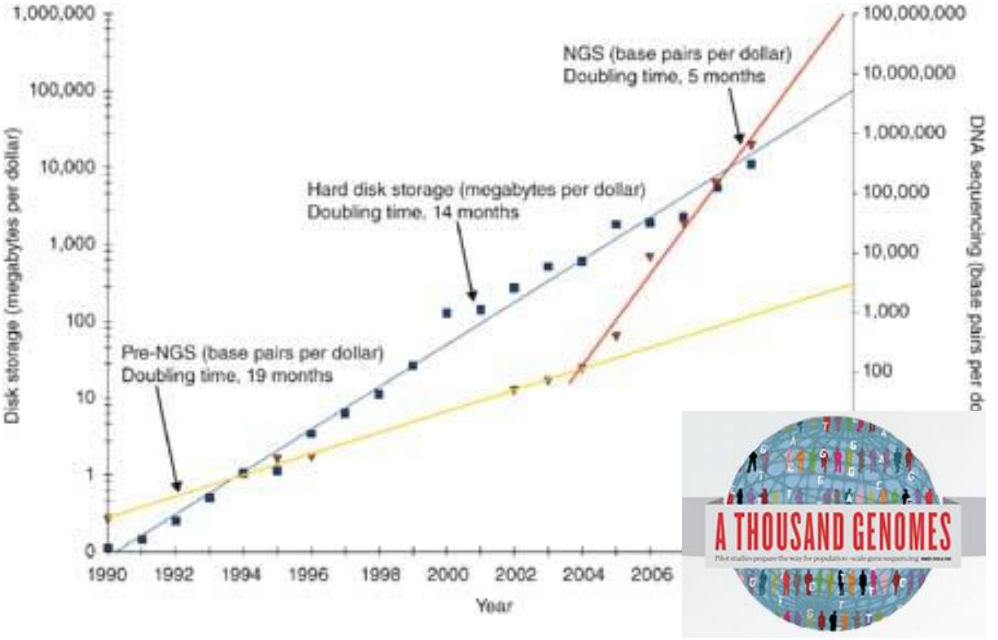
Proteomika: adatbázisok

Gáspári Zoltán, 2019

gaspari.zoltan@itk.ppke.hu

Miért kellenek adatbázisok?

- Adatok forrása bioinformatikai kutatásokhoz:
  - adatbázisok
  - kollaborációk (majdnem mindig kiegészítve adatbázisokból származó adatokkal a tágabb kontextusba helyezés céljából)
- A tárolandó és rendszerezendő adatmennyiség hatalmas:

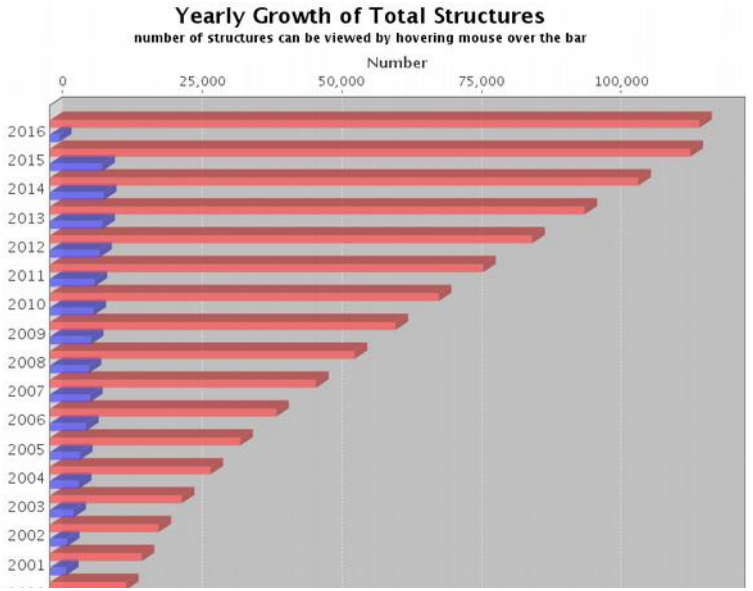


DNS szekvenciák:

- 1990: 25 ezer bp / hét
- 2000: 5 millió bp / hét
- 2010: 250 milliárd bp / hét (NGS)
- komplexitás: NGS eredmények többszörös illesztésként tárolva
- 1000 USD/emberi genom már realitás!

fehérjeszerkezetek:

- > 9 000 új szerkezet 2015-ben (összesen > 110 000)
- Növekvő méret és komplexitás (RNS-fehérje komplexek stb.)
- Dinamika mint új jelleg



