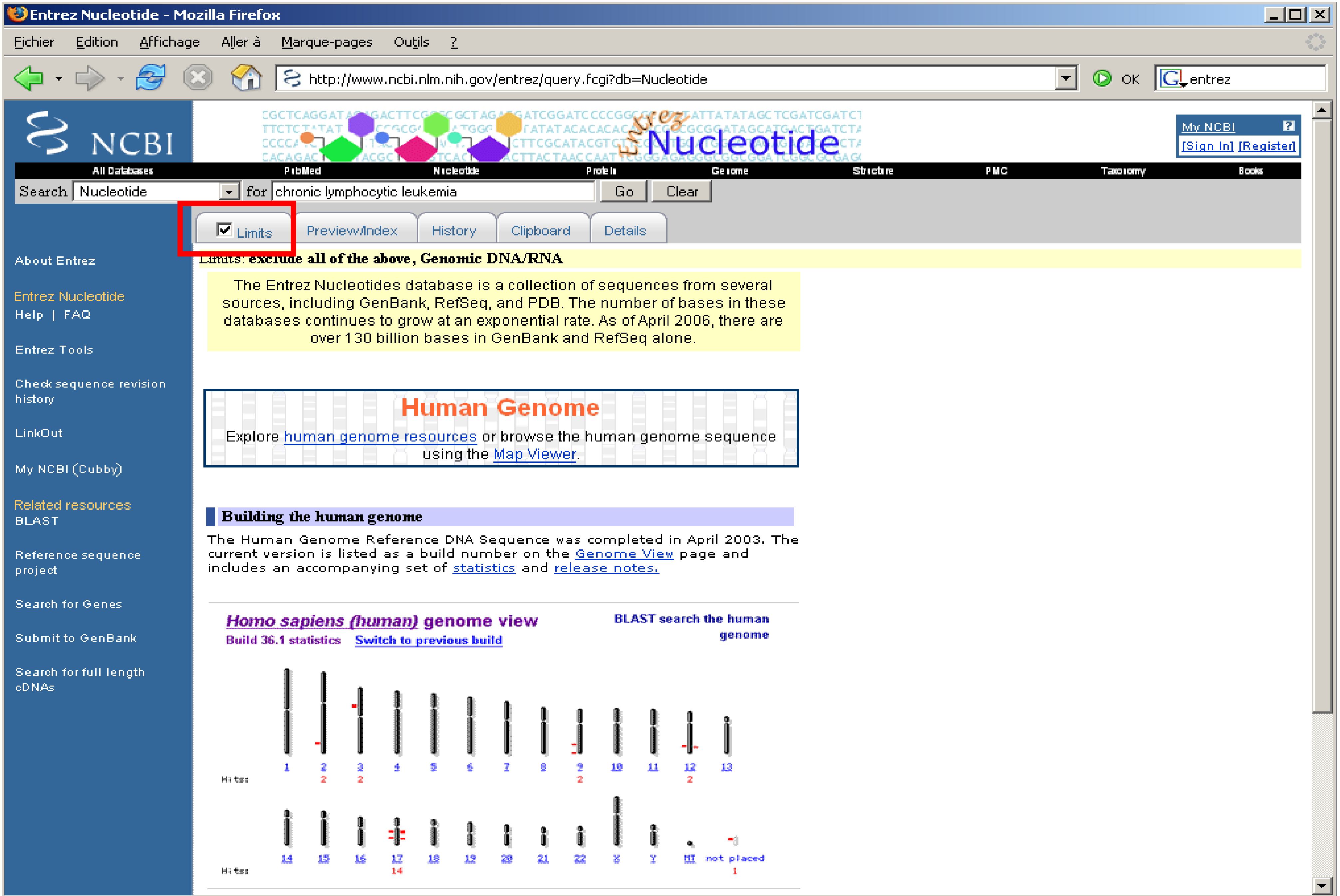
☐ 4: CS417897
Sequence 30 from Patent WO2006048776
gi|115420786|emb|CS417897.1||pat|WO|2006048776|30[115420786]

☐ 5: CS417896
Sequence 29 from Patent WO2006048776
gi|115420785|emb|CS417896.1||pat|WO|2006048776|29[115420785]

☐ 6: CS417894
Sequence 27 from Patent WO2006048776
gi|115420783|emb|CS417894.1||pat|WO|2006048776|27[115420783]

☐ 7: CS417892
Sequence 25 from Patent WO2006048776
gi|115420781|emb|CS417892.1||pat|WO|2006048776|25[115420781]

Affiner ses recherches par l'onglet limits



http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide

entrez

Search Nucleotide for chronic lymphocytic leukemia

Limits: exclude all of the above, Genomic DNA/RNA

The Entrez Nucleotides database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. The number of bases in these databases continues to grow at an exponential rate. As of April 2006, there are over 130 billion bases in GenBank and RefSeq alone.

Human Genome

Explore human genome resources or browse the human genome sequence using the Map Viewer.

Building the human genome

The Human Genome Reference DNA Sequence was completed in April 2003. The current version is listed as a build number on the Genome View page and includes an accompanying set of statistics and release notes.

Homo sapiens (human) **genome view**

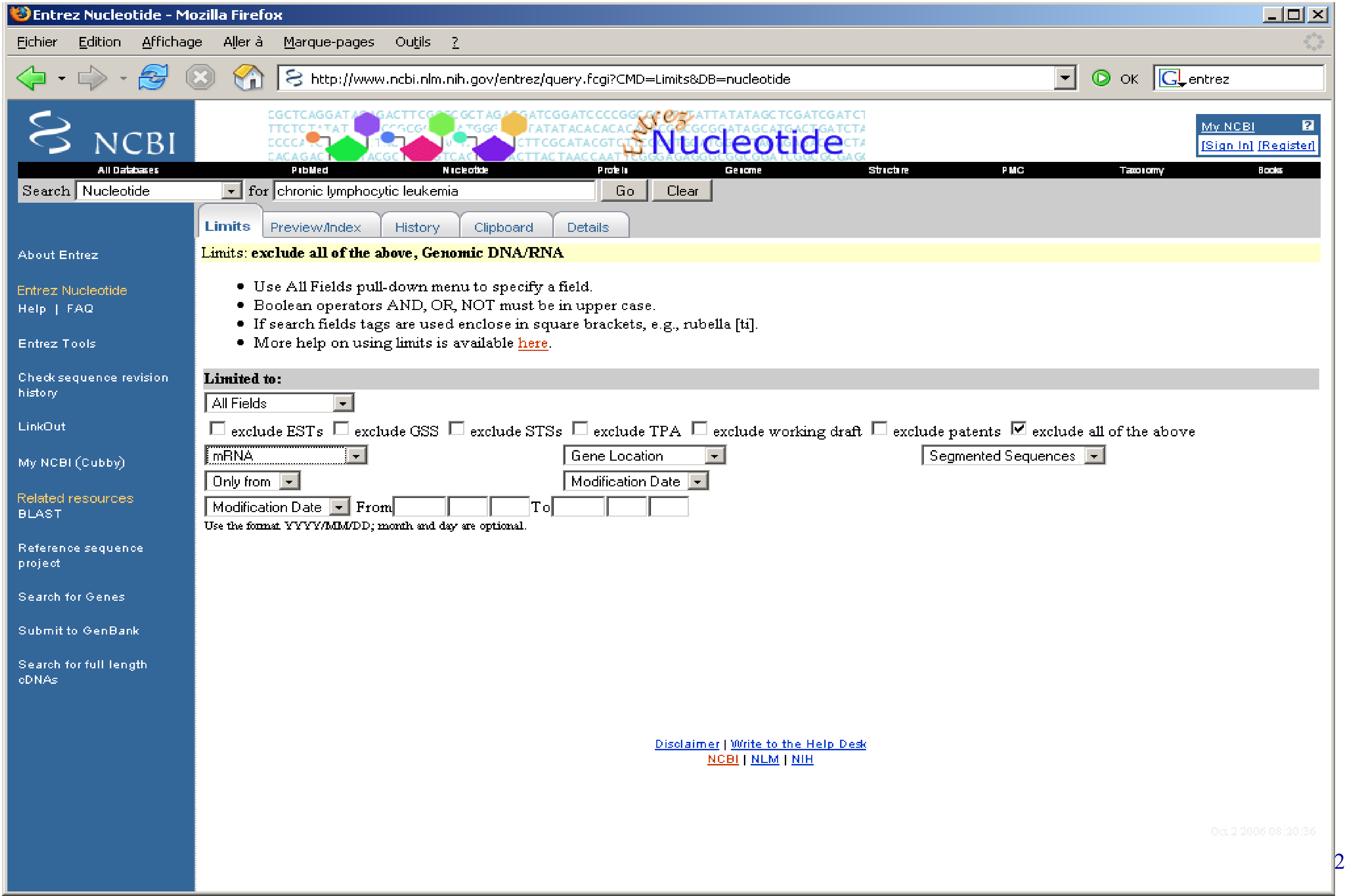
Build 36.1 statistics Switch to previous build

BLAST search the human genome

	1	2	3	4	5	6	7	8	9	10	11	12	13
Hits:		2	2						2			2	

	14	15	16	17	18	19	20	21	22	X	Y	MT	not placed
Hits:				14									1

Affiner ses recherches par l'onglet "Limits"   Procéder par étapes: 1) "chronic lymphocytic leukemia » dans tous les champs



http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Limits&DB=nucleotide

Search: Nucleotide for: chronic lymphocytic leukemia

Limits: **exclude all of the above, Genomic DNA/RNA**

- Use All Fields pull-down menu to specify a field.
- Boolean operators AND, OR, NOT must be in upper case.
- If search fields tags are used enclose in square brackets, e.g., rubella [ti].
- More help on using limits is available here.

