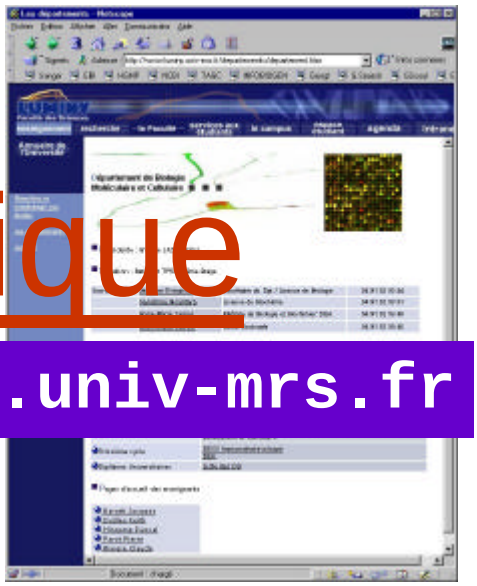
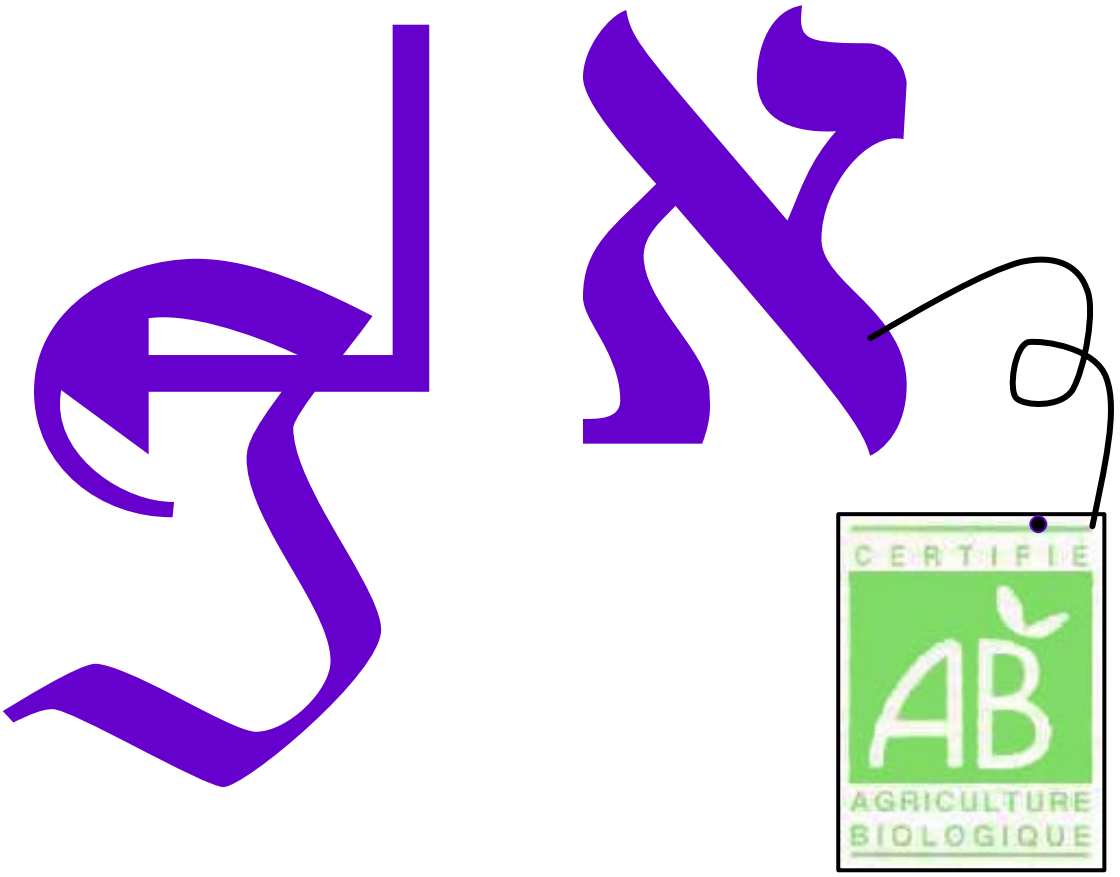
Bioinformatique



http://www.luminy.univ-mrs.fr

- Qu'est ce?
  - Ce que ce n'est pas...
  - Définition***s!***
  - Exemples
  - Science / Technologie ?
    (créateurs / utilisateurs?)
  - Bref historique
  - Catégories
- A quoi ça ressemble?
  - lignes de commande
  - logiciels interactifs
  - interfaces Internet

Informatique «Bio»



Enki Bilal (1977) Mémoires d'Outre Espace

éditions *Les humanoïdes Associés*

Bioinformatiques

Bioinfo de labo
support au clonage, séquençage, & PCR...

Analyse séquence
identification gènes, comparaisons de séquences, prédiction motifs...

Phylogénie
évolution à l'échelle moléculaire...

Structure des protéines
calcul, visualisation, prédiction...

Liaison génétique
gènes candidats de maladies génétiques...

Génomique fonctionnelle
transcriptome, protéome, interactome...

Bref historique

1970 Needlman & Wunsch
Alignement optimal de 2 séquences

1974 Chou & Fasman
Prédiction structure II protéines

1978 Dayhoff *et. al.*
Matrice de substitution (PAM)

1980 EMBL data library
Banque internationale de séquence

1981 Smith & Waterman
Similarités de séquences dans les banques

1981 Felsenstein *et. al.*
Arbres phylogénét. par comparaison de séq. (phylip)

1984 Devereux *et. al.*
Logiciel d'analyse de séquence (UW GCG)

1987 Feng & Doolittle
Alignement multiple progressif de séquences

1988 Bairoch *et. al.*
Banque de signatures protéiques (PROSITE)

1988 Pearson & Lipman
Sim. de séq. dans les banques (FASTA)

Bref historique (suite)

1990 Atschul *et. al.*
Sim. de séq. dans les banques (BLAST)

1991 Bowie *et. al.*
Prédiction struct. III protéines

1993 Borodovsky *et. al.*
Prédiction gènes génomes bactériens (GeneMark)

1995 Venter *et. al.*
Séquençage 'shotgun' génome Haemophilus *1.8Mb*

1997 Sonnhammer *et. al.*
Banque de domaines protéiques (PFam)

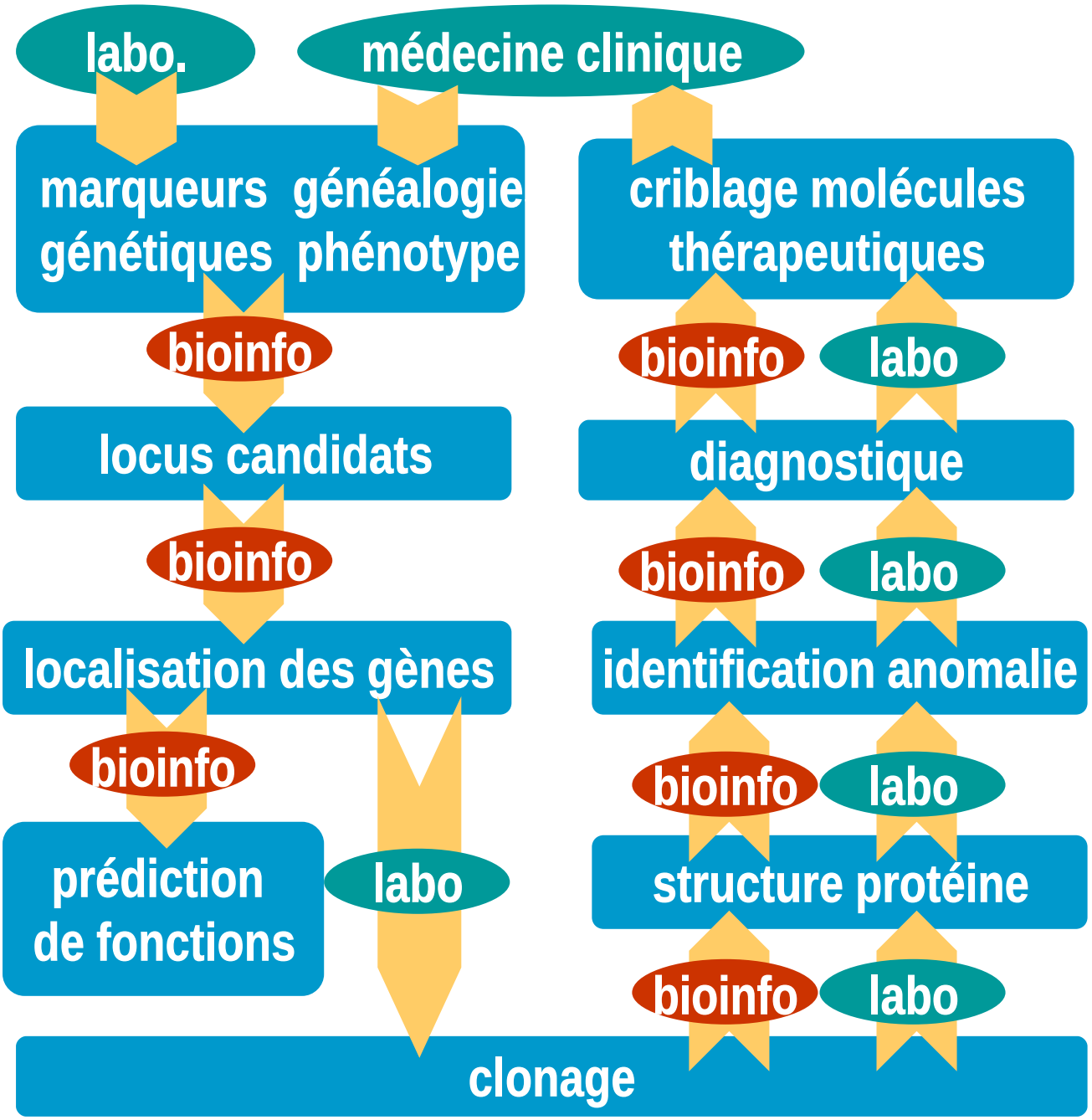
1997 Burge *et. al.*
Prédiction gènes génomes eucaryotes (GENSCAN)