- Fehérje-fehérje kölcsönhatások: adatbázisonként nagyságrendben 100 ezer
- > 23 millió kivonat tudományos közleményekből

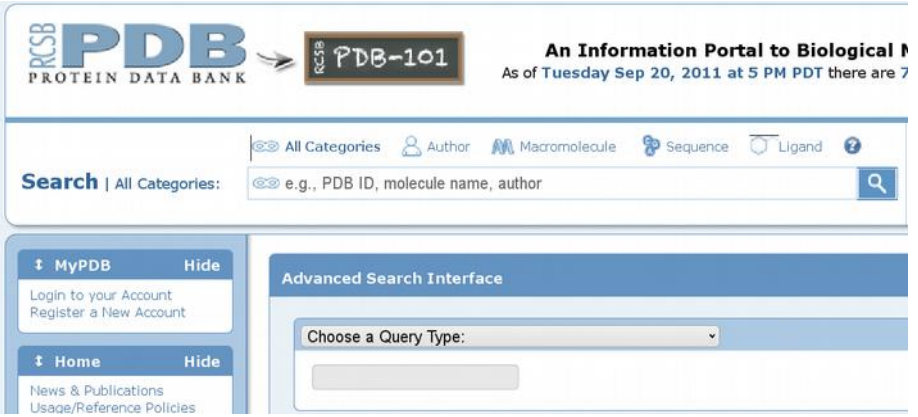
Mi van az adatbázisokban?

- Amit a kutatók beleraknak
- Az adatok nem lesznek attól megbízhatóak, hogy bekerülnek az adatbázisba
- Az adatbázisok messze vannak a teljességtől:
  - Tudásunk hiányos
  - Nem minden publikált adat kerül be (lustaság, időhiány, szándékosság)
  - (És viszont: nem minden adatbázis rekordhoz van publikáció...)
- Az adatbázisok hibákat is tartalmaznak
  - Emberek vagyunk...
  - Kísérleti hibák (van, amit nehéz kiszűrni/észrevenni!)
  - Egy kísérlet komoly ellenőrzése csak a reprodukálásával lehetséges – erre nyilvánvalóan nincs kapacitás
  - Tudatos csalások (pl. H.M. Krishna Murthy)
- Ki teszi be az adatokat?
  - Bárki, aki produkálta őket: elsődleges adatbázisok
    - pl. DNS szekvenciák, fehérjeszerkezetek
    - Az adatok minősége leginkább a kutató hozzáállását tükrözi (természetesen vannak törekvések az ellenőrzésre stb.)
    - Az adatok nyilvános adatbázisban való elhelyezését a legtöbb tudományos folyóirat megköveteli, de a kutatók megtalálják a módját, hogy ha nem akarják, nem teszik be (vagy az igazán érdekes részeket elmismásolják)
  - Kurátorok: másodlagos adatbázisok, illetve egyes elsődlegesek esetében annotáció
    - Adatok elsődleges adatbázisokból/irodalomból
    - Annotálás, esetleg részletes ellenőrzés
    - Adott esetben komoly automatizálás is lehet, kézi beavatkozás csak szükség esetén

