

RefSeq (NCBI Reference Sequences)

But :

- produire une banque de référence pour les séquences génomiques, les ARN, les protéines
- éviter la redondance => choix d'une séquence représentative
- corriger les erreurs
- ajouter des informations par des processus automatiques ou l'intervention d'experts

Différents statuts

Unknown
Model
Inferred
Predicted
Provisional
Validated
Reviewed

Traitements automatiques

Séquences annotées par des experts
Informations complémentaires ajoutées

RefSeq and LocusLink: NCBI gene-centered resources.
Pruitt & Maglott. *Nucleic acids research* 29:137-140 (2001)

RefSeq (NCBI Reference Sequences)

	Type	Accession format	Méthodes	Catégories
2 161 000	génomique	NC_	validation	Séquences génomiques complètes (chromosomes, organelles, plasmids)
		NG_	validation	Régions génomiques incomplètes
		NT_	automatique	Assemblage de contigs
		NZ_	automatique	Séquences provenant d'un projet de séquençage en cours
2 237 000	mRNA	NM_	validation	Genome annotation
		XM_	prédiction	
	RNA	NR_	validation	
10 641 000	protéique	XR_	prédiction	Genome annotation
		NP_	validation	Genome annotation
		XP_	prédiction	
		ZP_	prédiction	

Banques

Reviewed REFSEQ

LOCUS NG_000002 909203 bp DNA linear PRI 08-AUG-2002

DEFINITION Homo sapiens immunoglobulin lambda locus (IGL@) on chromosome 22.

ACCESSION NG_000002

VERSION NG_000002.1 GI:18860922

KEYWORDS .

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 909203)

REFERENCE

AUTHORS Hieter,P.A., Korsmeyer,S.J., Waldmann,T.A. and Leder,P.

TITLE Human immunoglobulin kappa light-chain genes are deleted or rearranged in lambda-producing B cells

JOURNAL Nature 290 (5805), 368-372 (1981)

MEDLINE 81173034

PUBMED 6783958

...

REFERENCE

9 (bases 1 to 909203)

AUTHORS Kawasaki,K., Minoshima,S., Nakato,E., Shibuya,K., Shintani,A., Schmeits,J.L., Wang,J. and Shimizu,N.

TITLE One-megabase sequence analysis of the human immunoglobulin lambda gene locus

JOURNAL Genome Res. 7 (3), 250-261 (1997)

MEDLINE 97228902

PUBMED 9074928

COMMENT **REVIEWED REFSEQ:** This record has been curated by NCBI staff. The reference sequence was derived from D86993.1, D86989.2, D86991.1, D86994.1, D86996.1, D86998.1, D86999.1, D87000.2, D87002.1, D87003.1, D87004.2, D87006.1, D87007.1, D87009.1, D87010.1, D87011.2, D87013.1, D87014.1, D87015.2, D87016.2, D87017.2, D87018.1, D87020.1, D87021.1, D87022.1, D87023.1, D87024.1, D88269.2, D88270.2 and D88271.2.

Summary: Immunoglobulins recognize foreign antigens and initiate immune responses such as phagocytosis and the complement system...

Nombreuses références biblio

Diminution de la redondance

Résumé sur le gène ou la région considérée

Banques

Protéine annotée dans RefSeq

LOCUS NP_147707 51 aa linear BCT 09-JUL-2001

DEFINITION S ribosomal protein L39E [Aeropyrum pernix].

ACCESSION NP_147707

VERSION NP_147707.1 GI:14602199

DBSOURCE REFSEQ: accession NC_000854.1

KEYWORDS .

SOURCE Aeropyrum pernix.

ORGANISM Aeropyrum pernix

Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales; Desulfurococceaceae; Aeropyrum.

1 (sites)

REFERENCE

AUTHORS Kawarabayasi,Y., Hino,Y., Horikawa,H., Yamazaki,S., Haikawa,Y., ...