1998 Eisen *et. al.*
Analyse des profils d'expression (CLUSTER)

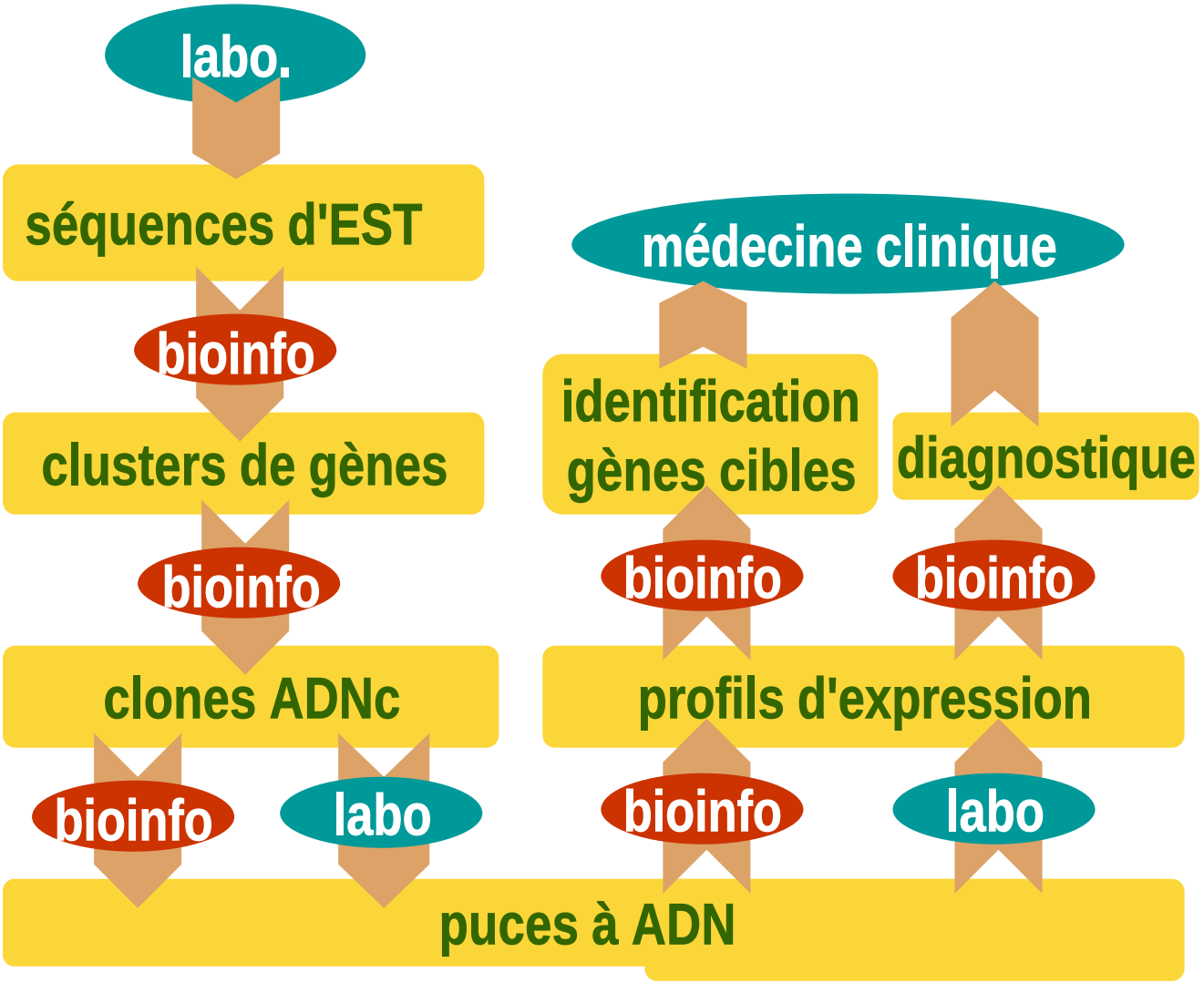
2000 Venter *et. al.*
'Shotgun' génome complet Drosophile *120 Mb*