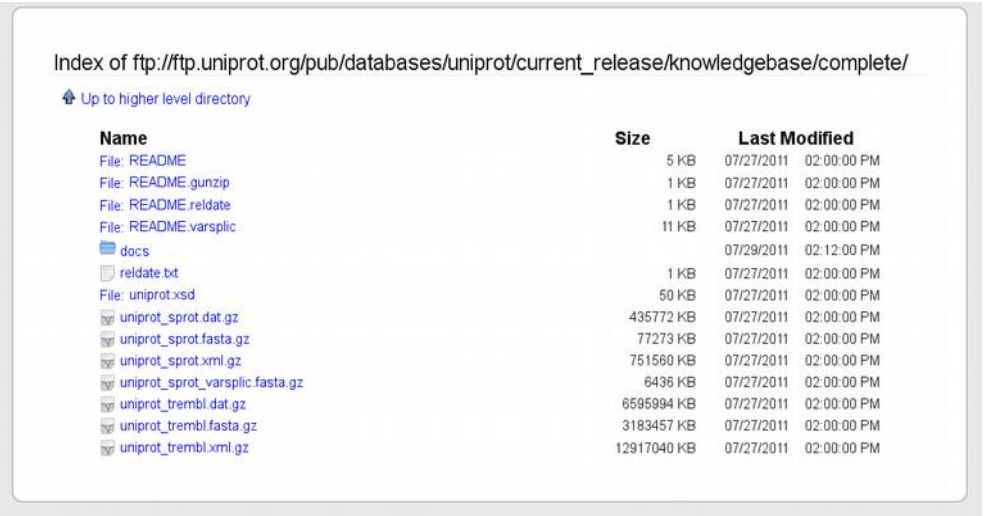
Adatbázisok működése és tudományetika

- Hozzáférés az adatokhoz
  - Teljesen szabad: nonprofit szervezetek, közösségi források segítségével fenntartott adatbázisok esetén
  - fizetős: céges adatbázisok.
  - vegyes: bizonyos részek ingyen, a teljes adatbázis pénzért – bevett gyakorlat
- Ki az adatok tulajdonosa?
  - Közösségi forrásokból finanszírozott kutatások esetén az adatokat általában kötelező szabad hozzáférésű adatbázisokba tenni
  - Céges adatbázisokból származó adatokat nem lehet akárkivel megosztani
- Az adatok felhasználása
  - Ha az adatok egyszer hozzáférhetőek lettek, a szerzőknek nincs beleszólásuk, ki és mire használja őket
  - Az adatok generálása általában nagyobb munka, mint az elemzésük (de az NGS esetében már ez fordítva igaz!), de az elemzés teszi az adatokat biológialag értelmezhetővé: konfliktus a kísérleteket és a feldolgozást végző kutatók között
  - embargó: az adatokat elhelyezik, de adott ideig nem hozzáférhetőek ill. nem használhatóak fel tudományos közleményekben
  - Valós eset: 2009 augusztusában megjelent egy cikk, ami Laura Bierut csoportjának a dbGaP (database of genotypes and phenotypes) adatbázisban elhelyezett adatait használta, bár az adatok szeptemberig embargó alatt voltak (mivel a cikket márciusban küldték be, az embargót 6 hónappal sértették meg)

Néhány gyakorlati szempont



- Egy vagy néhány adatbázisbejegyzés (rekord) használata:
  - Érdemes a webes keresőfelületet használni
  - Általában működik, a kívánt adat megtalálható, letölthető stb.
  - Részletes minőség-ellenőrzés lehetséges
- Nagy adatmennyiség elemzése
  - Meg lehet próbálni a webes keresést, de jó eséllyel nem teljes listát kapunk, elszalasztunk valamit, vagy csak túl sok lesz a találat
  - Ilyenkor sokszor szükséges a teljes adatbázis letöltése helyi elemzéshez, DE sokszor amit kapunk, távolról sem adatbázis, csak szöveges állományok
  - Akár magunknak kellhet belőle „igazi” adatbázist csinálni a hatékony munkához
  - Minőség-ellenőrzés nem könnyű, egyszerű automatizált megoldás kell
- Adatbázisok verziói
  - A legtöbb adatbázist többé-kevésbé rendszeresen frissítik
  - Ha túl ritkán, még nem lesz bent, amit keresünk
  - Ha túl gyakran, mire befejezzük az elemzést, új változatok lesznek kint, fontos, hogy legyen meg helyben az a verzió, amivel dolgoztunk, hogy ellenőrizni tudjuk az eredményeinket később is!



Adatbázisok és fájlformátumok

- Sokféle adatbázis még több féle adatot tárol
- Sok adatbázisnak saját adatformátuma van
  (megadhatnak egyes adatokat gyakran használt formátumban is)
- Több formátumot az adatbázisokról neveztek el (GenBank, PDB, EMBL...)
- Formátumkonverzió fontossága:
  - Csak első ránézésre egyszerű
  - Adatvesztés: csak az marad(hat) meg, amit mindkét formátum képes reprezentálni, ténylegesen szerepel a bemenetben ÉS a konverter is kezeli!
  - A formátumok összetettsége miatt nem mindig könnyű saját konvertert írni – ha már létezik elérhető, használjuk azt, de körültekintéssel!
  - Ritkán használt szintaktikai megoldások, kulcsszavak gondot okozhatnak a saját programjainknak is, de akár a „hivatalosaknak” is.
- Valós eset: FT ismétlődések
  - Egy kolléga kérte, hogy nézzek utána egy adott szekvenciában periodikusan előforduló Phe-Thr (FT) aminosavpárok jelentőségének
  - Kiderült, hogy az FT jelenléte hibás konverzió eredménye, az EMBL formátum 'feature' sorait vezeti be

```
FT                   /translation="EEKYTTMFDNVNLDEILANDRLLNNYVKCLLEDGEANCTADGKEL
FT                   KKAVPDALSNECAKCNDKQKEGTKKVLKHLINHKPDIWAQLKAKYDPDGTYSKKYEDKE
FT                   KELHE"
```