TITLE Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, Aeropyrum pernix K1

JOURNAL DNA Res. 6 (2), 83-101 (1999)

MEDLINE 99310339

PUBMED 10382966

REFERENCE

2 (residues 1 to 51)

AUTHORS NCBI Microbial Genomes Annotation Project.

TITLE Direct Submission

JOURNAL Submitted (19-JUN-2001) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA

FEATURES

Location/Qualifiers

source 1..51

/organism="Aeropyrum pernix"

/db_xref="taxon:56636"

Protein 1..51

/product="S ribosomal protein L39E"

CDS 1..51

/gene="APE1087a"

/coded_by="complement(NC_000854.1:691740..691895)"

/transl_table=11

ORIGIN

1 marnkplgsk lrlaralksn raipwvvir terrirfnll rrwrrsklk v

//

Banques Quelques banques majeures

- Séquences
 - nucléiques
 - protéiques
 - mixtes
- ➔ • Structures
 - Familles, domaines et sites protéiques
 - Ontologie
 - Cluster de transcrits

Banques Les banques de structures

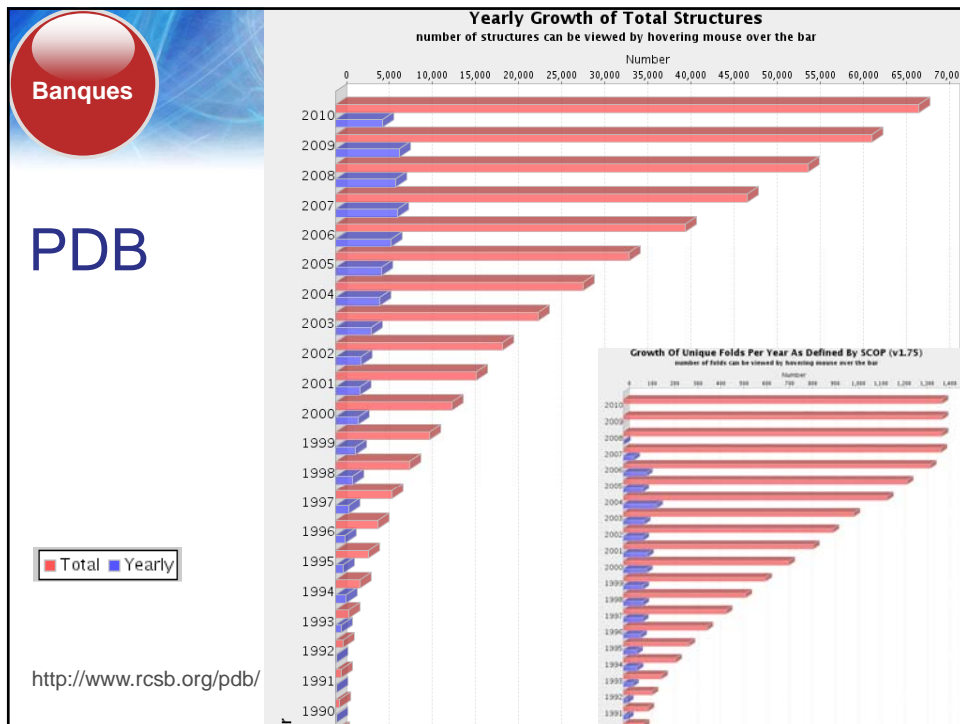
La Protein Data Bank (PDB)

Exp.Method	Proteins	Nucleic Acids	ProteinNA Complexes	Other	Total
X-RAY	54936	1227	2567	17	58747
NMR	7479	923	162	7	8571
ELECTRON MICROSCOPY	211	17	77	0	305
HYBRID	21	1	1	1	24
other	126	4	4	13	147
Total	62773	2172	2811	38	67794

- Séquences
- Structures secondaires
- Coordonnées atomiques

Dépôt obligatoire des coordonnées à la PDB avant publication
=> **LA** banque de structures