2001 Venter *et. al.*
'Shotgun' génome complet Humain *3000 Mb*

Exemple d ’application 1



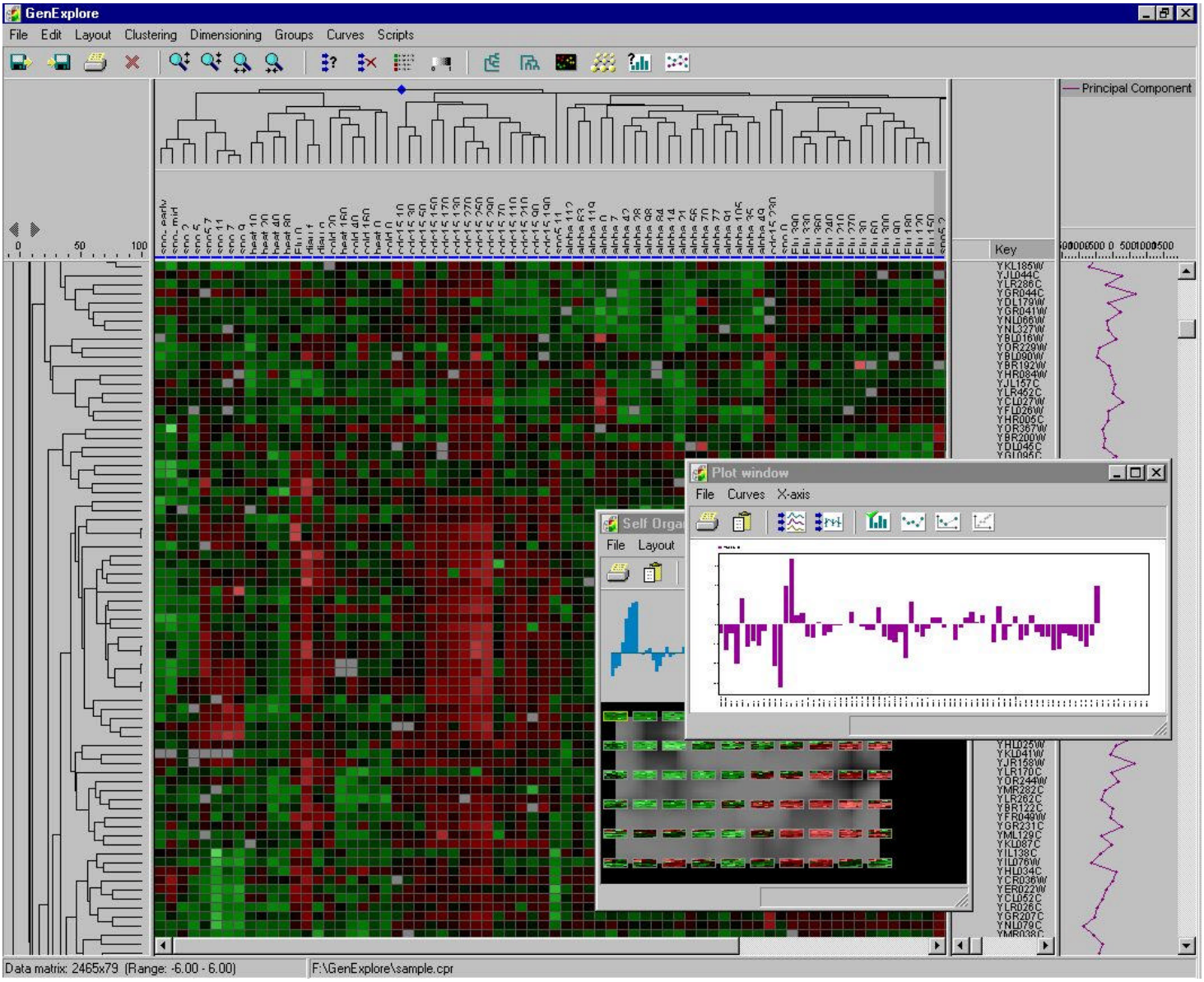
Exemple d ’application 2



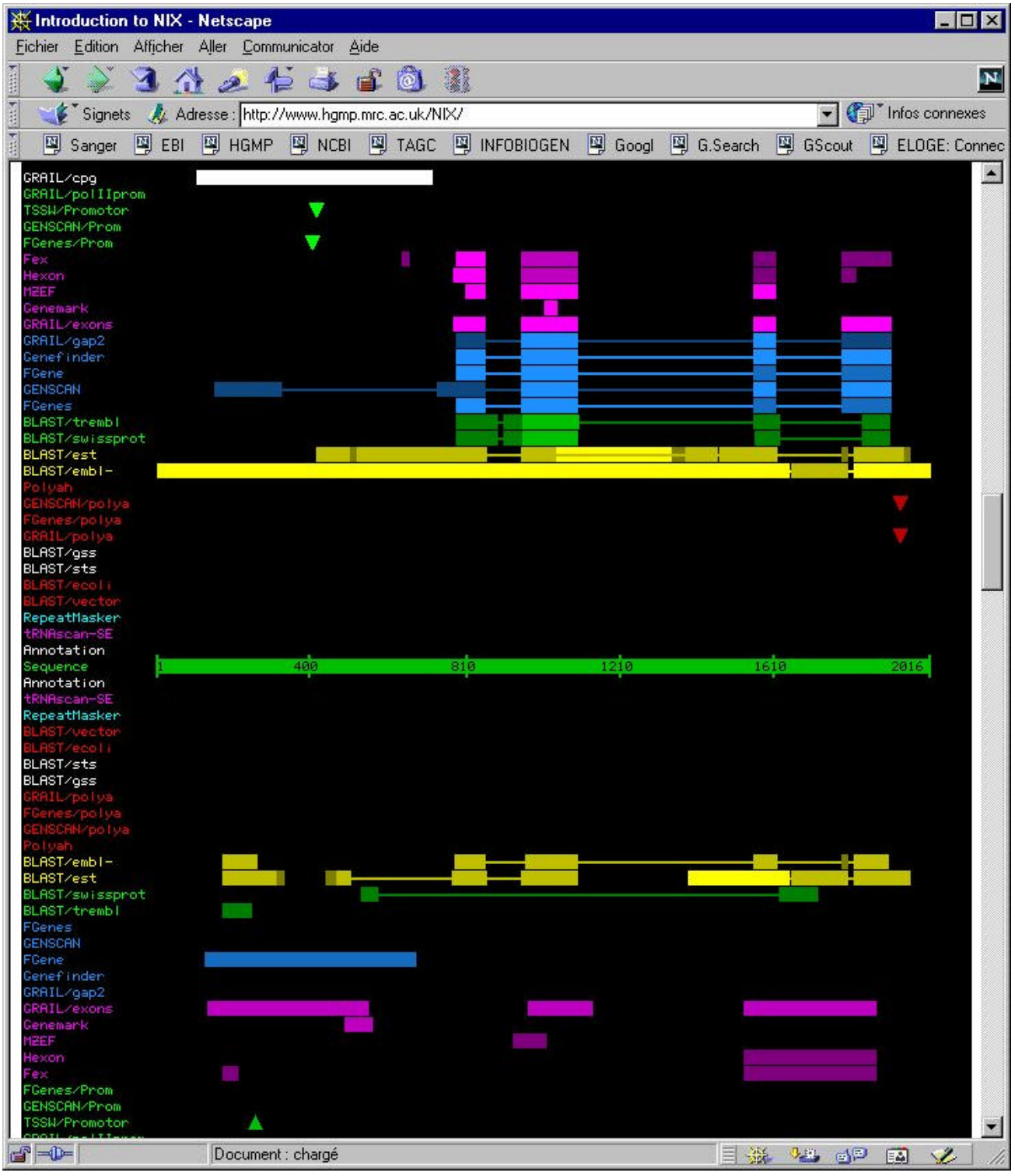
Ligne de commande

```
Unix % genefinder -corrfactor 0.8 -tablenamefile
/packages/genefinder_washu/genefinder/human/humtables.hex -cpenalty 6.35 -penaltyfactor
0.85 -maxintronlength 10500 -minintronlength 70 -penaltycluster 50 -seqfile
pax6.wild.type.seq

Genefinder Version 084, release 960118
Parameters:
-seqfile pax6.wild.type.seq, -genefile "", -debug 0, -syncodon FALSE,
-normalize FALSE, -exoncutoff 3, -genecutoff 0, -atgcutoff 0,
-intron3cutoff 0, -intron5cutoff 0, -orfcutoff 5, -minexonlength 33,

Sequence: hsa1280.embl,.C
Length: 22253, nuc counts (N,A,C,G,T): 0 6645 4848 4692 6068

 394 3' sites w/ scores > 0.00
 310 5' sites w/ scores > 0.00
 232 atg sites w/ scores > 0.00

Sequence: hsa1280.embl,
Length: 22253, nuc counts (N,A,C,G,T): 0 6068 4692 4848 6645

 408 3' sites w/ scores > 0.00
 429 5' sites w/ scores > 0.00
 249 atg sites w/ scores > 0.00

No predicted genes in hsa1280.embl,.C.

ORFs:

 ORF:   17183 16701   483   4.16   5.79     6.00
*TKIISTMSKRPTPRHCYHTEPAMLSPRECLELCLEGQLMSLGCGQKARQPFPSCRTRGLLQLSSRFFRLYQTVLLCPGNPDLQSTHIRA
QVHTYRPSAPTSLLFSHSHLLHLACDIRVSKVSAVYTTEPWAGTGCLTVVEITASEKEEVSQPAAVACTPT*
        17144 17109    36   1.12
  1  3' 17122     1.62
    atg 17114     2.07
  2  3' 17100     0.99
    atg 17066     1.10
  2  5' 17056     2.54
  1  3' 17056     0.92
  1  5' 17054     1.24
  2  3' 17010     1.55
       17009 16983    27   2.10
  0  3' 16997     0.85
  2  3' 16989     0.67
  2  3' 16977     0.94
  0  3' 16967     1.24
  0  3' 16931     1.53
       16853 16719   135   5.79
  1  3' 16798     1.52
  2  5' 16795     1.03
  0  5' 16737     2.49

 ORF:    5582  4998   585   5.66   4.06     6.00
*QSPPKNLTCGWGGGDCQLTQNFLATRRVKVKSPWTSSTSGPNHHRRTLADLTLTSRPQTAGDENPPESGATSGGSCCPRLLPNSKKRKG
GVSCAPSSWEPTHRGEPGVEALGSALPGLGHRALSPSPRLLGLSGPDSCQGRDHQDLRLLGAEGWTLVSAKFGCHLGVSSVYKWESYGFK
SQAGMGLPTLGGVNR*
        5507  5457    51   1.27
  2  3'  5445     2.97
        5408  5256   153   4.06
  0  3'  5405     3.67
  1  5'  5363     1.70
  0  3'  5324     1.78
```