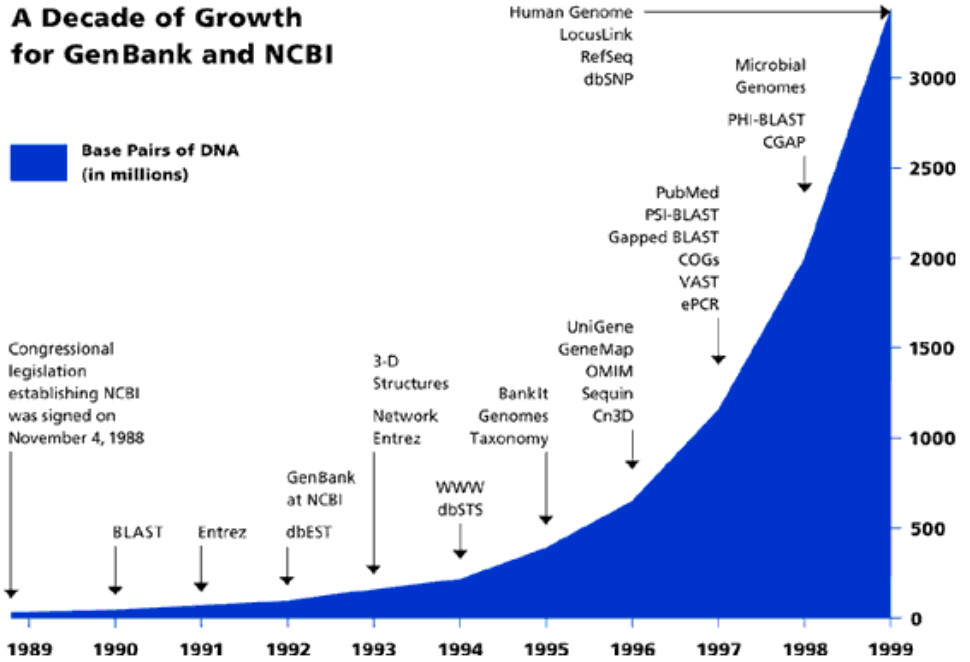
Egy EMBL formátumú file 'feature' részének részlete az adott DNS-szekvencia fordításával

Adatbázisok fajtái

- Nyers adatok (pl. a TrEMBL automatikusan fordított fehérjeszekvenciái)
- Annotált szekvenciák (Swiss-Prot, Uniprot)
- A nyers és az annotált adatbázisok is elsődlegesek
- A másodlagos adatbázisok tipikusan elsődlegesekből vett adatokon alapulnak.
- Szekvenciák esetén: a PFAM fehérjedomének, HMM-ek és illesztések gyűjteménye, a COG BLAST hasonlóság alapján csoportosított bakteriális szekvenciáké, az SBASE fehérjedomének kollekciója BLAST alapokon. Ezek mindegyike UniuProt alapú.
- 3D szerkezetek esetén: a SCOP és a CATH 3D szerkezetek hierarchikus adatbázisa, a PDB-n alapulnak. A SCOP főleg maunálisan, a CATH főleg automatikusan annotált / készített adatbázis

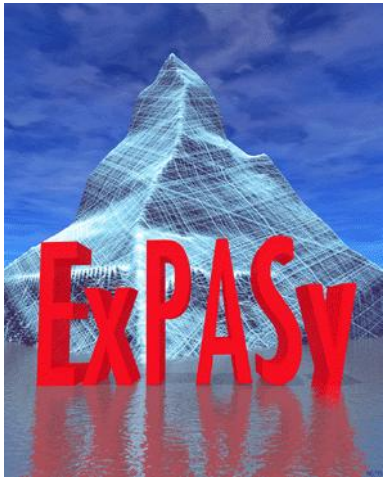
Bioinformatikai portálok: integrált források (adatok, eszközök stb.)