<http://www.rcsb.org/pdb/>



Banques

Banques dérivées

- **SCOP (Structural Classification Of Proteins)**

Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	284	507	871
All beta proteins	174	354	742
Alpha and beta proteins (a/b)	147	244	803
Alpha and beta proteins (a+b)	376	552	1055
Multi-domain proteins	66	66	89
Membrane and cell surface proteins	58	110	123
Small proteins	90	129	219
Total	1195	1962	3902

STRUCTURAL CLASSIFICATION OF PROTEINS

- **(a/b)** Mainly parallel beta sheets (beta-alpha-beta units)
- **(a+b)** Mainly antiparallel beta sheets (segregated alpha and beta regions)

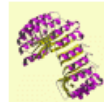
- CATH (Protein Structure classification)

Classification hiérarchique :

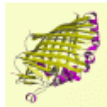
- Class (C)
- Architecture (A) : arrangement spatial des struct secondaires (sans regarder connexions)
- Topology (T) : prise en compte des connexions
- Homologous superfamily (H) : existence d'un ancêtre commun



Class 1: Mainly Alpha



Class 3: Mixed Alpha-Beta



Class 2: Mainly Beta



Class 4: Few Secondary Structures



Banques

- Séquences
 - nucléiques
 - protéiques
 - mixtes
- Structures
- ➔ ■ Familles, domaines et sites protéiques
- Ontologie
- Cluster de transcrits



PROSITE



- banque de **motifs** et de **profils** caractéristiques de domaines, de sites ou de familles de protéines
(Amos Bairoch, Swiss Institute of Bioinformatics)
- **Conservation** ⇒ **signification biologique** :
 - site actif d'une enzyme
 - site de fixation :
 - ions métalliques
 - ADN
 - ADP/ATP, GDP/GTP
 - protéine...
 - cystéines impliquées dans un pont disulfure...



Motif

- 1) alignement d'une famille
- 2) définition du motif
- 3) test sur Swiss-Prot

ABRU_DROME/544-568	RPDP.....KGGKIYRSANTTRTLEd.....KGE----
ABRU_DROME/573-601	YRQV.....LGGTVAKSINSHSHMSF.....QDRGIST
ACE1_TRIRE/400-429	KKQRE.....PQTRKGRPCDTRQEK.....MSSDPWK
ACE2_SCHPO/443-472	YVQLY.....NGGKRLARKYVVEHQLQ.....MISDRRP
ACE2_SCHPO/473-500	YRQD.....LGGAGVRRHDKQMLR.....MIRGRRP
ACE2_YEAST/603-632	FEELY.....PQGRVGRGRTLSRQLQ.....MIDRRP
ACE2_YEAST/633-662	YSDLE.....PQTRKAVRRHDITRKL.....SIRAKKY
ADNP_HUMAN/512-540	LSQP.....YRSTINDVEKMAADGR.....VLDDEEM
ADNP_MOUSE/231-259	LSQP.....YRSTINDVEKMAADGR.....VLDDEEM
ADNP_RAT/232-260	LSQP.....YRSTINDVEKMAADGR.....VLDDEEM
ADRI_YEAST/104-131	FVSG.....VETRAARQEHKQDER.....SIRNEKP
ADRI_YEAST/132-160	YPSG.....LQTRCIRRDLLDPAQK.....MISDHLG
AEF1_DROME/184-211	FQST.....VQDRRROLSTTRQWK.....MIRGEKP
AEF1_DROME/212-239	YKQH.....VQDKTROSSTTRMLK.....MIRGEKP
AEF1_DROME/240-267	YKQH.....FQPKLROLSTAMQWK.....MIRGEKP
AEF1_DROME/268-295	FEQV.....LQKQVROSSTTRMLK.....MIRMDKV
AIOL_HUMAN/118-145	HWQD.....VGLSCISFHVAVQKR.....SIRGRRP
AIOL_HUMAN/146-173	FQGH.....QGGASITQKGMLEMLK.....LIRGEKP
AIOL_HUMAN/174-201	FKQH.....LQNYACORRDAITGLR.....TISVEKE

Bon motif => retrouver toutes les protéines appartenant à la famille et uniquement ces protéines.