Logiciels interactifs



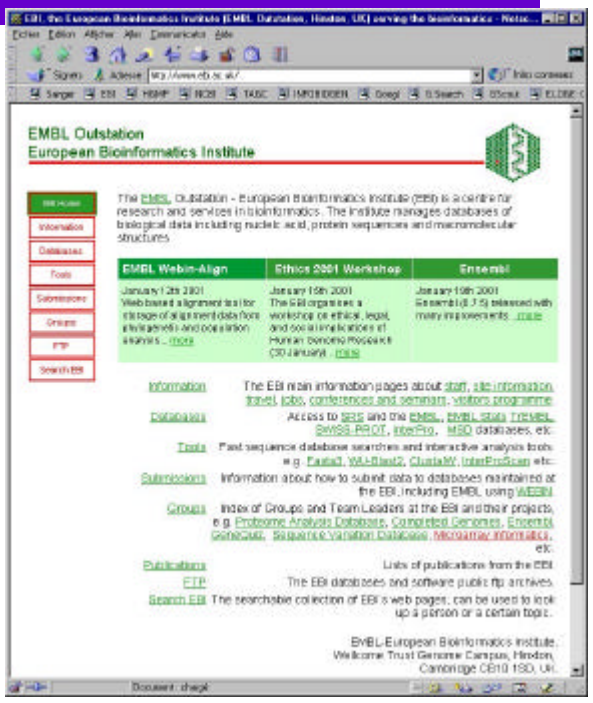
Interfaces Internet



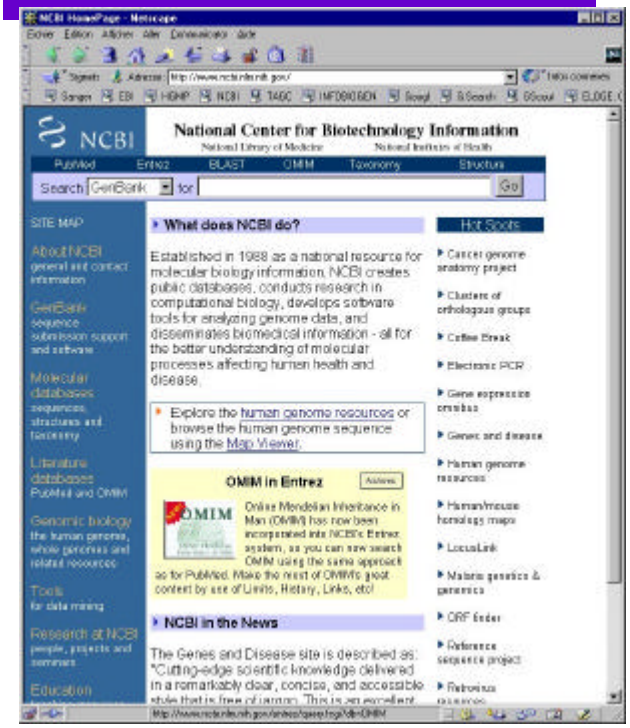
Centres de ressources

Infobiogen

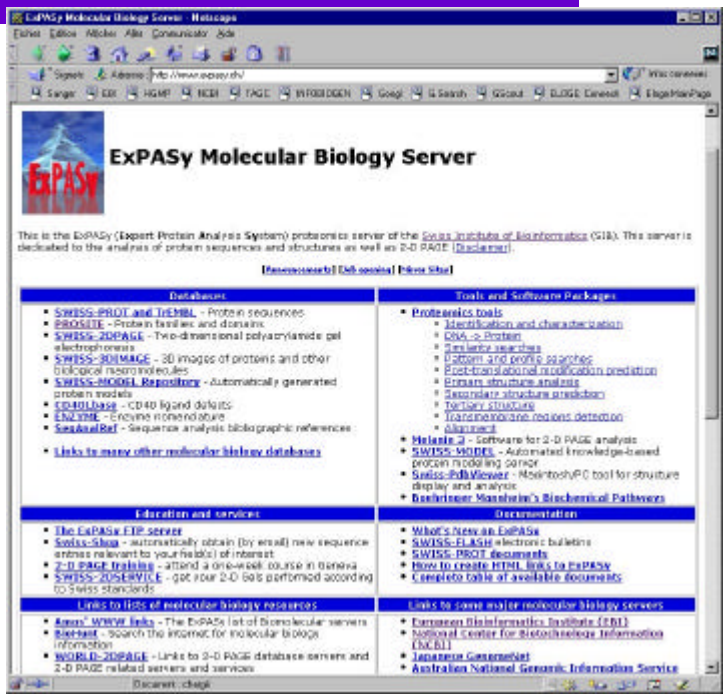
EBI



NCBI



Expasy



Plan du cours

1. Introduction
2. Banques de données
3. Matrices de substitutions
4. Alignement de 2 séquences
5. Recherche dans les banques par similarité de séquence
6. Alignements multiples
7. Motifs & signatures protéiques
8. Profils & domaines protéiques

Bases de données

- Modes de constitution

  Soumission directe:
  archives **primaires**

  Sélection & expertise:
  banques de **référence**

  redondantes

  incomplètes

- Catégories de données

  base de données moléculaires

  banques de connaissances

- Accès

  Numéros d'accession

  Annotation

  Formats

Banques de données moléculaires

Séquences nucléiques

- **EMBL/GENBANK**
- **dbEST**
- ***UNIGENE**
- ***ENSEMBL**
- ***RefSeq...**

Séquences protéiques

- **TrEMBL**
- ***SWISSPROT**
- ***PIR**

Structures protéines

- **PDB, * CATH, * SCOP...**

**** banques de référence expertisées (non redondantes)***

Banques de connaissances

Fonct/relat-ionnelles

- **LocusLink**
- **GeneCards**
- **PROSITE, PFam**
- **KEGG...**

Ontologies

- **Taxonomie**
- **Symboles gènes**
- **Enzymes**
- **Tissus...**

Organisme centrique

- **GDB**
- **MGI**
- **FlyBase**
- **AceDB...**

Bibliographiques

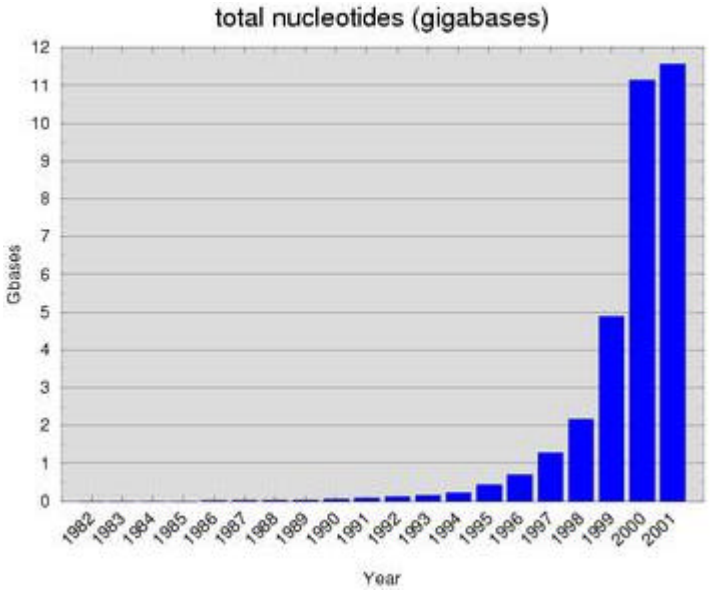
- **MEDLINE**
- **Journaux Online**
- **OMIM**

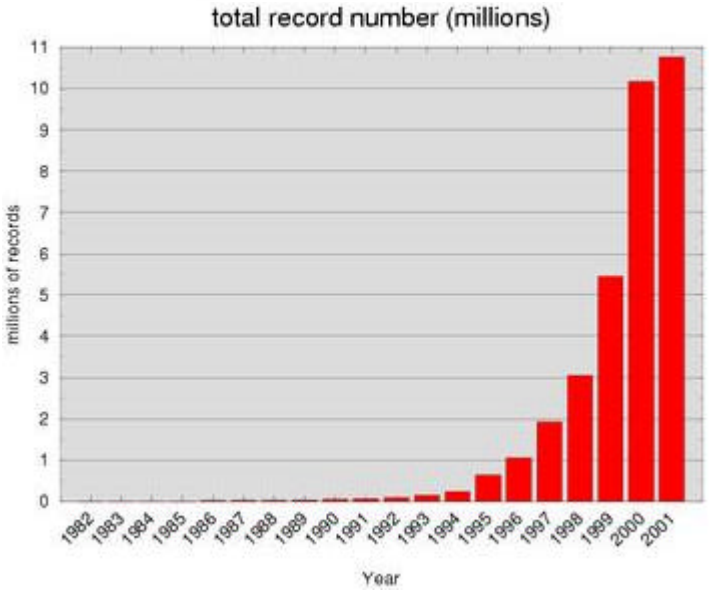
EMBL Outstation
European Bioinformatics Institute