- NCBI: National Center for Biotechnology Information

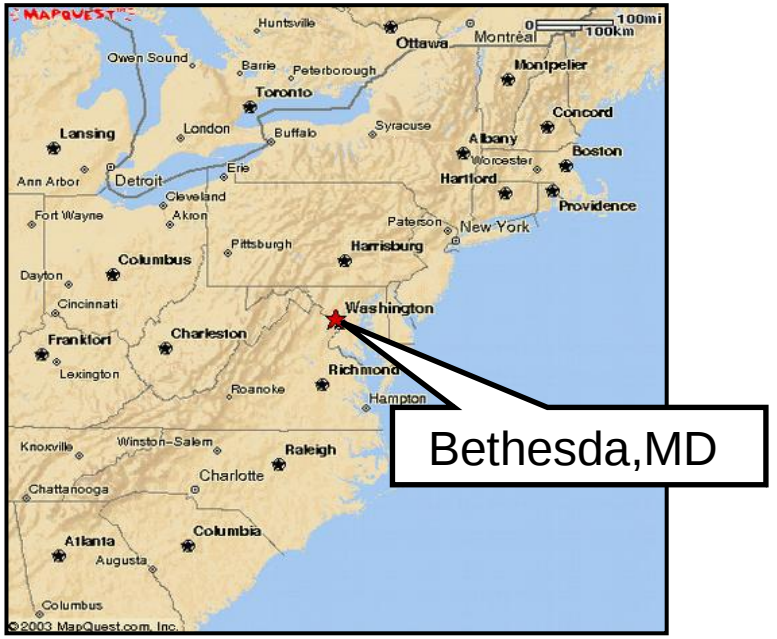


- EBI: European Bioinformatics Institute

- ExPASy: a Swiss Institute for Bioinformatics portálja
  - a UniProt konzorcium tagja



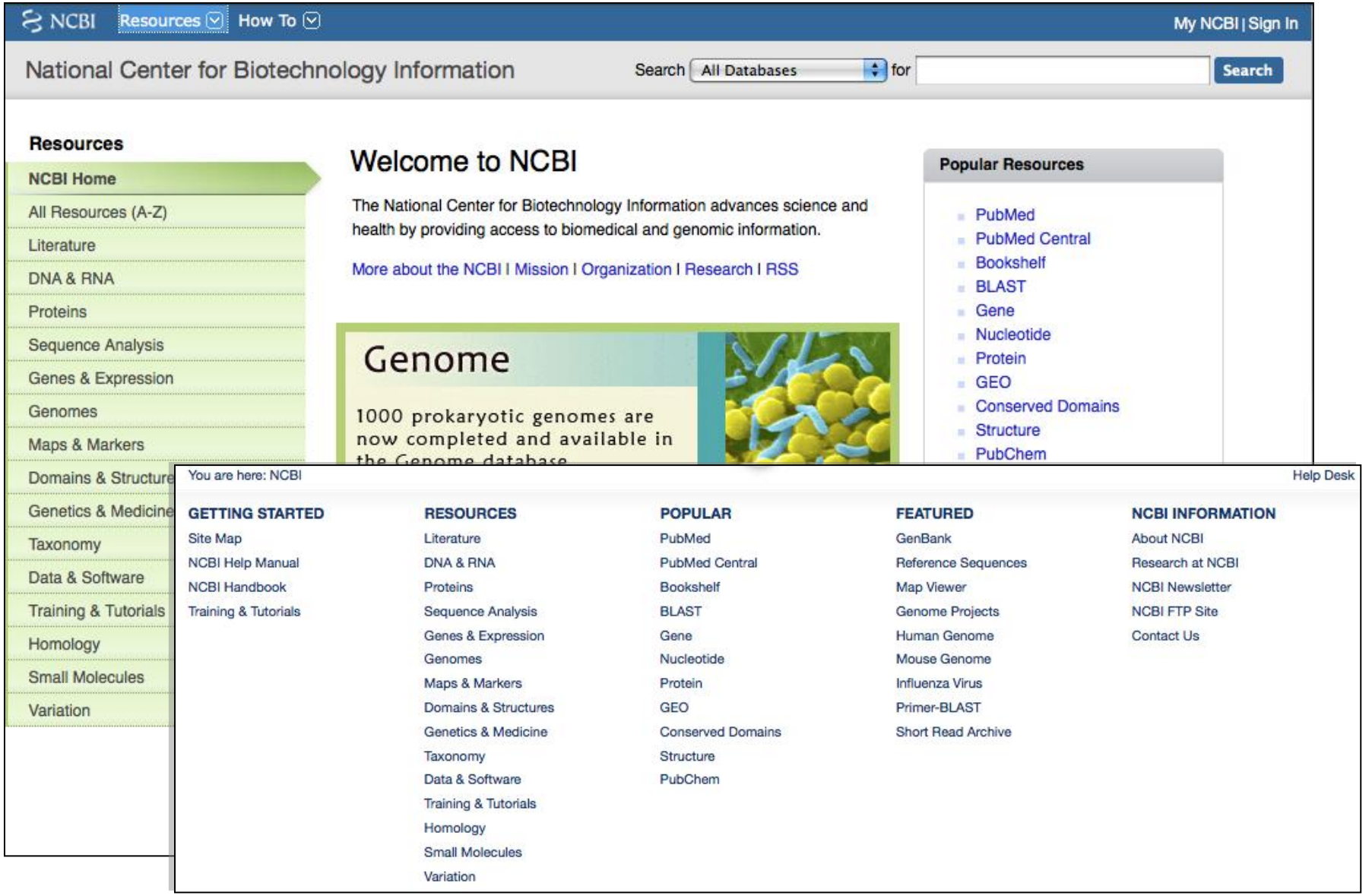
NCBI: The National Center for Biotechnology Information