« Faux positif (false positive) » :

protéine prédite comme ayant le motif mais qui n'appartient pas à la famille

« Faux négatif (false negative) » :

protéine prédite comme n'ayant pas le motif mais qui appartient à la famille



Description du Pattern

■ Ambiguïtés :

- x => un acide aminé quelconque
- [ALT] => Ala ou Leu ou Thr
- {AM} => un acide aminé quelconque sauf Ala ou Met

■ Répétition d'un acide aminé

- x(3) : équivaut à x-x-x
- x(2,4) : équivaut à x-x ou x-x-x ou x-x-x-x

Exemples:

[AC]-x-V-x(4)-{ED}. => [Ala or Cys]-any-Val-any-any-any-any-{any but Glu or Asp}
 A-x-[ST](2)-x(0,1)-V. => Ala-any-[Ser or Thr]-[Ser or Thr]-(any or none)-Val



PROSITE (Prosite.dat)

```

ID ZINC_FINGER_C2H2_1; PATTERN.
AC PS00028;
DT APR-1990 (CREATED); JUN-1994 (DATA UPDATE); SEP-2006 (INFO UPDATE).
DE Zinc finger C2H2 type domain signature.
PA C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H.
NR /RELEASE=50.7,232345;
NR /TOTAL=8007(1371); /POSITIVE=7856(1251); /UNKNOWN=6(5);
NR /FALSE_POS=145(115); /FALSE_NEG=32; /PARTIAL=3;
CC /TAXO-RANGE=??E?V; /MAX-REPEAT=35;
CC /SITE=1,zinc; /SITE=3,zinc; /SITE=7,zinc; /SITE=9,zinc;
CC /VERSION=1;
DR Q24174, ABRU_DROME , T; Q9P8W3, ACE1_TRIRE , T; O14258, ACE2_SCHPO , T;
DR P21192, ACE2_YEAST , T; Q9H2P0, ADNP_HUMAN , T; Q9Z103, ADNP_MOUSE , T;
DR Q9JKL8, ADNP_RAT , T; P07248, ADR1_YEAST , T; P39413, AEF1_DROME , T;
..
3D 1YUI; 1YUJ; 1ARD; 1ARE; 1ARF; 1PAA; 2ADR; 1NCS; 1ZFD; 2DRP; 1ZAA; 1AAY;
3D 1A1F; 1A1G; 1A1H; 1A1I; 1A1J; 1A1K; 1A1L; 1SP1; 1SP2; 1ZNF; 2GLI; 1UBD;
3D 1ZNM; 3ZNF; 4ZNF; 1BBO; 1TF3; 1TF6; 5ZNF; 7ZNF;
DO PDOC00028;
//