Limited to:

All Fields

☐ exclude ESTs ☐ exclude GSS ☐ exclude STSs ☐ exclude TPA ☐ exclude working draft ☐ exclude patents ☑ exclude all of the above

mRNA | Gene Location | Segmented Sequences

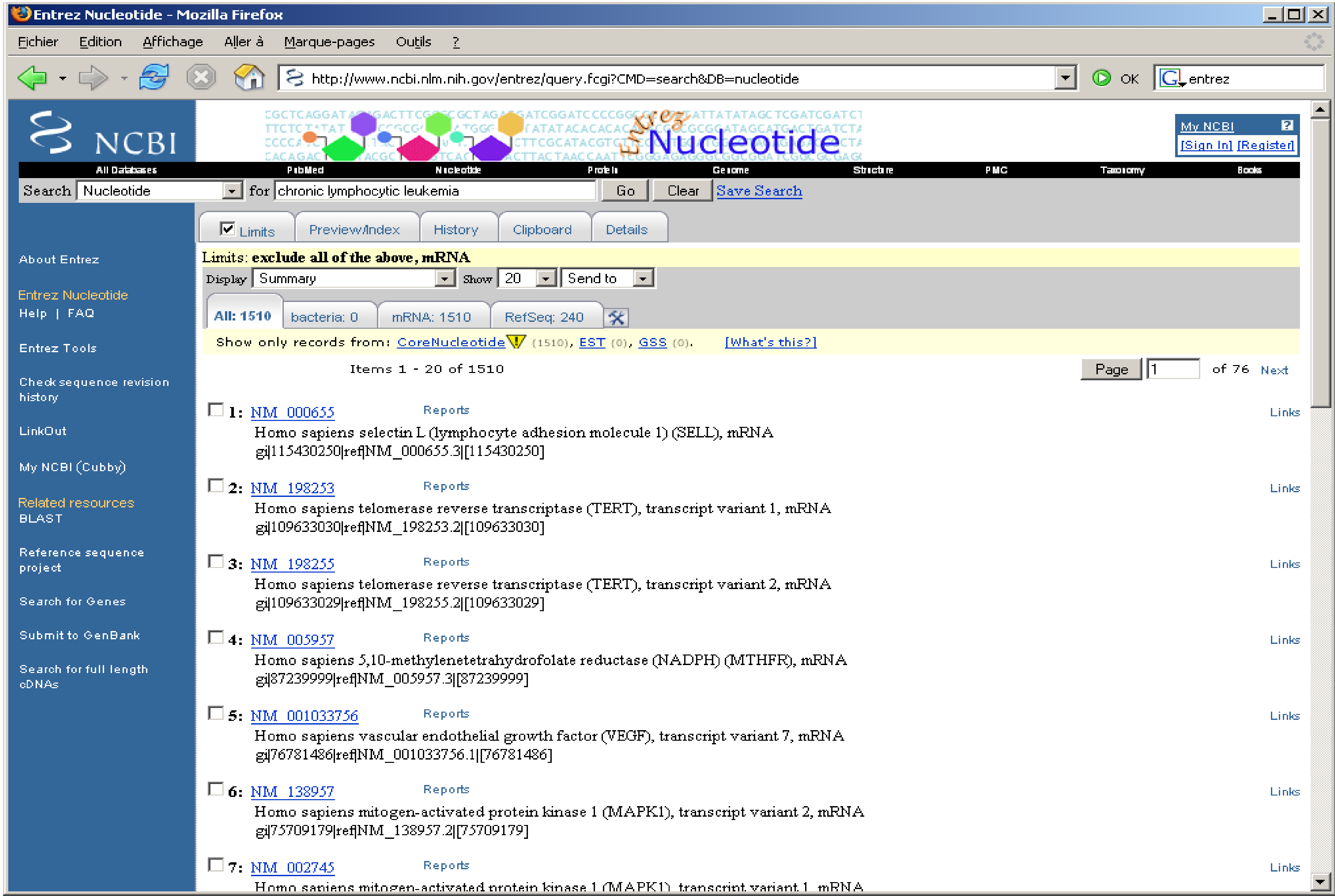
Only from | Modification Date

Modification Date From [ ] [ ] [ ] To [ ] [ ] [ ]

Use the format YYYY/MM/DD; month and day are optional.

Oct 2 2006 08:20:36

Affiner ses recherches par l'onglet limits



http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=nucleotide

Search Nucleotide for chronic lymphocytic leukemia

Limits: **exclude all of the above, mRNA**

Display Summary Show 20 Send to

All: 1510 | bacteria: 0 | mRNA: 1510 | RefSeq: 240

Show only records from: CoreNucleotide (1510), EST (0), GSS (0). [What's this?]

Items 1 - 20 of 1510 Page 1 of 76

☐ **1:** NM_000655 Reports Links
Homo sapiens selectin L (lymphocyte adhesion molecule 1) (SELL), mRNA
gi|115430250|ref|NM_000655.3|[115430250]

☐ **2:** NM_198253 Reports Links
Homo sapiens telomerase reverse transcriptase (TERT), transcript variant 1, mRNA
gi|109633030|ref|NM_198253.2|[109633030]

☐ **3:** NM_198255 Reports Links
Homo sapiens telomerase reverse transcriptase (TERT), transcript variant 2, mRNA
gi|109633029|ref|NM_198255.2|[109633029]

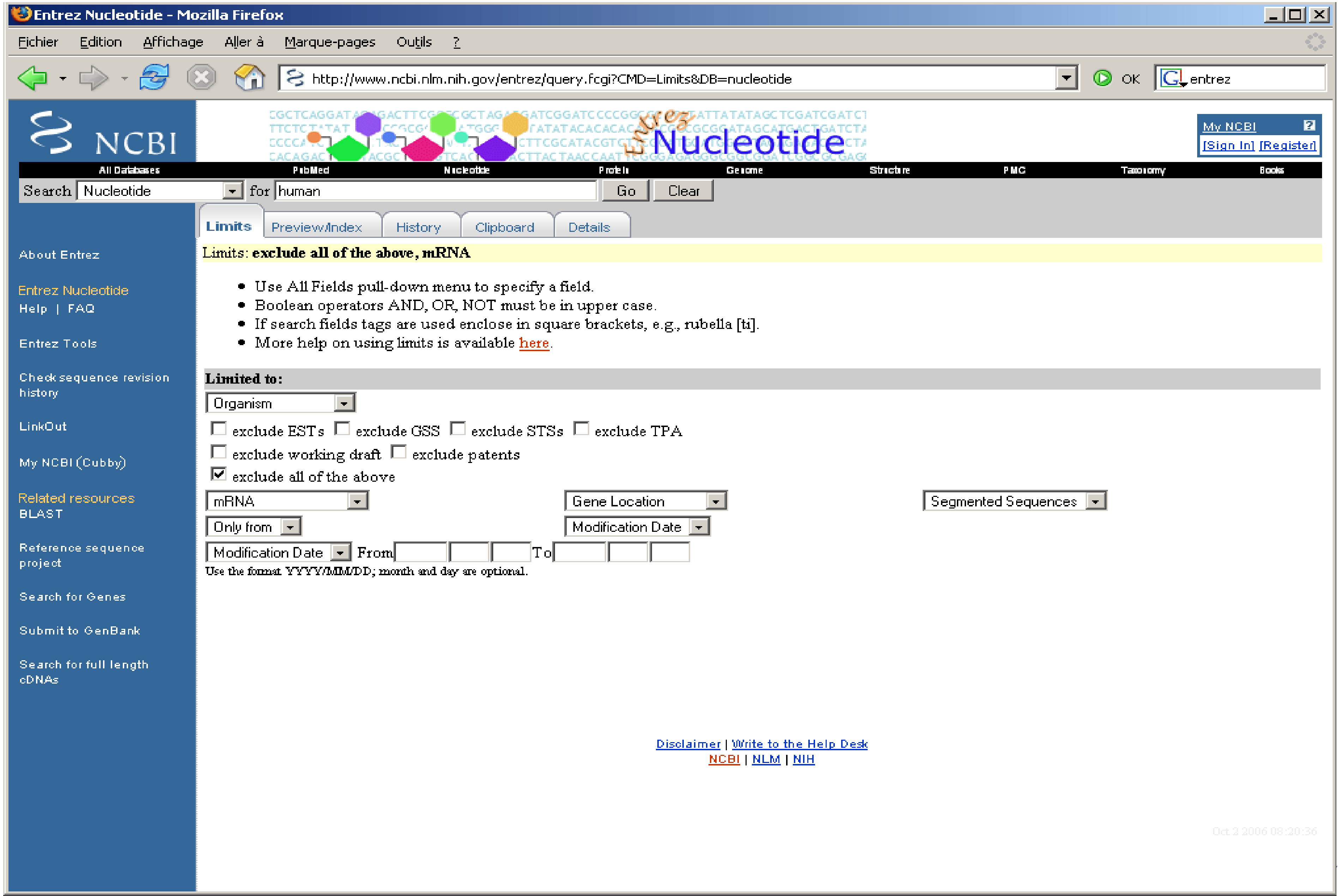
☐ **4:** NM_005957 Reports Links
Homo sapiens 5,10-methylenetetrahydrofolate reductase (NADPH) (MTHFR), mRNA
gi|87239999|ref|NM_005957.3|[87239999]

☐ **5:** NM_001033756 Reports Links
Homo sapiens vascular endothelial growth factor (VEGF), transcript variant 7, mRNA
gi|76781486|ref|NM_001033756.1|[76781486]

☐ **6:** NM_138957 Reports Links
Homo sapiens mitogen-activated protein kinase 1 (MAPK1), transcript variant 2, mRNA
gi|75709179|ref|NM_138957.2|[75709179]

☐ **7:** NM_002745 Reports Links
Homo sapiens mitogen-activated protein kinase 1 (MAPK1), transcript variant 1, mRNA

Affiner ses recherches par l’onglet “Limits” Procéder par étapes: 2) restreindre la recherche aux séquences ARNm humaines



http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Limits&DB=nucleotide

Search Nucleotide for human

Limits: **exclude all of the above, mRNA**

- Use All Fields pull-down menu to specify a field.
- Boolean operators AND, OR, NOT must be in upper case.
- If search fields tags are used enclose in square brackets, e.g., rubella [ti].
- More help on using limits is available here.

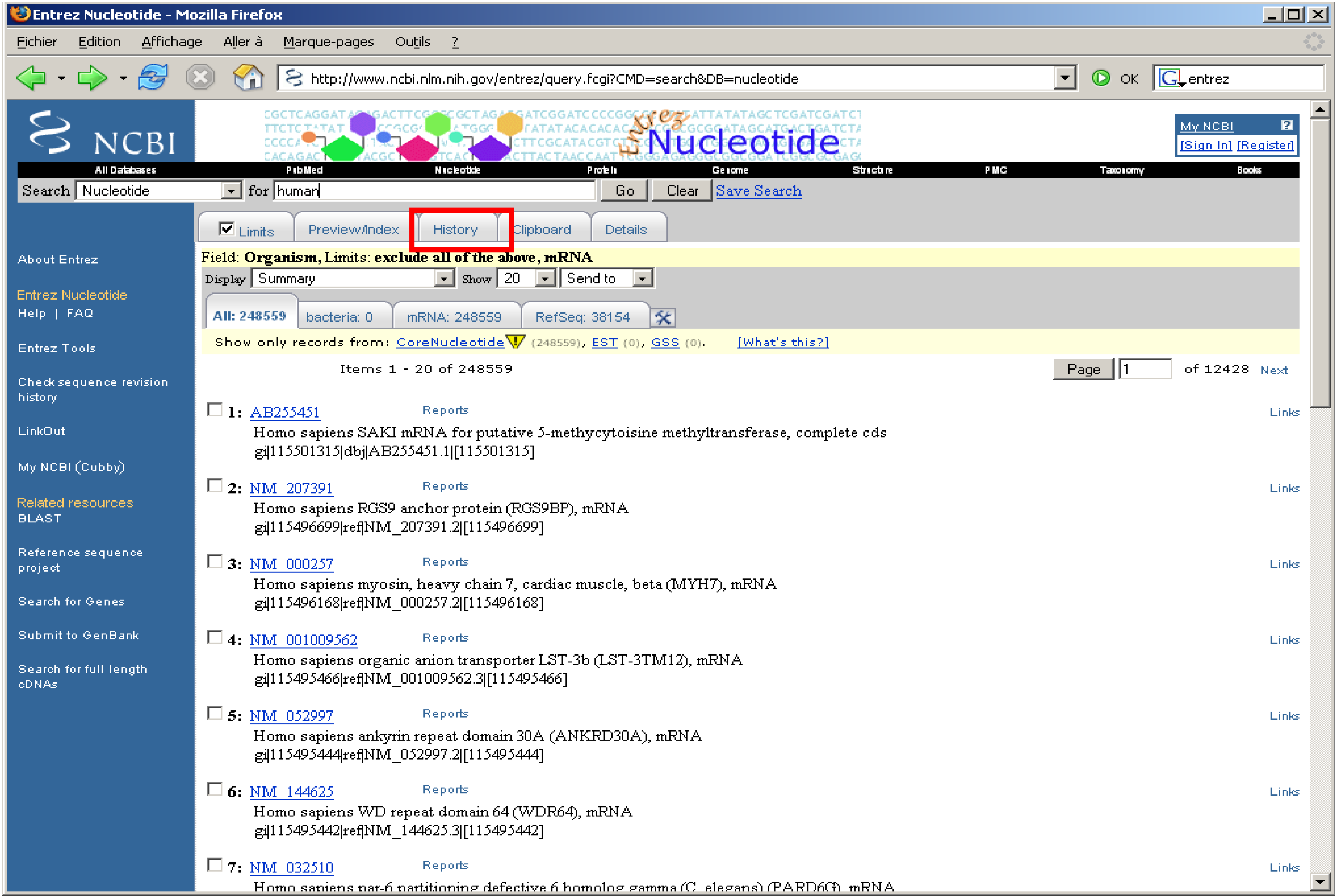
Limited to:

Organism

☐ exclude ESTs ☐ exclude GSS ☐ exclude STSs ☐ exclude TPA
☐ exclude working draft ☐ exclude patents
☑ exclude all of the above

mRNA Gene Location Segmented Sequences
Only from Modification Date
Modification Date From To

Use the format YYYY/MM/DD; month and day are optional.



Search Nucleotide for human

Field: **Organism**, Limits: **exclude all of the above, mRNA**

Display Summary Show 20 Send to

All: 248559 | bacteria: 0 | mRNA: 248559 | RefSeq: 38154

Show only records from: CoreNucleotide (248559), EST (0), GSS (0). [What's this?]

Items 1 - 20 of 248559 Page 1 of 12428

☐ **1:** AB255451 Reports Links
Homo sapiens SAKI mRNA for putative 5-methycytoisine methyltransferase, complete cds
gi|115501315|dbj|AB255451.1|[115501315]

☐ **2:** NM_207391 Reports Links
Homo sapiens RGS9 anchor protein (RGS9BP), mRNA
gi|115496699|ref|NM_207391.2|[115496699]

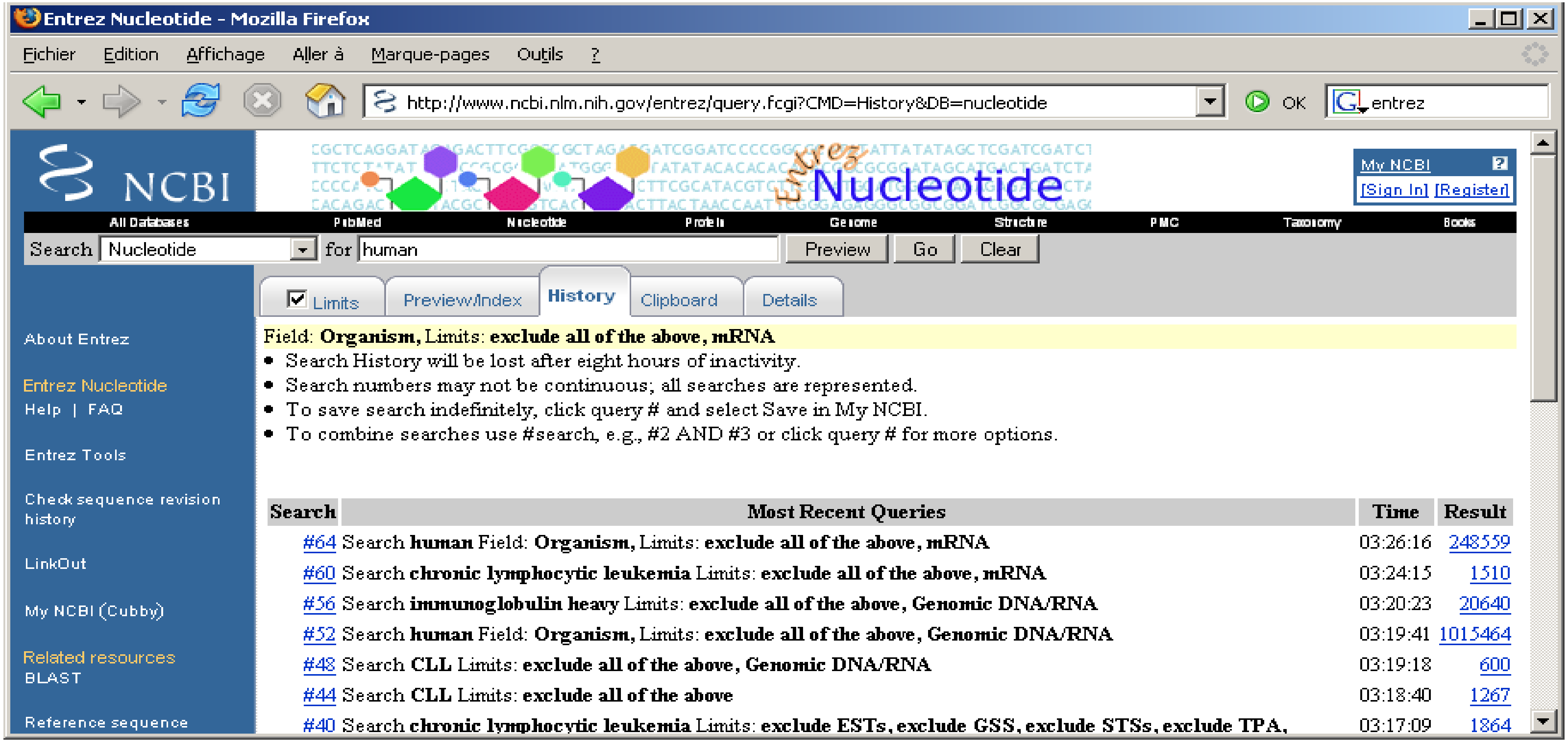
☐ **3:** NM_000257 Reports Links
Homo sapiens myosin, heavy chain 7, cardiac muscle, beta (MYH7), mRNA
gi|115496168|ref|NM_000257.2|[115496168]

☐ **4:** NM_001009562 Reports Links
Homo sapiens organic anion transporter LST-3b (LST-3TM12), mRNA
gi|115495466|ref|NM_001009562.3|[115495466]

☐ **5:** NM_052997 Reports Links
Homo sapiens ankyrin repeat domain 30A (ANKRD30A), mRNA
gi|115495444|ref|NM_052997.2|[115495444]

☐ **6:** NM_144625 Reports Links
Homo sapiens WD repeat domain 64 (WDR64), mRNA
gi|115495442|ref|NM_144625.3|[115495442]

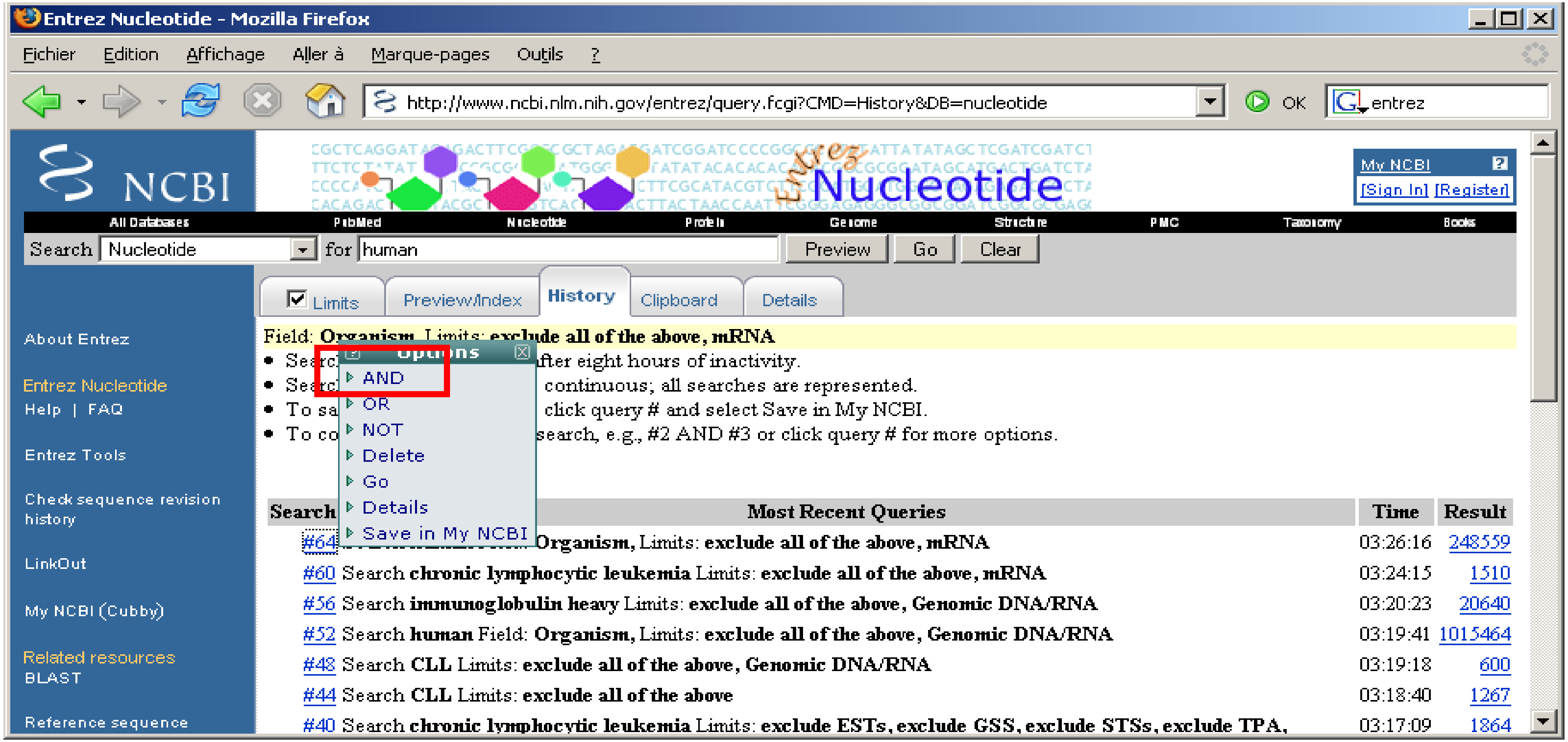
☐ **7:** NM_032510 Reports Links
Homo sapiens par-6 partitioning defective 6 homolog gamma (C. elegans) (PARD6G), mRNA



Field: **Organism,** Limits: **exclude all of the above, mRNA**

- Search History will be lost after eight hours of inactivity.
- Search numbers may not be continuous; all searches are represented.
- To save search indefinitely, click query # and select Save in My NCBI.
- To combine searches use #search, e.g., #2 AND #3 or click query # for more options.