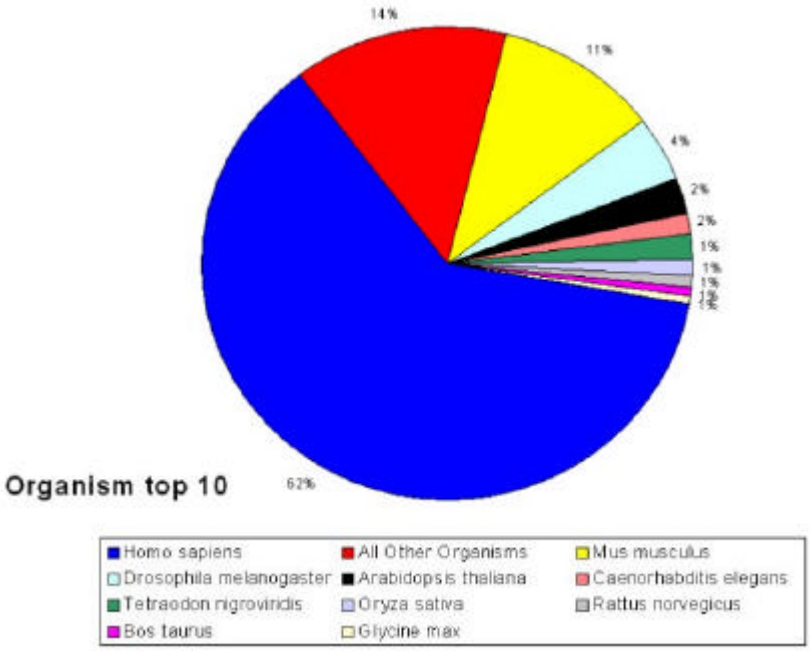
EMBL Nucleotide Sequence Database Statistics

EMBL Database growth

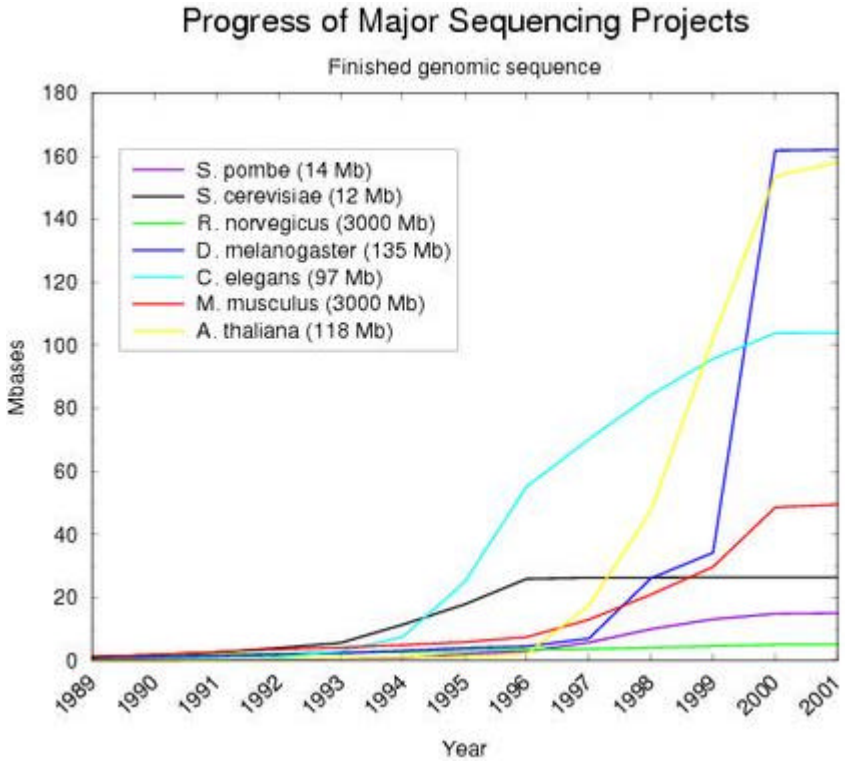
This morning, the EMBL Database contained **11,571,085,945** bases in **10,770,275** records.

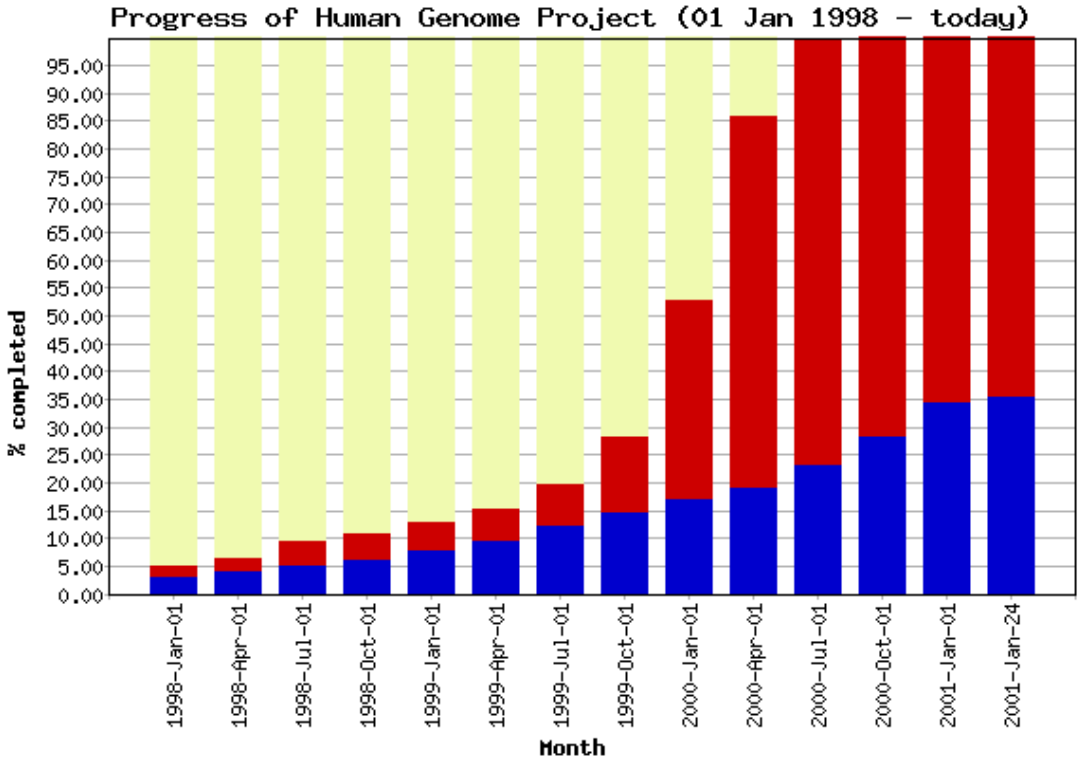






EMBL: génomes





Séquence brute

```
GGGTTTTATACGGATCTCCCTTCTCGTTTGATAATTATGCCATTAAAGGTTTTACCAGTTCATAAATTTA
GTAAAAATGAACCCCATAAAAAACAAAAGAGGTTCATCTACTTTTAAGGAATTAAACCAAGGAATTTAAT
TCATATTAAATAGCCATGGTTTCCAGTTTTTTACTGGCAGAGTACAAAAACCTAATAGTGAATCCCTCTG
AGCATTTCAAAATCTCAGTGAATGAAGATAATTTGACTGAATGGGATGTCATCTTAAAAGGCCCACCTGA
CACTCTTTATGAGGGAGGCTTATTCAAAGCAAAGATTGTCTTTCCTCCAAAATACCCATATGAACCACCC
AGATTAACATTCACCTCTGAAATGTGGCATCCCAATATCTACTCTGATGGGAAATTATGTATTTCTATCT
TGCATGGAGACAATGCTGAAGAACAGGGAATGACTTGGTCTCCGGCTCAAAAGATTGATACCGTACTTCT
TAGTGTAATTTCTCTGCTCAATGAGCCAAATCCAGATTCTCCAGCAAATGTAGATGCAGCTAAAAGCTAC
CGTAAATATCTATATAAAGAGGATTTAGAATCATACCCCATGGAAGTTAAAAAGACTGTCAAAAAATCAT
TGGATGAGTGTTCAGCGGAAGACATAGAATATTTTAAAAATGTTCCAGTGAATGTTCTACCAGTACCCAG
TGATGATTATGAAGATGAAGAAATGGAGGATGGCACCTATATCTTAACCTATGATGATGAGGATGAAGAA
GAGGATGAAGAGATGGATGATGAGTAGTGCTGATTTTAATGCATAACATATTAGTTACTTACACTTTAGT
GCTTAGATTTTAGTGTTTAAACTTTAGTGATTAGATTTTAGTGCTTAGATTTTAGTGTTTAAACTTTAGT
GATTAGATTTTAGTGCTTAGATTTTAATGTTTAAACTTTAGTGATTAGATTTTAGTGCTTAGATTTTAA
```