```



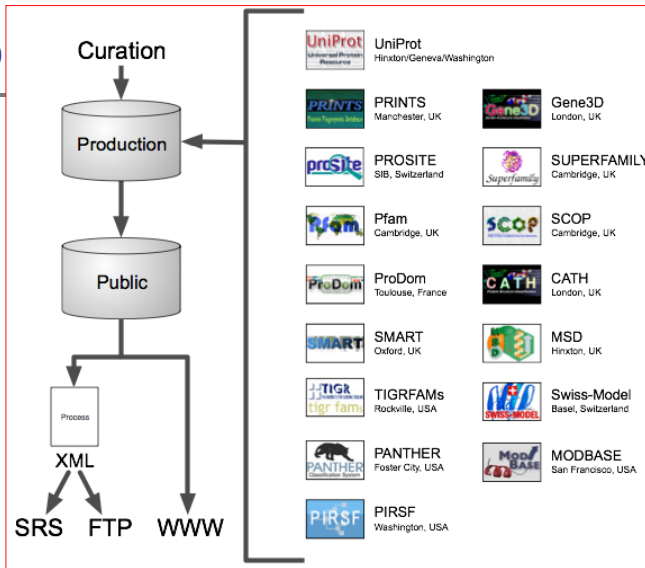
InterPro

- **InterPro** (<http://www.ebi.ac.uk/interpro>)
Integrated Resource of Protein Domains and Functional Sites
- **Banque qui intègre plusieurs banques de :**
 - motifs
 - domaines
 - familles protéiques
 - repeats
 => couvre ~80% des entrées Uniprot
- **Recherche**
 - par interrogation sur le texte
 - par InterProScan (comparaison d'une protéine à la banque)

<http://www.ebi.ac.uk/interpro>



InterPro

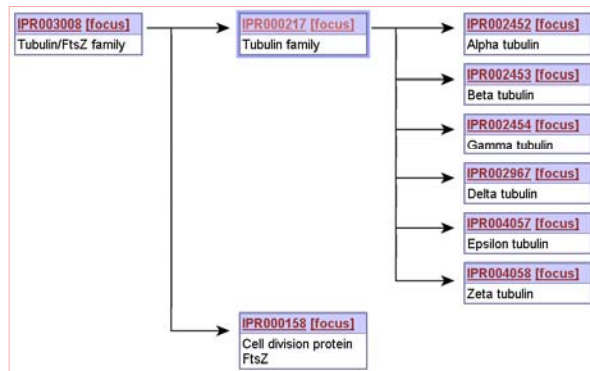


Organisation de la banque InterPro

■ Relations parents-enfants :

“**parent**” : motifs qui décrivent une famille

“**enfant**” : motifs spécifiques d'une sous-famille



Organisation de la banque InterPro

■ Relations “contains/found in” :

■ “found in” : un domaine peut être trouvé dans différentes familles

Ex: le domaine IPR000130 est trouvé dans 15 familles de protéines : neprilysin, astacin...

■ “contains” : une famille protéique peut être multidomaine

Database	InterPro
Accession	IPR001723; Strdhormone_receptor (matches 471 proteins)
Name	Steroid hormone receptor
Type	Family ⓘ
Dates	08-OCT-1999 (created) 15-JAN-2001 (last modified)
Signatures	PR00398; STRDHORMONER (471 proteins)
Contains ⓘ	IPR000536; Ligand-binding domain of nuclear hormone receptor (908 proteins) IPR001628; C4-type steroid receptor zinc finger (991 proteins)

Une entrée InterPro

InterPro IPR007087 Zinc finger, C2H2-type [Click here for help!](#)

Matches
 Overview: [sorted by AC](#), [sorted by name](#), [of known structure](#), [proteins with splice variants](#)
 Detailed: [sorted by AC](#), [sorted by name](#), [of known structure](#), [proteins with splice variants](#)
 Table: [For all matching proteins](#), [of known structure](#)
[Architectures](#)

Accession IPR007087 Znf_C2H2 Matches: 11312 proteins

Secondary IPR000822

Type Domain

Signatures

Database	ID	Name	Proteins
ProDom	PD000003	Znf_C2H2	4616
Pfam	PF00096	zF-C2H2	8276
PROSITE pattern	PS00028	ZINC_FINGER_C2H2_1	10860
PROSITE profile	PS0157	ZINC_FINGER_C2H2_2	8799
SMART	SM00355	ZnF_C2H2	11105


Children
[Tree](#)
 IPR007086 Zinc finger, C2H2-subtype

Found in
 IPKUU36U4 Zinc finger, U1-type
 IPR006865 Protein of unknown function DUF629
 IPR009170 Predicted retinoblastoma binding protein (RIZ)

Function
 GO:0003676 nucleic acid binding
 GO:0008270 zinc ion binding

Component
 GO:0005622 intracellular

...