1988-ban hozták létre a
National Library of Medicine, NIH részeként

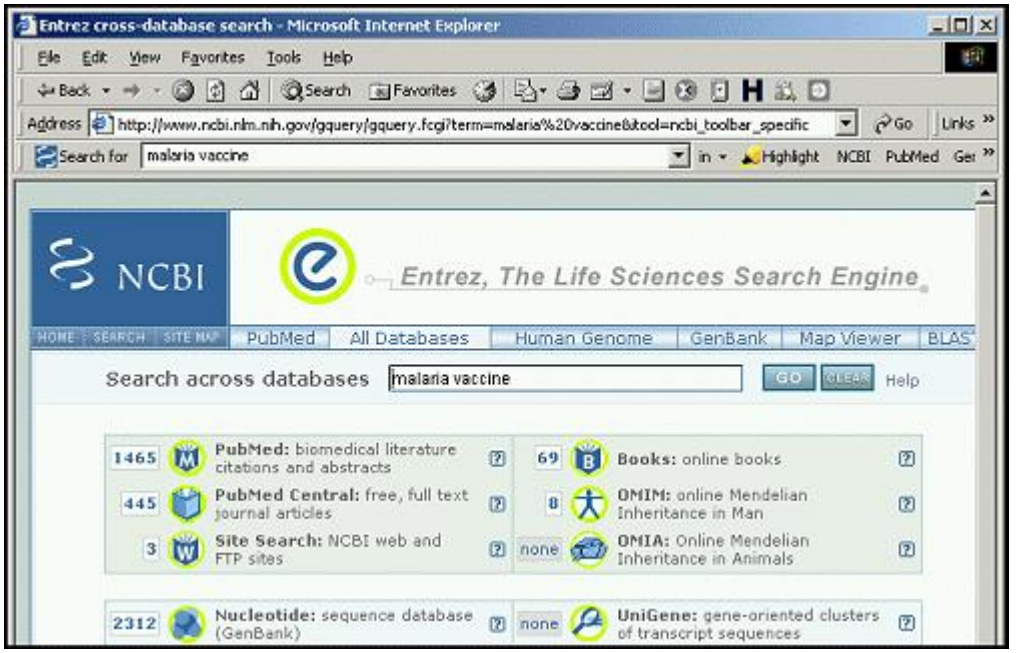
- Publikus adatbázisok létrehozása
- Számítógépes biológiai kutatások
- Programfejlesztés szekvenciaelemzéshez
- Információmegosztás