Format FASTA

```
>X62440 African swine fever virus DNA for ubiquitin conjugating enzyme
GGGTTTTATACGGATCTCCCTTCTCGTTTGATAATTATGCCATTAAAGGTTTTACCAGTTCATAAATTTA
GTAAAAATGAACCCCATAAAAAACAAAAGAGGTTCATCTACTTTTAAGGAATTAAACCAAGGAATTTAAT
TCATATTAAATAGCCATGGTTTCCAGTTTTTTACTGGCAGAGTACAAAAACCTAATAGTGAATCCCTCTG
AGCATTTCAAAATCTCAGTGAATGAAGATAATTTGACTGAATGGGATGTCATCTTAAAAGGCCCACCTGA
CACTCTTTATGAGGGAGGCTTATTCAAAGCAAAGATTGTCTTTCCTCCAAAATACCCATATGAACCACCC
AGATTAACATTCACCTCTGAAATGTGGCATCCCAATATCTACTCTGATGGGAAATTATGTATTTCTATCT
TGCATGGAGACAATGCTGAAGAACAGGGAATGACTTGGTCTCCGGCTCAAAAGATTGATACCGTACTTCT
TAGTGTAATTTCTCTGCTCAATGAGCCAAATCCAGATTCTCCAGCAAATGTAGATGCAGCTAAAAGCTAC
CGTAAATATCTATATAAAGAGGATTTAGAATCATACCCCATGGAAGTTAAAAAGACTGTCAAAAAATCAT
TGGATGAGTGTTCAGCGGAAGACATAGAATATTTTAAAAATGTTCCAGTGAATGTTCTACCAGTACCCAG
TGATGATTATGAAGATGAAGAAATGGAGGATGGCACCTATATCTTAACCTATGATGATGAGGATGAAGAA
GAGGATGAAGAGATGGATGATGAGTAGTGCTGATTTTAATGCATAACATATTAGTTACTTACACTTTAGT
GCTTAGATTTTAGTGTTTAAACTTTAGTGATTAGATTTTAGTGCTTAGATTTTAGTGTTTAAACTTTAGT
GATTAGATTTTAGTGCTTAGATTTTAATGTTTAAACTTTAGTGATTAGATTTTAGTGCTTAGATTTTAA
```

Fiche EMBL

```
ID   ASFVUCE    standard; DNA; VRL; 979 BP.
XX
AC   X62440;
XX
SV   X62440.1
XX
DT   15-JAN-1992 (Rel. 30, Created)
DT   27-APR-1992 (Rel. 31, Last updated, Version 2)
XX
DE   African swine fever virus DNA for ubiquitin conjugating enzyme
XX
KW   ubiquitin; ubiquitin-conjugating enzyme.
XX
OS   African swine fever virus
OC   Viruses; dsDNA viruses, no RNA stage; Asfarviridae;
OC   African swine fever-like viruses.
XX
RN   [1]
RX   MEDLINE; 92155177.
RA   Hingamp P.M., Arnold J.E., Mayer R.J., Dixon L.K.;
RT   "A Ubiquitin Conjugating  Enzyme Encoded by African Swine Fever Virus";
RL   EMBO J. 11:361-366(1992).
XX
RN   [2]
RP   1-979
RA   Dixon L.K.;
RT   ;
RL   Submitted (04-OCT-1991) to the EMBL/GenBank/DDBJ databases.
RL   L.K. Dixon, Institute of Animal Health, Pirbright Laboratory, Ash Road,
RL   Pirbright, Woking, Surrey GU24 0NF, UK
XX
DR   SWISS-PROT; P25869; UBC_ASFM2.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..979
FT                   /db_xref="taxon:10497"
FT                   /organism="African swine fever virus"
FT                   /cell_line="pig blood"
FT                   /clone_lib="lambda EMBL1 Lmw"
FT                   /clone="Lmw22"
FT                   /isolate="Malawi LIL20/1"
FT   5'UTR           1..155
FT   CDS             156..797
FT                   /db_xref="SWISS-PROT:P25869"
FT                   /gene="ASFVUBC"
FT                   /product="ubiquitin conjugating  enzyme"
FT                   /protein_id="CAA44305.1"
FT                   /translation="MVSSFLLAEYKNLIVNPSEHFKISVNEDNLTEWDVILKGPPDTLY
FT                   EGGLFKAKIVFPPKYPYEPPRLTFTSEMWHPNIYSDGKLCISILHGDNAEEQGMTWSPA
FT                   QKIDTVLLSVISLLNEPNPDSPANVDAAKSYRKYLYKEDLESYPMEVKKTVKKSLDECS
FT                   AEDIEYFKNVPVNVLPVPSDDYEDEEMEDGTYILTYDDEDEEEDEEMDDE"
FT   misc_feature    408..410
FT                   /note="active site cysteine"
FT   3'UTR           798..979
XX
SQ   Sequence 979 BP; 329 A; 151 C; 182 G; 317 T; 0 other;
     gggttttata cggatctccc ttctcgtttg ataattatgc cattaaaggt tttaccagtt        60
     cataaattta gtaaaaatga accccataaa aaacaaaaga ggttcatcta cttttaagga       120
                                [...]
     aaactttagt gattagattt tagtgcttag attttaatgt ttaaacttta gtgattagat       960
     tttagtgctt agattttaa                                                   979
//
```