Protein matches for protein BCL6_HUMAN(P41182) from the the UniProt/Swiss-Prot database.

InterPro Entry	Method accession	Graphical match	Method name
IPR000210	PF00651		BTB
IPR000210	PS50097		BTB
IPR000210	SM00225		BTB
IPR007087	PF00096		zF-C2H2
IPR007087	PS00028		ZINC_FINGER_C2H2_1
IPR007087	PS0157		ZINC_FINGER_C2H2_2
IPR007087	SM00355		ZnF_C2H2

Classification [PDB Chain/Domain ID & View 3D](#) [PDB Chain/Structural Domains](#)

Key:	Database	True match	False/Uncertain match
Prosite profile			
Pfam			
Smart			
PDB			
Prosite pattern			






Interproscan (sequence search)

YOUR EMAIL		RESULTS	
<input type="text"/>		interactive	
APPLICATIONS TO RUN <input type="radio"/> Clear all <input checked="" type="radio"/> Check all			
<input checked="" type="checkbox"/> BlastProDom	<input checked="" type="checkbox"/> FPrintScan	<input checked="" type="checkbox"/> HMMPIR	<input checked="" type="checkbox"/> HMMPfam
<input checked="" type="checkbox"/> HMMTigr	<input checked="" type="checkbox"/> ProfileScan	<input checked="" type="checkbox"/> ScanRegExp	<input checked="" type="checkbox"/> SuperFamily
<input checked="" type="checkbox"/> TMMHM	<input checked="" type="checkbox"/> HMMPanther	<input checked="" type="checkbox"/> Gene3D	<input checked="" type="checkbox"/> SignalPHMM
TRANSLATION TABLE (DNA/RNA only)		MIN. OPEN READING FRAME SIZE	
None		100	
Enter or Paste a <input type="text" value="PROTEIN"/> Sequence in any format: <input type="button" value="Help"/>			
Upload a file: <input type="text"/>		<input type="button" value="Parcourir..."/> <input type="button" value="Submit Job"/> <input type="button" value="Reset"/>	



Interproscan (sequence search)

InterPro search Results.

1 Query Sequence sw Length 937 aa.

InterPro	Results of PPsearch against PROSITE	Results of PFScan against PROSITE	Results of FingerPRINTScan against PRINTS	Results of HMMDecipher against PFAM-A
IPR001412 Aminoacyl-transfer RNA synthetases class-I	PS00178 [57-68]			
IPR002300 tRNA synthetase, class Ia				PF00133 [26-69]
IPR002301 Isoleucyl-tRNA synthetase			PR00984 [50-62] [237-261] [400-416] [525-539] [560-570]	
IPR002308 CysteinyI-tRNA synthetase				PF01406 [599-633]

IPR001412	PS00178	●
IPR002300	PF00133	●
IPR002301	PR00984	●
IPR002308	PF01406	●

Aminoacyl-transfer RNA synthetases class-I

tRNA synthetase, class Ia

Isoleucyl-tRNA synthetase

CysteinyI-tRNA synthetase



Banques

- Séquences
 - nucléiques
 - protéiques
 - mixtes
- Structures
- Familles, domaines et sites protéiques
- ■ Ontologie
- Cluster de transcrits



Gene Ontology (GO)

<http://www.geneontology.org/>

Ontologie :

spécification des concepts d'un domaine et de leurs relations

GO => appliqué à la connaissance du rôle des gènes et des protéines

Avantages :

La standardisation :

- la définition claire d'une fonction
- la génomique comparative au niveau fonctionnel

Les descriptions concises et non ambiguës peuvent être exploitées par un logiciel et par un humain...