NCBI web: *www.ncbi.nlm.nih.gov*



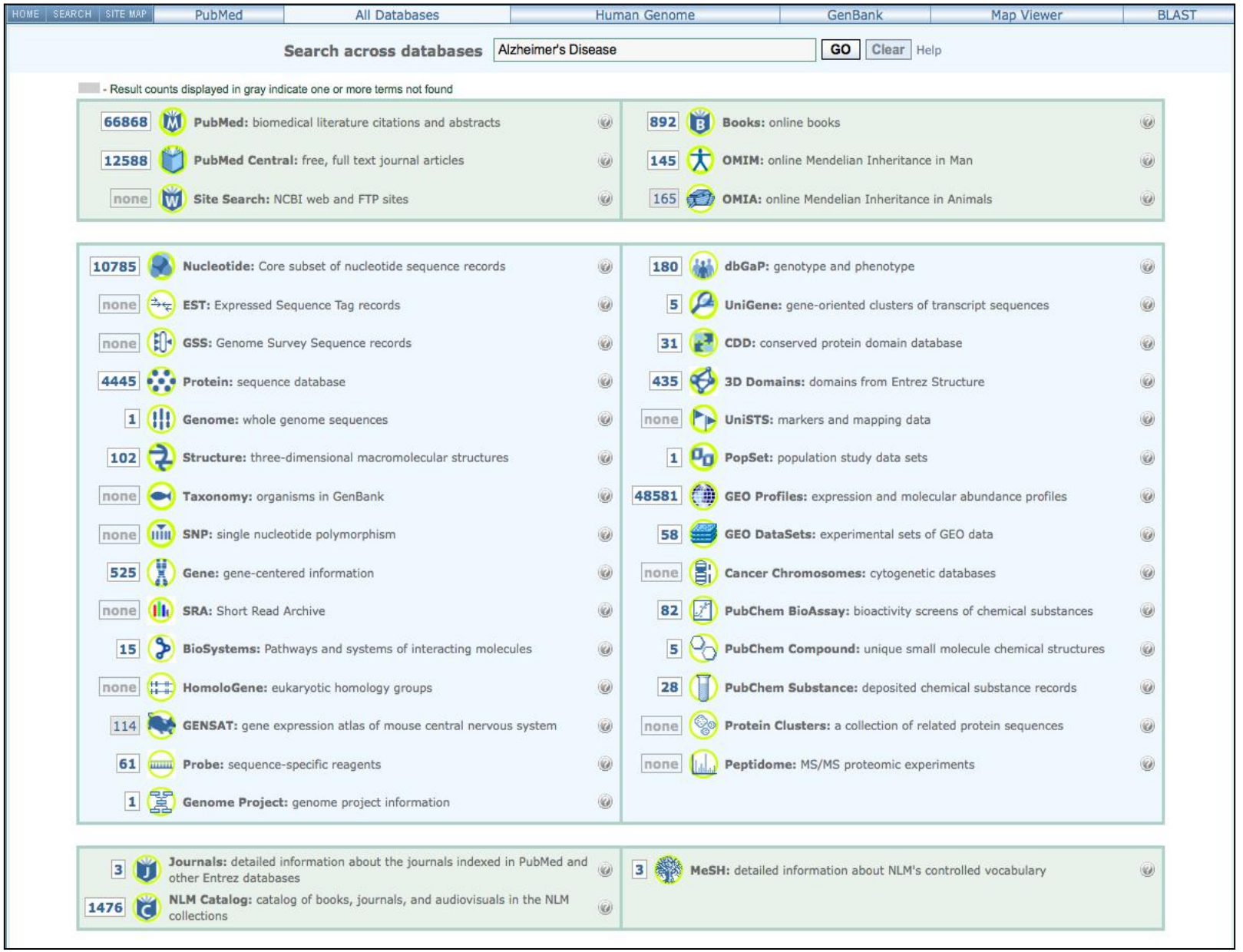
Néhány NCBI adatbázis és szolgáltatás

- GenBank: elsődleges (DNS) szekvencia adatbázis (NGS adatok mennyiségével jelentősége valamelyest csökken, bár fontos referencia maradt!)
- Szabad hozzáférés biomedicinális irodalomhoz
  - PubMed: szabad Medline (3 millió keresés naponta)
  - PubMed Central: Ingyenes teljes szövegű hozzáférés (open access publikációk + a NIH által támogatott kutatásokból származó, kötelezően beteendő cikkek)
- Entrez: integrált molekuláris és irodalmi adatbázisok



- BLAST: leggyakrabban használt szekvenciahasonlóság-kereső szolgáltatás (100 – 200 ezer keresés naponta)