Search	Most Recent Queries	Time	Result
#64	Search **human** Field: **Organism,** Limits: **exclude all of the above, mRNA**	03:26:16	248559
#60	Search **chronic lymphocytic leukemia** Limits: **exclude all of the above, mRNA**	03:24:15	1510
#56	Search **immunoglobulin heavy** Limits: **exclude all of the above, Genomic DNA/RNA**	03:20:23	20640
#52	Search **human** Field: **Organism,** Limits: **exclude all of the above, Genomic DNA/RNA**	03:19:41	1015464
#48	Search **CLL** Limits: **exclude all of the above, Genomic DNA/RNA**	03:19:18	600
#44	Search **CLL** Limits: **exclude all of the above**	03:18:40	1267
#40	Search **chronic lymphocytic leukemia** Limits: **exclude ESTs, exclude GSS, exclude STSs, exclude TPA,**	03:17:09	1864



http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=History&DB=nucleotide

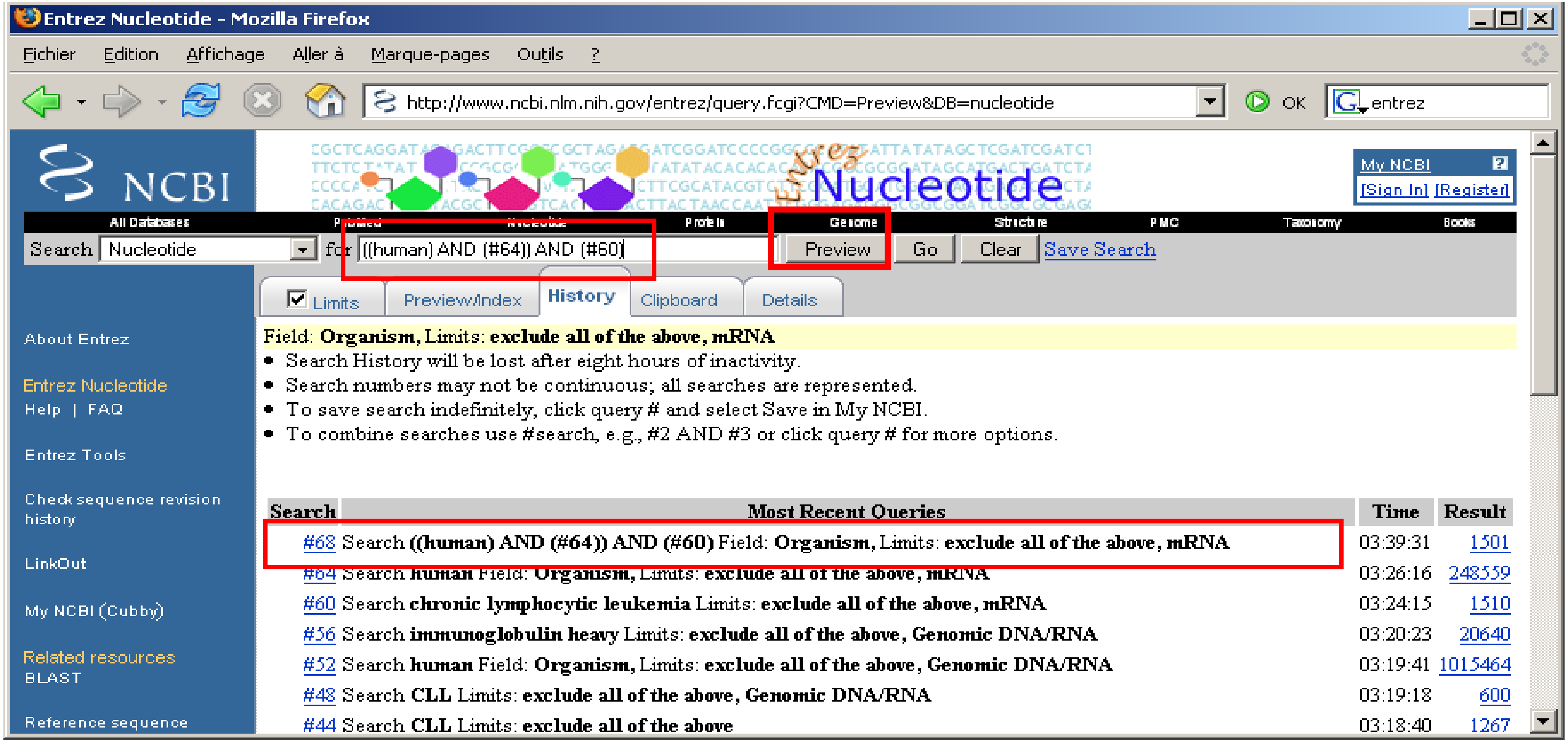
Search: Nucleotide for: human

☑ Limits

Field: **Organism**, Limits: **exclude all of the above, mRNA**

- Searc after eight hours of inactivity.
- Searc continuous; all searches are represented.
- To sa click query # and select Save in My NCBI.
- To co search, e.g., #2 AND #3 or click query # for more options.

Search	Most Recent Queries	Time	Result
#64	**Organism,** Limits: **exclude all of the above, mRNA**	03:26:16	248559
#60	Search **chronic lymphocytic leukemia** Limits: **exclude all of the above, mRNA**	03:24:15	1510
#56	Search **immunoglobulin heavy** Limits: **exclude all of the above, Genomic DNA/RNA**	03:20:23	20640
#52	Search **human** Field: **Organism,** Limits: **exclude all of the above, Genomic DNA/RNA**	03:19:41	1015464
#48	Search **CLL** Limits: **exclude all of the above, Genomic DNA/RNA**	03:19:18	600
#44	Search **CLL** Limits: **exclude all of the above**	03:18:40	1267
#40	Search **chronic lymphocytic leukemia** Limits: **exclude ESTs, exclude GSS, exclude STSs, exclude TPA,**	03:17:09	1864



Search Nucleotide for ((human) AND (#64)) AND (#60)

Field: **Organism,** Limits: **exclude all of the above, mRNA**

- Search History will be lost after eight hours of inactivity.
- Search numbers may not be continuous; all searches are represented.
- To save search indefinitely, click query # and select Save in My NCBI.
- To combine searches use #search, e.g., #2 AND #3 or click query # for more options.

Search	Most Recent Queries	Time	Result
#68	Search **((human) AND (#64)) AND (#60)** Field: **Organism,** Limits: **exclude all of the above, mRNA**	03:39:31	1501
#64	Search **human** Field: **Organism,** Limits: **exclude all of the above, mRNA**	03:26:16	248559
#60	Search **chronic lymphocytic leukemia** Limits: **exclude all of the above, mRNA**	03:24:15	1510
#56	Search **immunoglobulin heavy** Limits: **exclude all of the above, Genomic DNA/RNA**	03:20:23	20640
#52	Search **human** Field: **Organism,** Limits: **exclude all of the above, Genomic DNA/RNA**	03:19:41	1015464
#48	Search **CLL** Limits: **exclude all of the above, Genomic DNA/RNA**	03:19:18	600
#44	Search **CLL** Limits: **exclude all of the above**	03:18:40	1267

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=nucleotide

entrez

Search Nucleotide for ((human) AND (#64)) AND (#60)

Field: **Organism,** Limits: **exclude all of the above, mRNA**

Display Summary Show 20 Send to

All: 1501 bacteria: 0 mRNA: 1501 RefSeq: 231

Show only records from: CoreNucleotide (1501), EST (0), GSS (0). [What's this?]

Items 1 - 20 of 1501

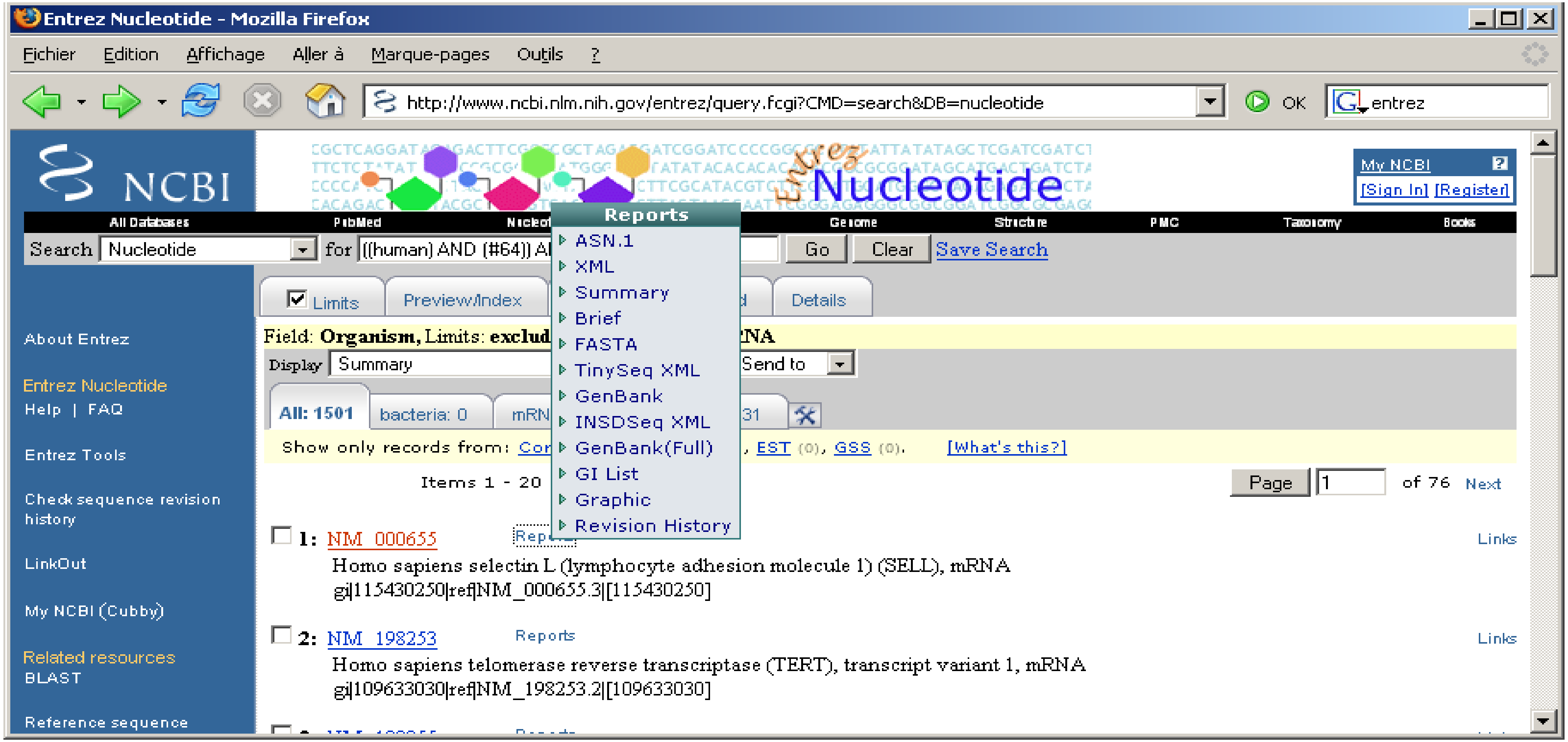
Page 1 of 76

☐ **1:** NM_000655

Homo sapiens selectin L (lymphocyte adhesion molecule 1) (SELL), mRNA
gi|115430250|ref|NM_000655.3|[115430250]

☐ **2:** NM_198253

Homo sapiens telomerase reverse transcriptase (TERT), transcript variant 1, mRNA
gi|109633030|ref|NM_198253.2|[109633030]



http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=nucleotide

entrez

Search Nucleotide for [(human) AND (#64)] Al

Field: **Organism,** Limits: **exclud** **NA**

Display Summary

Show only records from: Cor , EST (0), GSS (0).