AmiGO

Search GO

Exact Match
 Terms
 Gene Symbol/Name

Envooyer
 Advanced Query

Gene Product Filters

Species
 All
 A. aeolicus
 A. fulgidus

Datasource
 All
 FlyBase
 SGD

Evidence Code
 All
 Curator Approved
 IC

Set Filters

XML
 Flat File
 Permalink
 Last updated:
 2006-09-10

all : all (242881)

GO:0008150 : biological_process (190685)

- GO:0000004 : biological process unknown (37392)
- GO:0009987 : cellular process (115109)
- GO:0007275 : development (32144)
- GO:0040007 : growth (17773)**
 - GO:0035264 : body growth (563)
 - GO:0016049 : cell growth (797)
 - GO:0048589 : developmental growth (181)
 - GO:0030447 : filamentous growth (402)
 - GO:0048590 : non-developmental growth (32)
 - GO:0035265 : organ growth (63)
 - GO:0040008 : regulation of growth (16567)
- GO:0051704 : interaction between organisms (1722)
- GO:0021700 : maturation (0)
- GO:0007582 : physiological process (133101)
- GO:0043473 : pigmentation (128)
- GO:0050789 : regulation of biological process (33931)
- GO:0000003 : reproduction (22676)
- GO:0050896 : response to stimulus (18401)
- GO:0016032 : viral life cycle (428)
- GO:0005575 : cellular_component (140804)
- GO:0003674 : molecular_function (180826)
- obsolete_biological_process : obsolete_biological_process (221)
- obsolete_cellular_component : obsolete_cellular_component (95)
- obsolete_molecular_function : obsolete_molecular_function (2380)

<http://www.genedb.org/amigo/perl/go.cgi>

GoA

Annotation@EBI

The GOA project :
 to provide high-quality GO annotations to proteins in the UniProt Knowledgebase

GO Annotation Source	Number of Associations	Number of Distinct Proteins
Electronic GO annotation using InterPro to GO mapping	21309610	4095299
Electronic GO annotation using hamap to GO mapping :	504776	194922
Electronic GO annotation using Swiss-Prot keyword to GO mapping	8807230	2688032
Electronic GO annotation using UniProt Subcellular Location to go mapping:	587074	534762
Electronic GO annotation using EC to GO mapping	668970	655687
Electronic GO annotation using Compara projections :	147865	35511
Total Electronic GO annotation:	32025525	4322683
Total Manual GO annotation by Proteome Inc:	18179	5959
Total Manual GO annotation by UniProt:	79365	16712
Total Manual GO annotation:	467825	78668
Total UniProt GOA annotation:	32493350	4335231



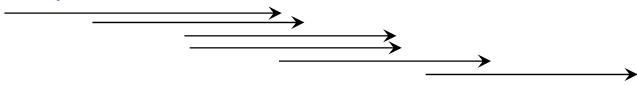
Banques

- Séquences
 - nucléiques
 - protéiques
 - mixtes
- Structures
- Familles, domaines et sites protéiques
- Ontologie
- Cluster de transcrits



UniGene

ORGANIZED VIEW OF THE TRANSCRIPTOME

- Banque de clusters de transcrits (regroupés par gène d'un même organisme)

- limiter la redondance des séquences d'ARNm et d'EST de GenBank
- meilleure information sur le gène
 - confrontation des EST chevauchantes
 - données sur l'expression du gène (tissus, stade de développement, conditions physiologiques)
 - données sur le splicing alternatif

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>

UniGene

1) Filtrage des séquences

- recherche des contaminants (*E. coli*, vecteurs de clonage)
- recherche des repeats, régions de faible complexité
- élimination des séquences de moins de 100 pb informatives

2) Clusters initiaux

- comparaison ARNm/ARNm

3) Ajout des EST

- comparaison EST/clusters initiaux
- Comparaison EST/EST

4) Réunions de clusters

2 clusters contenant des EST non chevauchants mais provenant d'un même clone (extrémité 5' et 3') sont réunis

5) Elimination de certains clusters

les clusters doivent contenir au moins une séquence correspondant à la région 3' du gène

Animaux:

- *Anopheles gambiae*
- *Bos taurus*
- *Danio rerio*
- *Drosophila melanogaster*
- *Homo sapiens*
- *Mus musculus*
- *Rattus norvegicus*
- *Xenopus laevis*...