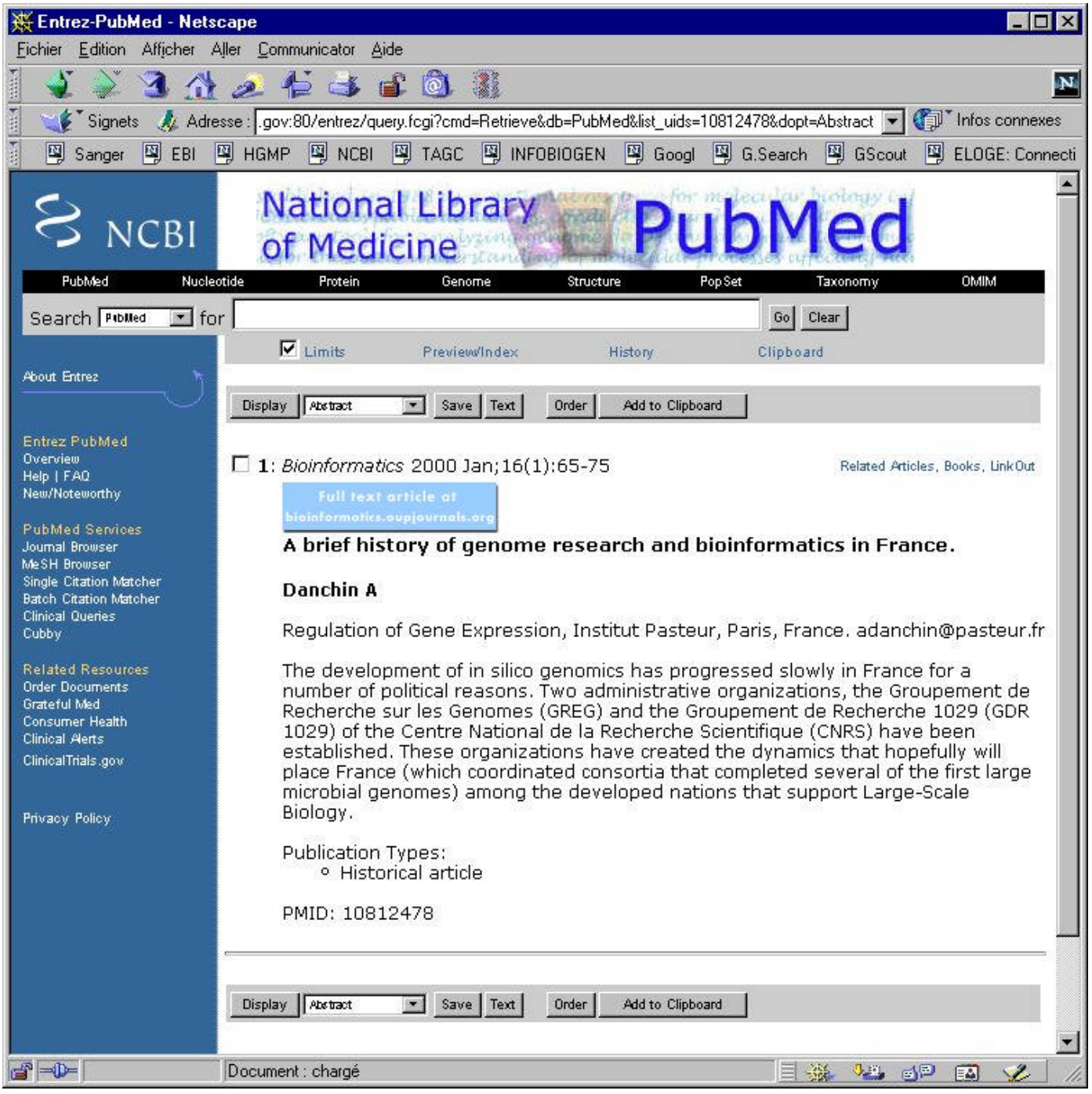
Fiche GENBANK

```
  LOCUS       ASFVUCE       979 bp    DNA             VRL       27-APR-1992
DEFINITION  African swine fever virus DNA for ubiquitin conjugating enzyme.
ACCESSION   X62440
VERSION     X62440.1  GI:58649
KEYWORDS    ubiquitin; ubiquitin-conjugating enzyme.
SOURCE      African swine fever virus.
  ORGANISM  African swine fever virus
            Viruses; dsDNA viruses, no RNA stage; African swine fever-like
            viruses.
REFERENCE   1  (bases 1 to 979)
  AUTHORS   Hingamp,P.M., Arnold,J.E., Mayer,R.J. and Dixon,L.K.
  TITLE     A ubiquitin conjugating enzyme encoded by African swine fever virus
  JOURNAL   EMBO J. 11 (1), 361-366 (1992)
  MEDLINE   92155177
REFERENCE   2  (bases 1 to 979)
  AUTHORS   Dixon,L.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-1991) L.K. Dixon, Institute of Animal Health,
            Pirbright Laboratory, Ash Road, Pirbright, Woking, Surrey GU24 0NF,
            UK
FEATURES             Location/Qualifiers
     source          1..979
                     /organism="African swine fever virus"
                     /isolate="Malawi LIL20/1"
                     /db_xref="taxon:10497"
                     /clone="Lmw22"
                     /cell_line="pig blood"
                     /clone_lib="lambda EMBL1 Lmw"
     5'UTR           1..155
     gene            156..797
                     /gene="ASFVUBC"
     CDS             156..797
                     /gene="ASFVUBC"
                     /codon_start=1
                     /product="ubiquitin conjugating  enzyme"
                     /protein_id="CAA44305.1"
                     /db_xref="GI:58650"
                     /db_xref="SWISS-PROT:P25869"
                     /translation="MVSSFLLAEYKNLIVNPSEHFKISVNEDNLTEWDVILKGPPDTL
                     YEGGLFKAKIVFPPKYPYEPPRLTFTSEMWHPNIYSDGKLCISILHGDNAEEQGMTWS
                     PAQKIDTVLLSVISLLNEPNPDSPANVDAAKSYRKYLYKEDLESYPMEVKKTVKKSLD
                     ECSAEDIEYFKNVPVNVLPVPSDDYEDEEMEDGTYILTYDDEDEEEDEEMDDE"
     misc_feature    408..410
                     /gene="ASFVUBC"
                     /note="active site"
     3'UTR           798..979
BASE COUNT      329 a    151 c    182 g    317 t
ORIGIN
        1 gggttttata cggatctccc ttctcgtttg ataattatgc cattaaaggt tttaccagtt
       61 cataaattta gtaaaaatga accccataaa aaacaaaaga ggttcatcta cttttaagga
                                  […]
      901 aaactttagt gattagattt tagtgcttag attttaatgt ttaaacttta gtgattagat
      961 tttagtgctt agattttaa
//
```

Fiche SWISSPROT

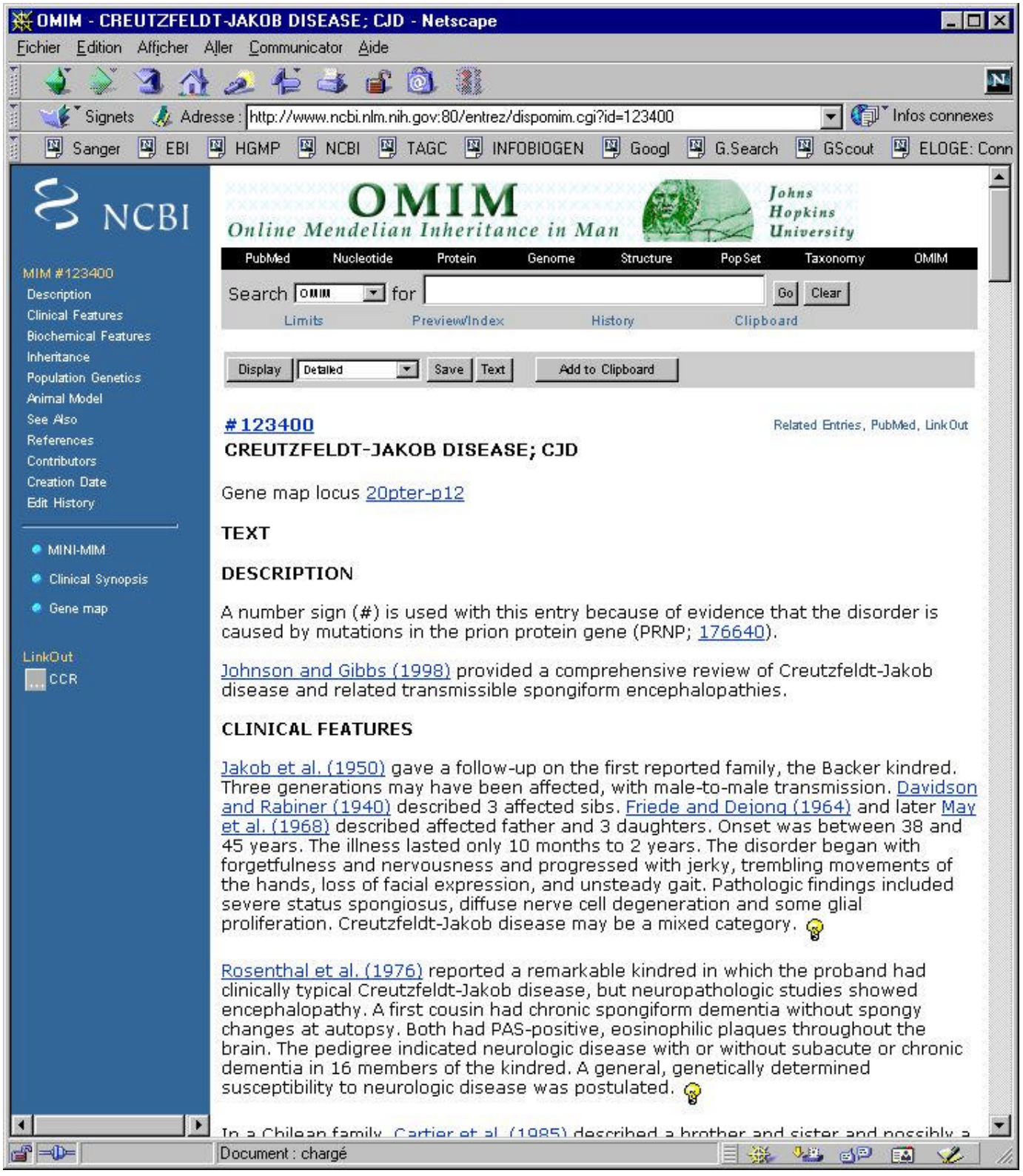
```
ID   UBC_ASFM2      STANDARD;      PRT;   213 AA.
AC   P25869;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   UBIQUITIN-CONJUGATING ENZYME E2-21 KDA (EC 6.3.2.19)
DE   (UBIQUITIN-PROTEIN LIGASE) (UBIQUITIN CARRIER PROTEIN).
GN   ASFV-UBC.
OS   African swine fever virus (isolate Malawi Lil 20/1) (ASFV).
OC   Viruses; dsDNA viruses, no RNA stage; Asfarviridae;
OC   African swine fever-like viruses.
OX   NCBI_TaxID=10500;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92155177; PubMed=1310934;
RA   Hingamp P.M., Arnold J.E., Mayer R.J., Dixon L.K.;
RT   "A ubiquitin conjugating enzyme encoded by African swine fever
RT   virus.";
RL   EMBO J. 11:361-366(1992).
CC   -!- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO
CC       OTHER PROTEINS.
CC   -!- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE = AMP +
CC       PYROPHOSPHATE + PROTEIN N-UBIQUITYLLYSINE.
CC   -!- PATHWAY: SECOND STEP IN UBIQUITIN CONJUGATION.
CC   -!- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR
CC       UBIQUITIN-THIOLESTER FORMATION.
CC   -!- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
DR   EMBL; X62440; CAA44305.1; -.
DR   EMBL; X71982; CAA50851.1; -.
DR   PIR; S19158; S19158.
DR   HSSP; Q02159; 2UCZ.
DR   InterPro; IPR000608; -.
DR   Pfam; PF00179; UQ_con; 1.
DR   PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; 1.
DR   PROSITE; PS50127; UBIQUITIN_CONJUGAT_2; 1.
KW   Ubiquitin conjugation; Ligase.
FT   BINDING      85     85       UBIQUITIN (BY SIMILARITY).
FT   DOMAIN      183    213       ASP/GLU-RICH (BASIC).
SQ   SEQUENCE   213 AA;  24468 MW;  F9671BC7385D6DCE CRC64;
     MVSSFLLAEY KNLIVNPSEH FKISVNEDNL TEWDVILKGP PDTLYEGGLF KAKIVFPPKY
     PYEPPRLTFT SEMWHPNIYS DGKLCISILH GDNAEEQGMT WSPAQKIDTV LLSVISLLNE
     PNPDSPANVD AAKSYRKYLY KEDLESYPME VKKTVKKSLD ECSAEDIEYF KNVPVNVLPV
     PSDDYEDEEM EDGTYILTYD DEDEEEDEEM DDE
//
```