Items 1 - 20

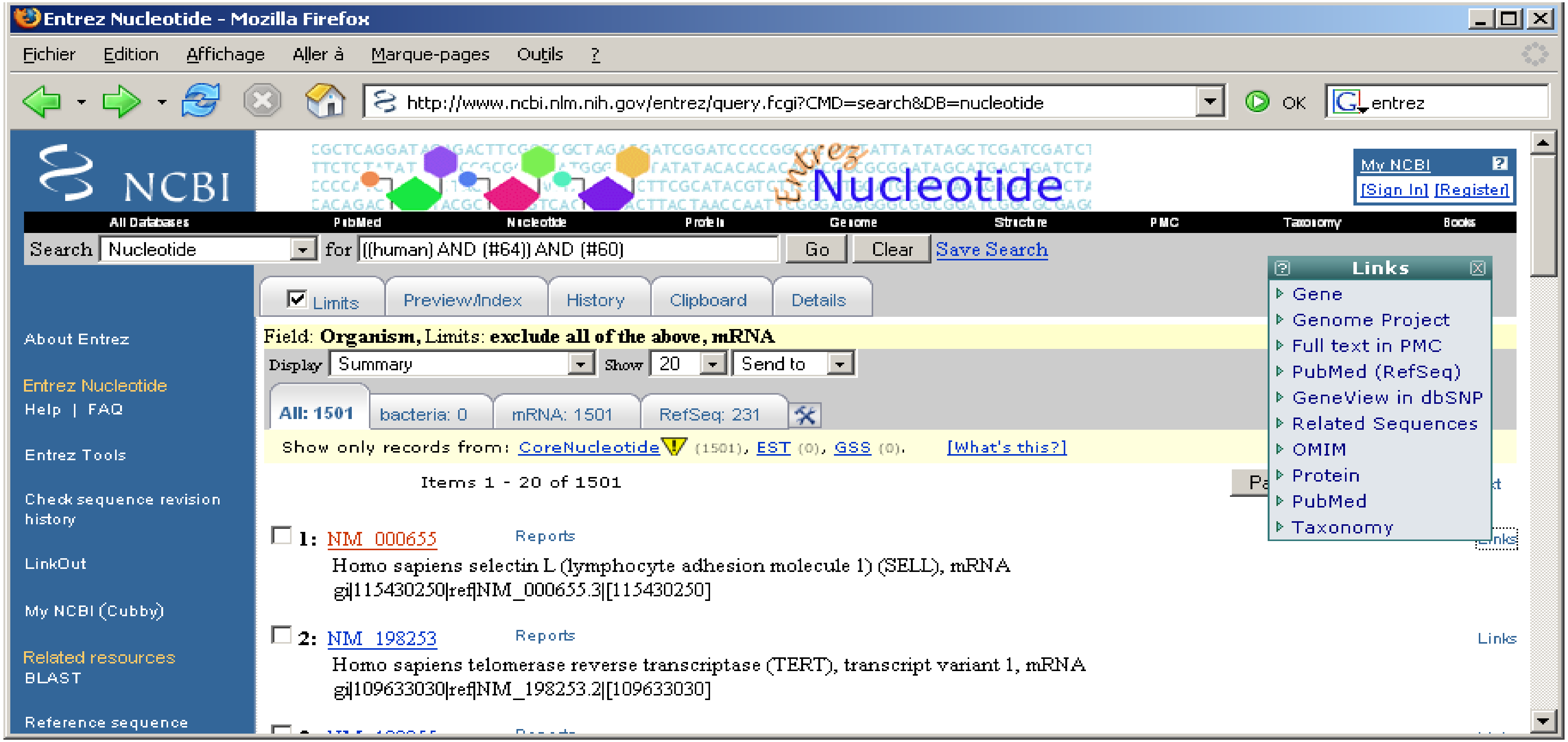
Page 1 of 76

☐ **1:** NM_000655 Reports

Homo sapiens selectin L (lymphocyte adhesion molecule 1) (SELL), mRNA
gi|115430250|ref|NM_000655.3|[115430250]

☐ **2:** NM_198253 Reports

Homo sapiens telomerase reverse transcriptase (TERT), transcript variant 1, mRNA
gi|109633030|ref|NM_198253.2|[109633030]



Search Nucleotide for ((human) AND (#64)) AND (#60)

☑ Limits

Field: **Organism,** Limits: **exclude all of the above, mRNA**

Display Summary Show 20 Send to

All: 1501 bacteria: 0 mRNA: 1501 RefSeq: 231

Show only records from: CoreNucleotide (1501), EST (0), GSS (0). [What's this?]

Items 1 - 20 of 1501

☐ 1: NM_000655 Reports

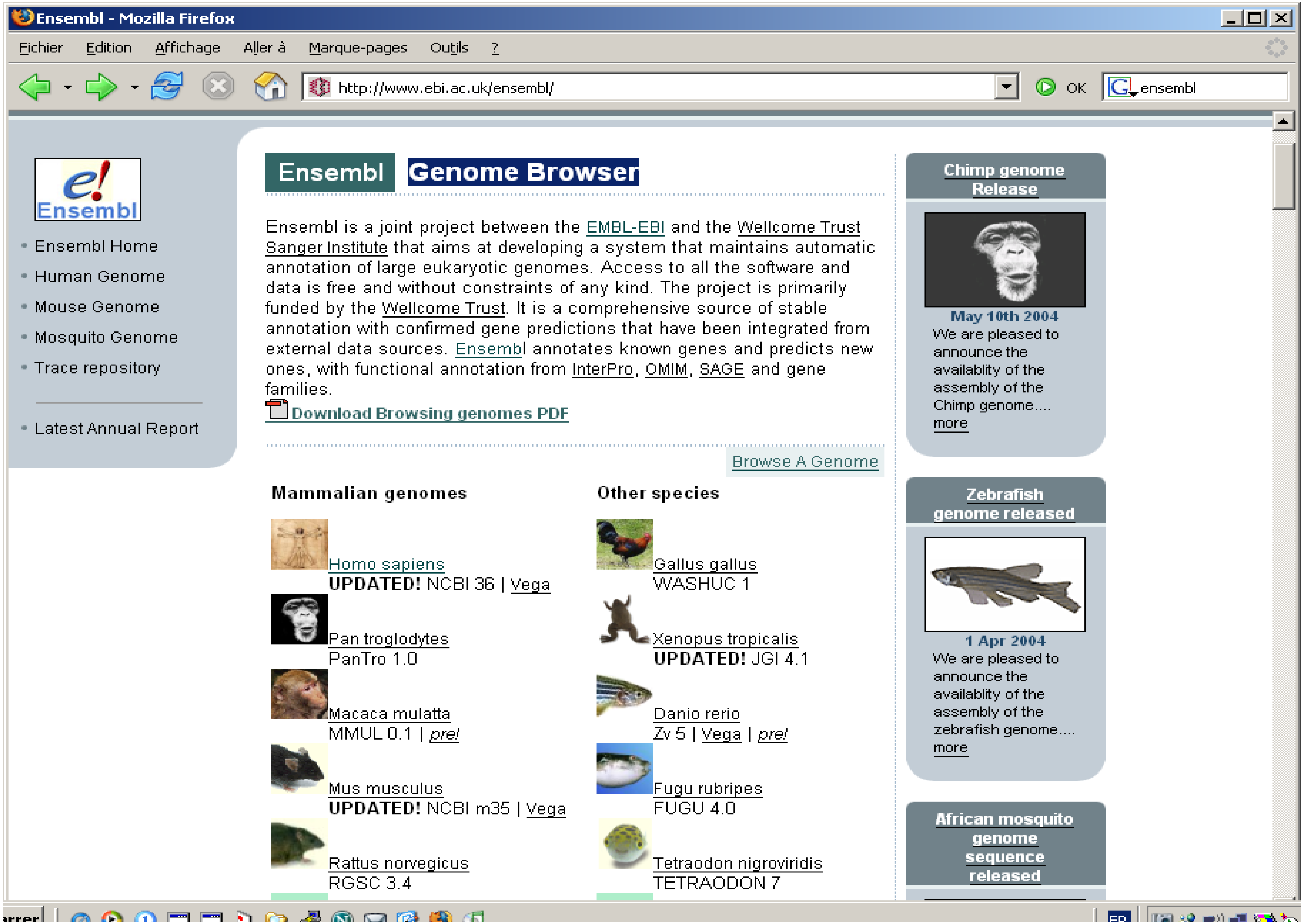
Homo sapiens selectin L (lymphocyte adhesion molecule 1) (SELL), mRNA
gi|115430250|ref|NM_000655.3|[115430250]

☐ 2: NM_198253 Reports Links

Homo sapiens telomerase reverse transcriptase (TERT), transcript variant 1, mRNA
gi|109633030|ref|NM_198253.2|[109633030]

Ensembl Genome Browser

http://www.ebi.ac.uk/ensembl



- Ensembl Home
- Human Genome
- Mouse Genome
- Mosquito Genome
- Trace repository

- Latest Annual Report

Ensembl Genome Browser

Ensembl is a joint project between the EMBL-EBI and the Wellcome Trust Sanger Institute that aims at developing a system that maintains automatic annotation of large eukaryotic genomes. Access to all the software and data is free and without constraints of any kind. The project is primarily funded by the Wellcome Trust. It is a comprehensive source of stable annotation with confirmed gene predictions that have been integrated from external data sources. Ensembl annotates known genes and predicts new ones, with functional annotation from InterPro, OMIM, SAGE and gene families.