Plantes:

- *Arabidopsis thaliana*
- *Hordeum vulgare*
- *Oryza sativa*
- *Triticum aestivum*
- *Zea mays*...

GENE EXPRESSION		UniGene Hs.449884 Homo sapiens complexin 4 CPLX4	
Tissues and development stages from			
Restricted Expression:	eye [show more like this] retinoblastoma [show more like this] adult (17 years old and older) [show more like this]		
Expression Profile:	View expression levels using UniGene's EST ProfileViewer [Show more entries with profiles like this]		
GEO profiles:	Gene expression profiles in the NCBI Gene Expression Omnibus database		
cDNA Sources:	eye, brain, uncharacterized tissue, muscle, pineal gland		
MAPPING POSITION			
Genomic location specified by transcript mapping, radiation hybrid mapping, genetic mapping or cytogenetic mapping.			
Chromosome:	18		
Map position:	18q21.32		
SEQUENCES			
Sequences representing this gene; mRNAs, ESTs, and gene predictions supported by transcribed sequences.			
mRNA sequences (9)		P Has similarity to known Proteins (after translation) A Contains a poly-Adenylation signal M Clone is putatively CDS-complete	
AY266502.1	Homo sapiens complexin IV mRNA, complete cds		
NM_181654.2	Homo sapiens complexin 4 (CPLX4), mRNA		
BX648518.1	Homo sapiens mRNA; cDNA DKFZp686A0185 (from clone DKFZp686A0185)		A
BX640655.1	Homo sapiens mRNA; cDNA DKFZp686O0683 (from clone DKFZp686O0683), complete cds		A
AK123184.1	Homo sapiens cDNA FL41190 fis, clone BRACE2044816		
BC065562.1	Homo sapiens cDNA clone IMAGE 6144778, partial cds		PA
BC106888.1	Homo sapiens complexin 4, mRNA (cDNA clone MGC:125784 IMAGE:40030309), complete cds		
BC106887.1	Homo sapiens complexin 4, mRNA (cDNA clone MGC:125783 IMAGE:40030305), complete cds		
BC106886.1	Homo sapiens complexin 4, mRNA (cDNA clone MGC:125769 IMAGE:40029631), complete cds		
EST sequences (10 of 16) [Show all sequences]			
BX510095.1	Clone DKFZp686B1896	muscle	5' read P
DA089405.1	Clone BRACE2044816	brain	5' read
DB298248.1	Clone BRACE2044816	brain	3' read
H92112.1	Clone IMAGE:221503	eye	5' read
BG472915.1	Clone IMAGE:4646600	eye	5' read
BM661830.1	Clone UI-E-CK1-abn-b-01-0-UI	eye	3' read A
BM661846.1	Clone UI-E-CK1-abn-d-02-0-UI	eye	3' read A
BM689790.1	Clone UI-E-CK1-abn-b-01-0-UI	eye	5' read
BM689812.1	Clone UI-E-CK1-abn-d-02-0-UI	eye	5' read
BU187675.1	Clone IMAGE:6144778	eye	5' read M



Conclusions

➔ Une quantité énorme de données hétérogènes

- ✓ Des banques généralistes incontournables :
Genbank / EMBL / DDBJ
Uniprot, Refseq
PDB
- ✓ Effort d'intégration des données
(ex : Interpro, Go, banques génomiques)
- ✓ Effort de validation (Swissprot, RefSeq)
- ✓ Effort de diminution des redondances
(ex : RefSeq, Unigene)