Abstracts PubMed

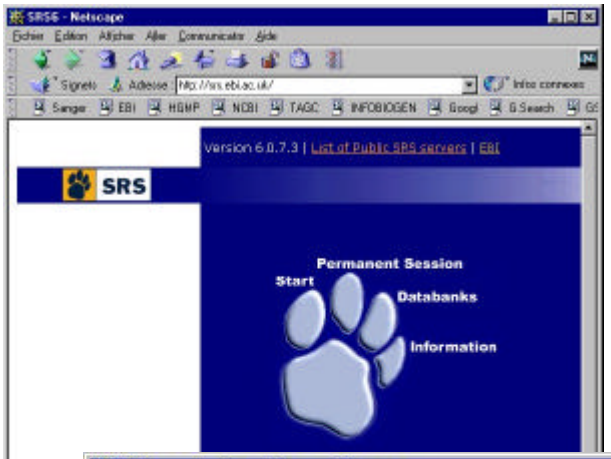
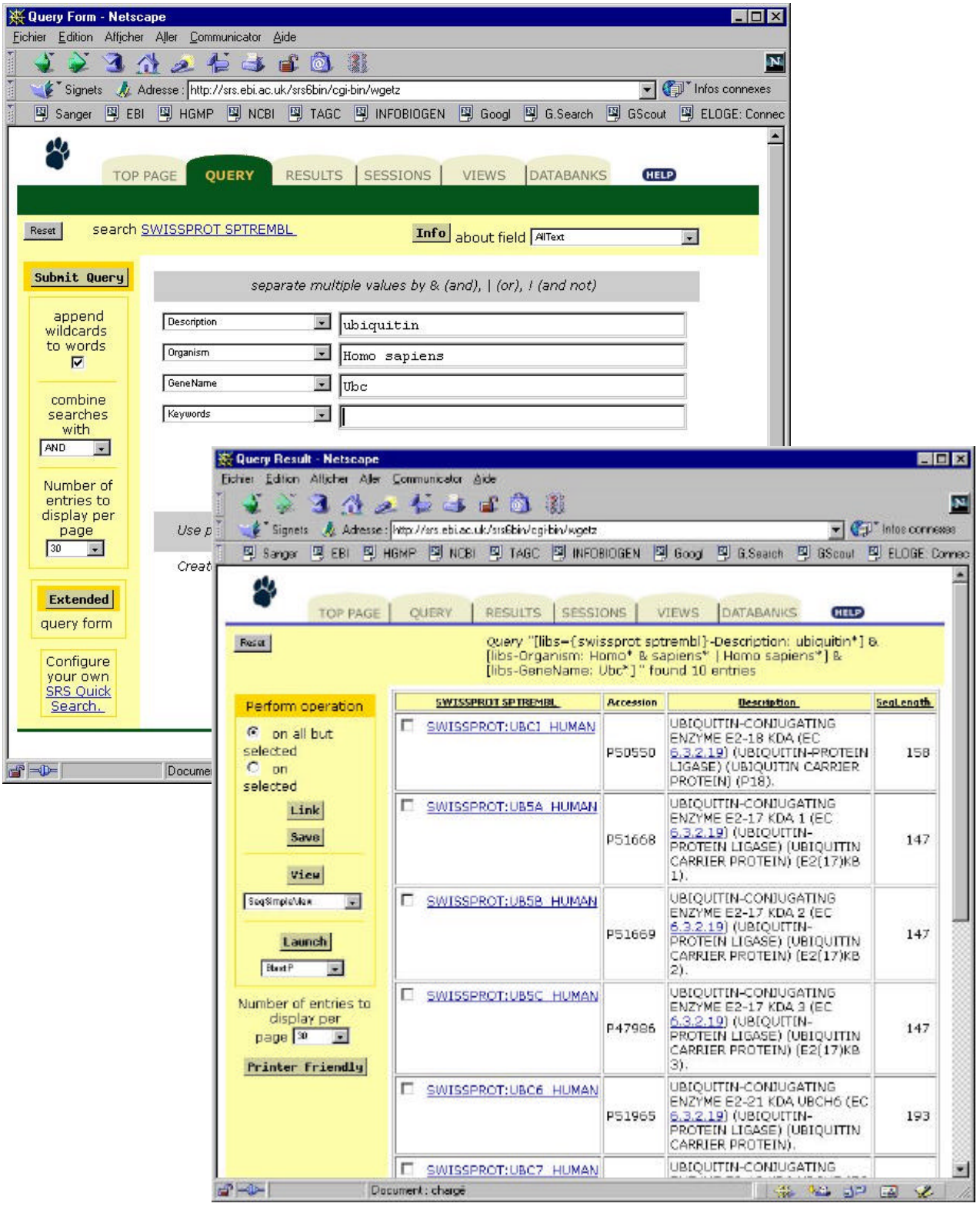


Journaux ‘Online’

OMIM


OMIM - CREUTZFELDT-JAKOB DISEASE; CJD - Netscape
Fichier Edition Afficher Aller Communicator Aide
Signets Adresse : http://www.ncbi.nlm.nih.gov:80/entrez/dispomim.cgi?id=123400
Infos connexes
Sanger EBI HGMP NCBI TAGC INFOBIOGEN Googl G.Search GScout ELOGE: Conn
NCBI
OMIM
Online Mendelian Inheritance in Man
Johns Hopkins University
PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM
Search OMIM for Go Clear
Limits Preview/Index History Clipboard
Display Detailed Save Text Add to Clipboard
MIM #123400
Description
Clinical Features
Biochemical Features
Inheritance
Population Genetics
Animal Model
See Also
References
Contributors
Creation Date
Edit History
MINI-MIM
Clinical Synopsis
Gene map
LinkOut
CCR
#123400
Related Entries, PubMed, LinkOut
CREUTZFELDT-JAKOB DISEASE; CJD
Gene map locus 20pter-p12
TEXT
DESCRIPTION
A number sign (#) is used with this entry because of evidence that the disorder is caused by mutations in the prion protein gene (PRNP; 176640).
Johnson and Gibbs (1998) provided a comprehensive review of Creutzfeldt-Jakob disease and related transmissible spongiform encephalopathies.
CLINICAL FEATURES
Jakob et al. (1950) gave a follow-up on the first reported family, the Backer kindred. Three generations may have been affected, with male-to-male transmission. Davidson and Rabiner (1940) described 3 affected sibs. Friede and Dejong (1964) and later May et al. (1968) described affected father and 3 daughters. Onset was between 38 and 45 years. The illness lasted only 10 months to 2 years. The disorder began with forgetfulness and nervousness and progressed with jerky, trembling movements of the hands, loss of facial expression, and unsteady gait. Pathologic findings included severe status spongiosus, diffuse nerve cell degeneration and some glial proliferation. Creutzfeldt-Jakob disease may be a mixed category.
Rosenthal et al. (1976) reported a remarkable kindred in which the proband had clinically typical Creutzfeldt-Jakob disease, but neuropathologic studies showed encephalopathy. A first cousin had chronic spongiform dementia without spongy changes at autopsy. Both had PAS-positive, eosinophilic plaques throughout the brain. The pedigree indicated neurologic disease with or without subacute or chronic dementia in 16 members of the kindred. A general, genetically determined susceptibility to neurologic disease was postulated.
In a Chilean family, Cartier et al. (1985) described a brother and sister and possibly a
Document : chargé

Recherche par l'annotation



Recherche par l'annotation