Download Browsing genomes PDF

Browse A Genome

Mammalian genomes

Homo sapiens
UPDATED! NCBI 36 | Vega

Pan troglodytes
PanTro 1.0

Macaca mulatta
MMUL 0.1 | *pre!*

Mus musculus
UPDATED! NCBI m35 | Vega

Rattus norvegicus
RGSC 3.4

Other species

Gallus gallus
WASHUC 1

Xenopus tropicalis
UPDATED! JGI 4.1

Danio rerio
Zv 5 | Vega | *pre!*

Fugu rubripes
FUGU 4.0

Tetraodon nigroviridis
TETRAODON 7

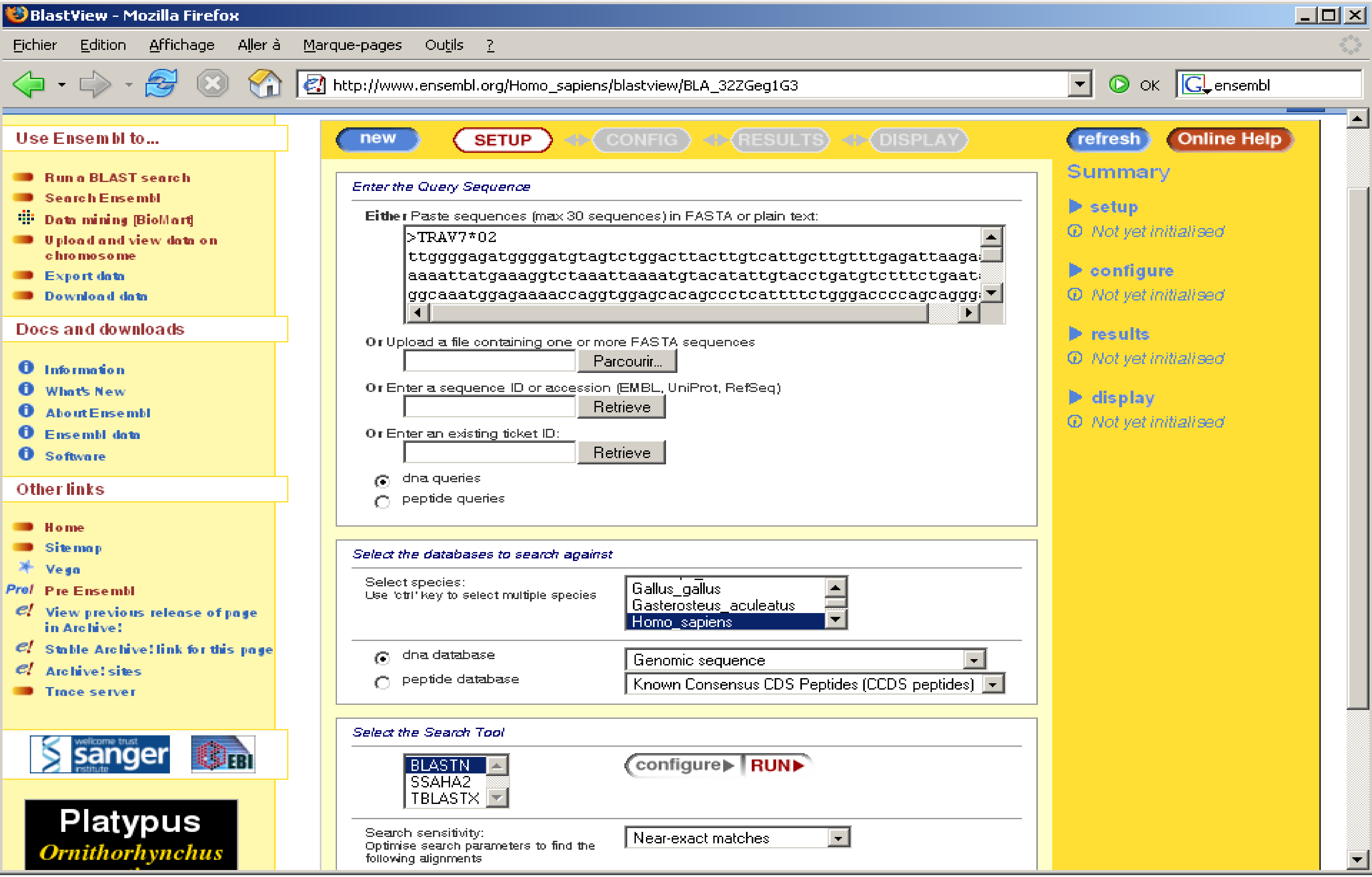
Chimp genome Release

May 10th 2004
We are pleased to announce the availablity of the assembly of the Chimp genome....
more

Zebrafish genome released

1 Apr 2004
We are pleased to announce the availablity of the assembly of the zebrafish genome....
more

African mosquito genome sequence released



http://www.ensembl.org/Homo_sapiens/blastview/BLA_32ZGeg1G3

Enter the Query Sequence

Either Paste sequences (max 30 sequences) in FASTA or plain text:

```
>TRAV7*02
ttggggagatggggatgtagtctggacttacttgtcattgcttgtttgagattaaga
aaaattatgaaaggtctaaattaaaatgtacatattgtacctgatgtcttttctgaat
ggcaaatggagaaaaccaggtggagcacagccctcattttctgggaccccagcaggg
```

Or Upload a file containing one or more FASTA sequences

Or Enter a sequence ID or accession (EMBL, UniProt, RefSeq)

Or Enter an existing ticket ID:

● dna queries

○ peptide queries

Select the databases to search against

Select species:
Use 'ctrl' key to select multiple species

Gallus_gallus
Gasterosteus_aculeatus
Homo_sapiens

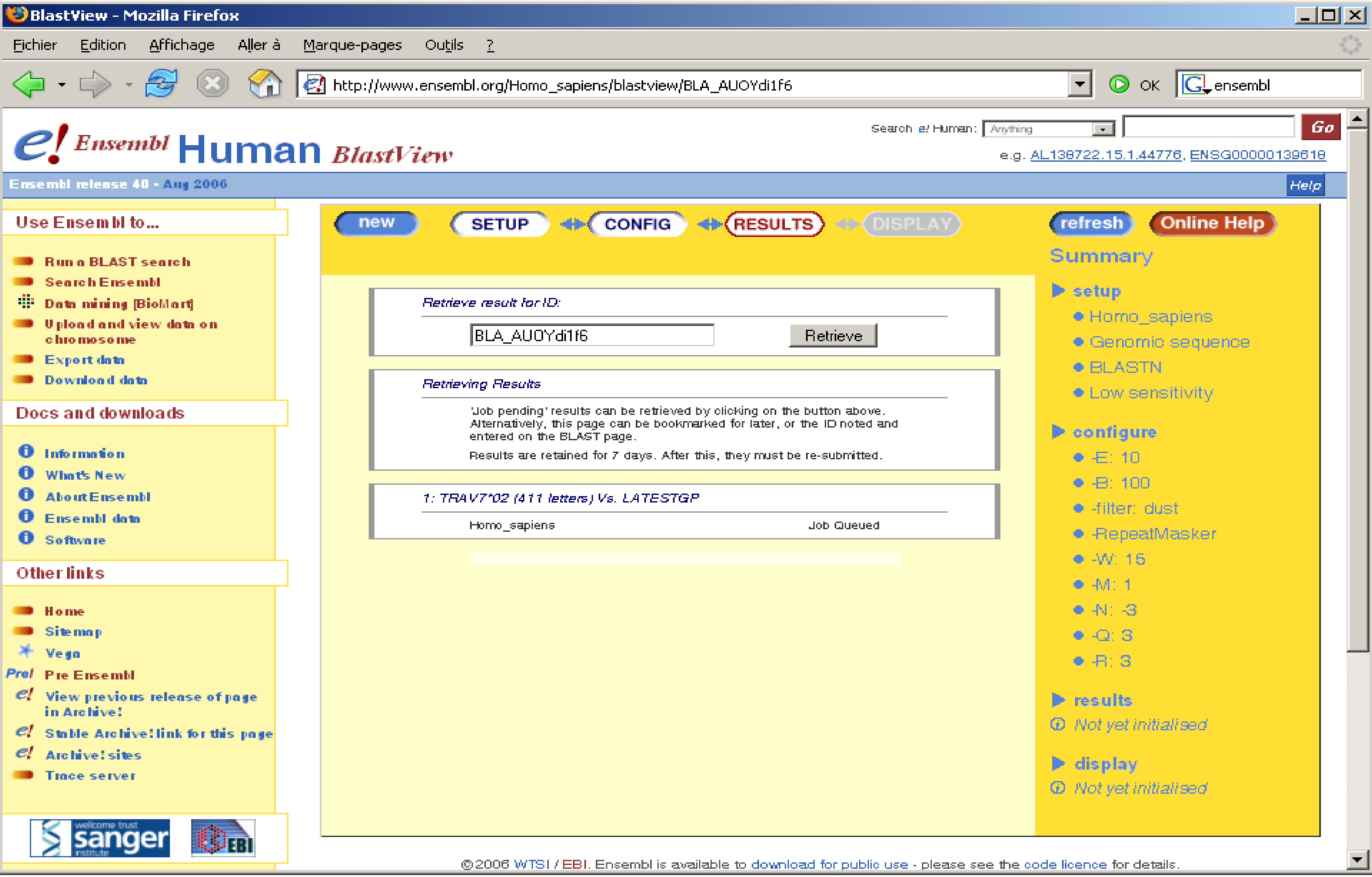
● dna database — Genomic sequence

○ peptide database — Known Consensus CDS Peptides (CCDS peptides)

Select the Search Tool

BLASTN
SSAHA2
TBLASTX

Search sensitivity:
Optimise search parameters to find the following alignments — Near-exact matches



http://www.ensembl.org/Homo_sapiens/blastview/BLA_AUOYdi1f6

e.g. AL139722.15.1.44776, ENSG00000139618

Ensembl release 40 - Aug 2006

Retrieve result for ID:

BLA_AUOYdi1f6

Retrieving Results

'Job pending' results can be retrieved by clicking on the button above. Alternatively, this page can be bookmarked for later, or the ID noted and entered on the BLAST page.

Results are retained for 7 days. After this, they must be re-submitted.

*1: TRAV7*02 (411 letters) Vs. LATESTGP*

Homo_sapiens Job Queued

Summary

setup
- Homo_sapiens
- Genomic sequence
- BLASTN
- Low sensitivity

configure
- -E: 10
- -B: 100
- -filter: dust
- -RepeatMasker
- -W: 15
- -M: 1
- -N: -3
- -Q: 3
- -R: 3

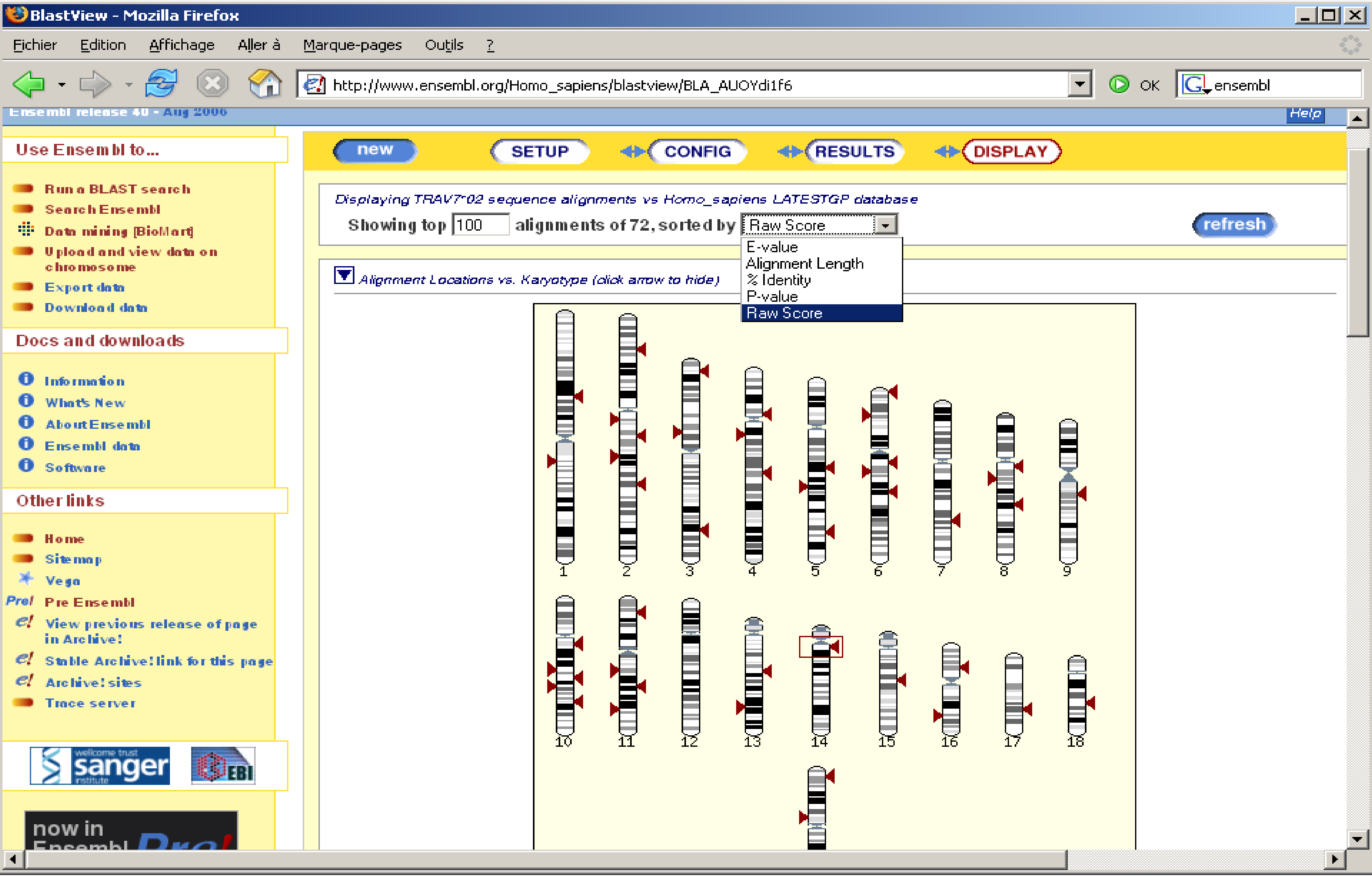
results

Not yet initialised

display

Not yet initialised

© 2006 WTSI / EBI. Ensembl is available to download for public use - please see the code licence for details.

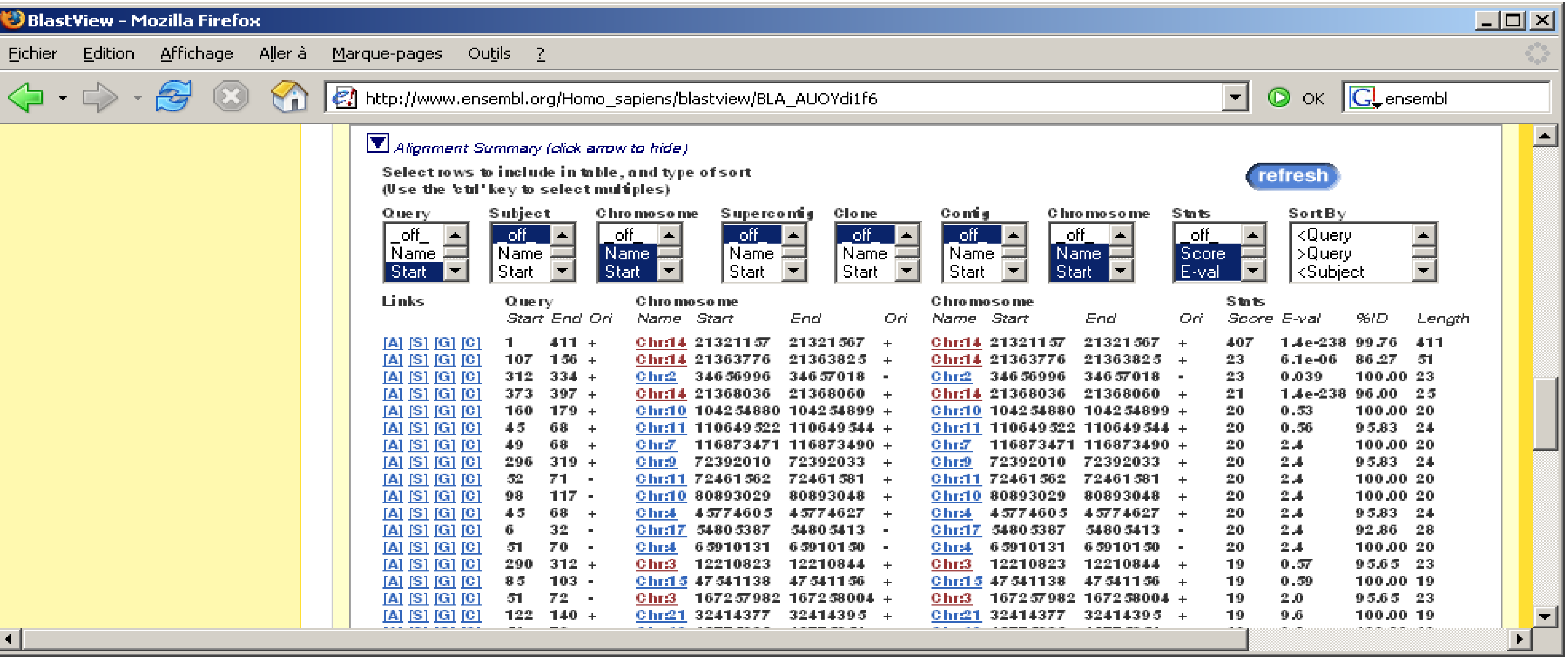


http://www.ensembl.org/Homo_sapiens/blastview/BLA_AUOYdi1f6

Displaying TRAV7-02 sequence alignments vs Homo_sapiens LATESTGP database

Showing top 100 alignments of 72, sorted by Raw Score

Alignment Locations vs. Karyotype (click arrow to hide)

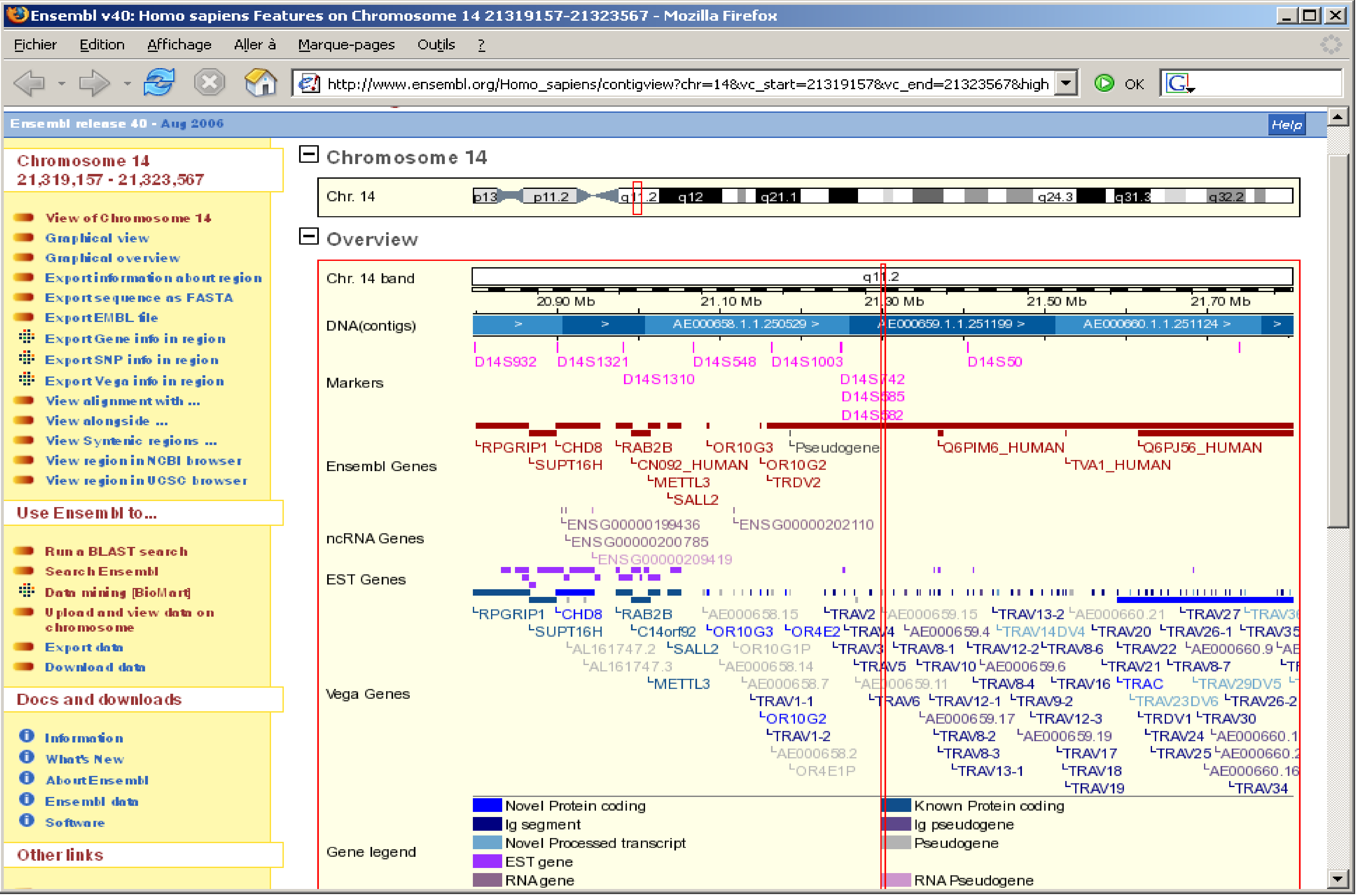


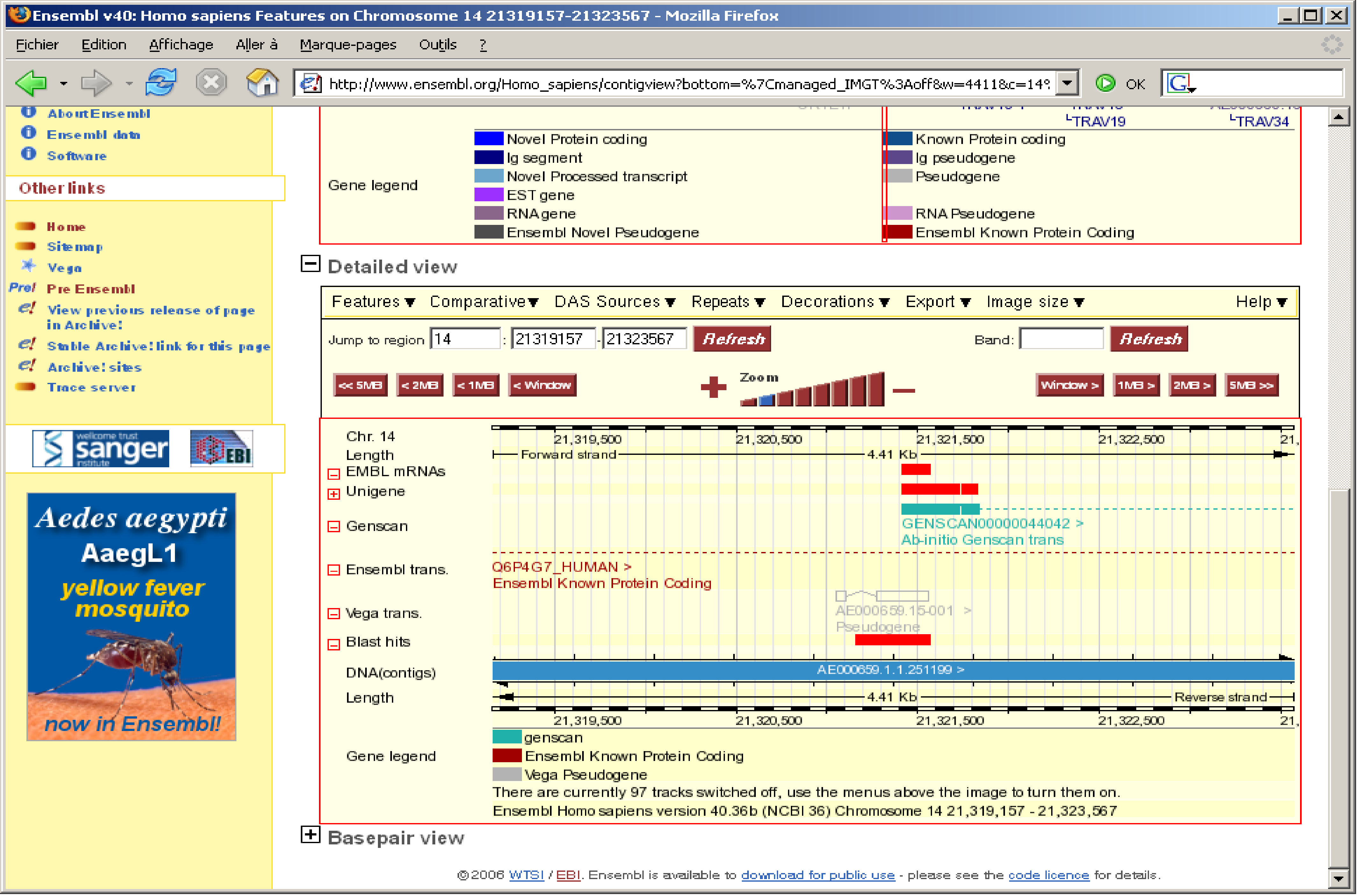
Alignment Summary (click arrow to hide)

Select rows to include in table, and type of sort
(Use the 'ctrl' key to select multiples)

Query: _off_, Name, Start
Subject: _off_, Name, Start
Chromosome: _off_, Name, Start
Supercontig: _off_, Name, Start
Clone: _off_, Name, Start
Contig: _off_, Name, Start
Chromosome: _off_, Name, Start
Stats: _off_, Score, E-val
Sort By: <Query, >Query, <Subject

Links	Query			Chromosome				Chromosome				Stats			
	Start	*End*	*Ori*	*Name*	*Start*	*End*	*Ori*	*Name*	*Start*	*End*	*Ori*	*Score*	*E-val*	*%ID*	*Length*
[A] [S] [G] [C]	1	411	+	Chr:14	21321157	21321567	+	Chr:14	21321157	21321567	+	407	1.4e-238	99.76	411
[A] [S] [G] [C]	107	156	+	Chr:14	21363776	21363825	+	Chr:14	21363776	21363825	+	23	6.1e-06	86.27	51
[A] [S] [G] [C]	312	334	+	Chr:2	34656996	34657018	-	Chr:2	34656996	34657018	-	23	0.039	100.00	23
[A] [S] [G] [C]	373	397	+	Chr:14	21368036	21368060	+	Chr:14	21368036	21368060	+	21	1.4e-238	96.00	25
[A] [S] [G] [C]	160	179	+	Chr:10	104254880	104254899	+	Chr:10	104254880	104254899	+	20	0.53	100.00	20
[A] [S] [G] [C]	45	68	+	Chr:11	110649522	110649544	+	Chr:11	110649522	110649544	+	20	0.56	95.83	24
[A] [S] [G] [C]	49	68	+	Chr:7	116873471	116873490	+	Chr:7	116873471	116873490	+	20	2.4	100.00	20
[A] [S] [G] [C]	296	319	+	Chr:9	72392010	72392033	+	Chr:9	72392010	72392033	+	20	2.4	95.83	24
[A] [S] [G] [C]	52	71	-	Chr:11	72461562	72461581	+	Chr:11	72461562	72461581	+	20	2.4	100.00	20
[A] [S] [G] [C]	98	117	-	Chr:10	80893029	80893048	+	Chr:10	80893029	80893048	+	20	2.4	100.00	20
[A] [S] [G] [C]	45	68	+	Chr:4	45774605	45774627	+	Chr:4	45774605	45774627	+	20	2.4	95.83	24
[A] [S] [G] [C]	6	32	-	Chr:17	54805387	54805413	-	Chr:17	54805387	54805413	-	20	2.4	92.86	28
[A] [S] [G] [C]	51	70	-	Chr:4	65910131	65910150	-	Chr:4	65910131	65910150	-	20	2.4	100.00	20
[A] [S] [G] [C]	290	312	+	Chr:3	12210823	12210844	+	Chr:3	12210823	12210844	+	19	0.57	95.65	23
[A] [S] [G] [C]	85	103	-	Chr:15	47541138	47541156	+	Chr:15	47541138	47541156	+	19	0.59	100.00	19
[A] [S] [G] [C]	51	72	-	Chr:3	167257982	167258004	+	Chr:3	167257982	167258004	+	19	2.0	95.65	23
[A] [S] [G] [C]	122	140	+	Chr:21	32414377	32414395	+	Chr:21	32414377	32414395	+	19	9.6	100.00	19


Ensembl v40: Homo sapiens Features on Chromosome 14 21319157-21323567 - Mozilla Firefox
Fichier
Edition
Affichage
Aller à
Marque-pages
Outils
?
http://www.ensembl.org/Homo_sapiens/contigview?chr=14&vc_start=21319157&vc_end=21323567&high
OK
Ensembl release 40 - Aug 2006
Help
Chromosome 14
21,319,157 - 21,323,567
View of Chromosome 14
Graphical view
Graphical overview
Export information about region
Export sequence as FASTA
Export EMBL file
Export Gene info in region
Export SNP info in region
Export Vega info in region
View alignment with ...
View alongside ...
View Syntenic regions ...
View region in NCBI browser
View region in UCSC browser
Use Ensembl to...
Run a BLAST search
Search Ensembl
Data mining [BioMart]
Upload and view data on chromosome
Export data
Download data
Docs and downloads
Information
What's New
About Ensembl
Ensembl data
Software
Other links
Chromosome 14
Chr. 14
p13
p11.2
q11.2
q12
q21.1
q24.3
q31.3
q32.2
Overview
Chr. 14 band
q11.2
20.90 Mb
21.10 Mb
21.30 Mb
21.50 Mb
21.70 Mb
DNA(contigs)
AE000658.1.1.250529 >
AE000659.1.1.251199 >
AE000660.1.1.251124 >
Markers
D14S932
D14S1321
D14S1310
D14S548
D14S1003
D14S742
D14S585
D14S582
D14S50
Ensembl Genes
RPGRIP1
SUPT16H
CHD8
RAB2B
CN092_HUMAN
METTL3
SALL2
OR10G3
OR10G2
TRDV2
Pseudogene
Q6PIM6_HUMAN
TVA1_HUMAN
Q6PJ56_HUMAN
ncRNA Genes
ENSG00000199436
ENSG00000200785
ENSG00000209419
ENSG00000202110
EST Genes
Vega Genes
RPGRIP1
SUPT16H
AL161747.2
AL161747.3
CHD8
RAB2B
C14orf92
SALL2
METTL3
AE000658.15
OR10G3
OR10G1P
AE000658.14
AE000658.7
TRAV1-1
OR10G2
TRAV1-2
AE000658.2
OR4E2
OR4E1P
TRAV2
TRAV4
TRAV3
TRAV5
AE000659.11
TRAV6
AE000659.15
AE000659.4
TRAV8-1
TRAV10
TRAV12-1
AE000659.17
TRAV8-2
TRAV8-3
TRAV13-1
TRAV13-2
TRAV14DV4
TRAV12-2
AE000659.6
TRAV8-4
TRAV9-2
TRAV12-3
AE000659.19
TRAV17
TRAV18
TRAV19
AE000660.21
TRAV20
TRAV8-6
TRAV16
TRAV21
TRAV22
TRAC
TRAV23DV6
TRDV1
TRAV24
TRAV25
TRAV27
TRAV26-1
AE000660.9
TRAV8-7
TRAV29DV5
TRAV26-2
TRAV30
AE000660.1
AE000660.2
AE000660.16
TRAV34
TRAV3
TRAV35
AE
TF
Gene legend
Novel Protein coding
Ig segment
Novel Processed transcript
EST gene
RNA gene
Known Protein coding
Ig pseudogene
Pseudogene
RNA Pseudogene



Ensembl v40: Homo sapiens Features on Chromosome 14 21319157-21323567 - Mozilla Firefox

http://www.ensembl.org/Homo_sapiens/contigview?bottom=%7Cmanaged_IMGT%3Aoff&w=4411&c=14%

Gene legend

Novel Protein coding	Known Protein coding
Ig segment	Ig pseudogene
Novel Processed transcript	Pseudogene
EST gene	
RNA gene	RNA Pseudogene
Ensembl Novel Pseudogene	Ensembl Known Protein Coding

Jump to region 14 : 21319157 - 21323567

Band:

Gene legend

- genscan
- Ensembl Known Protein Coding
- Vega Pseudogene

There are currently 97 tracks switched off, use the menus above the image to turn them on.

Ensembl Homo sapiens version 40.36b (NCBI 36) Chromosome 14 21,319,157 - 21,323,567

© 2006 WTSI / EBI. Ensembl is available to download for public use - please see the code licence for details.