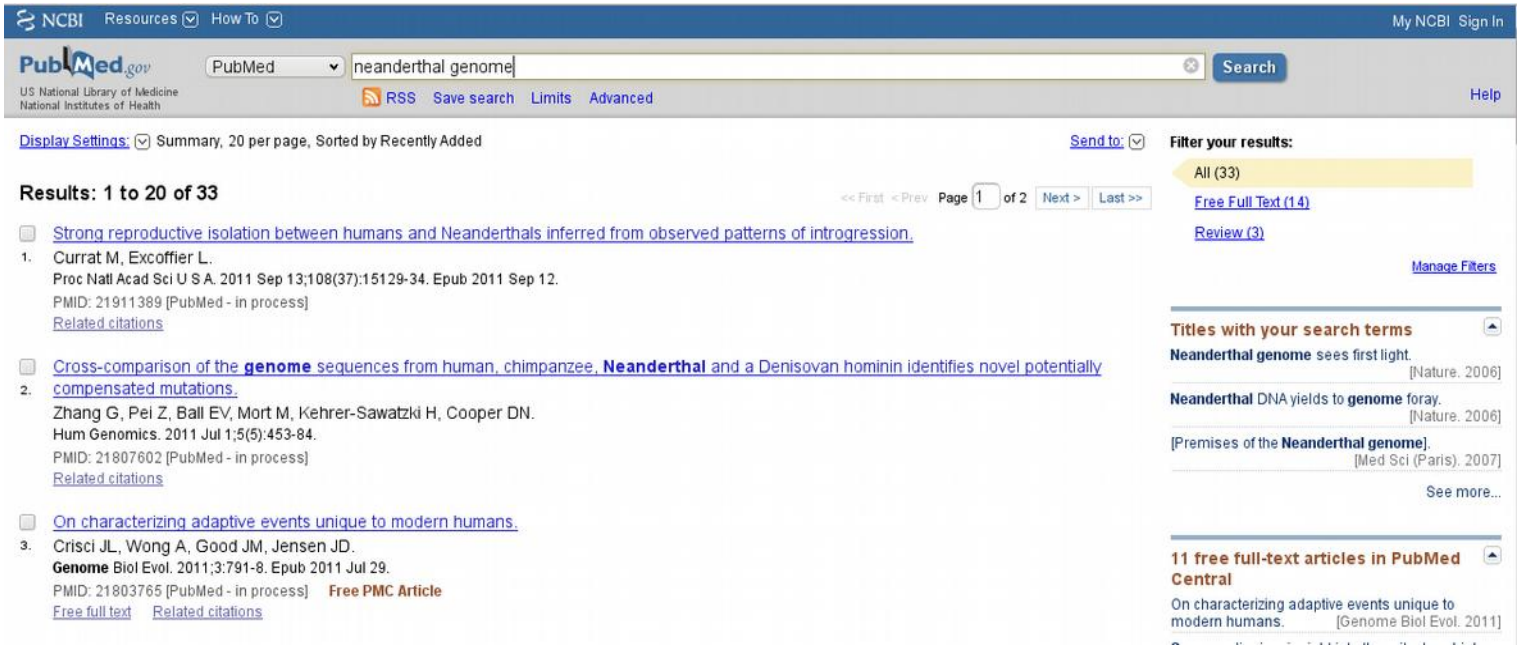
Entrez: az összes NCBI adatbázis egyidfejű keresése



- 40 (és növekvő számú) integrált adatbázis

PubMed: cikkek kivonatai (biomedicinális területen)

> 21 millió kivonat összetett kereshetősége



Kulcsszó: ”neanderthal genome” (más eredmény, mint a ”neandertal genome” kulcsszóra!)

Az ExPASy szerver

http://www.expasy.org

- Az első molekuláris biológiai szerver (1993 augusztus), proteomikai fókusszal
  - Adatbázisok: **UniProtKB**, **PROSITE**, ENZYME, **Swiss-2DPAGE**, stb.;
  - Sokféle 2D/MS fehéjeazonosító/jellemző és szekvenciaelemző eszköz

- Szekvenciaelemzés (Blast, ScanProsite, ProtParam, ProtScale, RandSeq, Translate, etc.);
- Proteomika (AACompIdent, FindMod, FindPept, Aldente, PeptideMass, TagIdent, etc.);
- 3D szerkezetelemzés és -megjelenítés (Swiss-Model, Swiss-PDBviewer)

Az ExPASy WWW szerver: genomikai és proteomikai eszközök és adatbázisok

Visual Guidance

Categories

- proteomics
- **genomics**
  - sequence alignment
  - similarity search
  - characterisation/annotation
- structural bioinformatics
- systems biology
- phylogeny/evolution
- population genetics
- transcriptomics
- biophysics
- imaging

Databases

- STRING • protein-protein interactions • [more]
- EPD • collection of eukaryotic promoters • [more]
- SwissRegulon • annotations of regulatory sites • [more]
- smirnaDB • miRNA expression profiles analysis • [more]

- CLIPZ • binding sites of RNA-binding proteins • [more]
- EIMMo • miRNA target predictions • [more]
- GPSDB • gene and protein synonyms • [more]
- ImmunoDB • insect immune-related genes and gene families • [more]
- miROrtho • catalogue of animal microRNA genes • [more]
- MyHits • protein domains database and tools • [more]
- OMA • orthology inference among complete genomes. • [more]
- OpenFlu • Influenza genetic and epidemiological data • [more]
- OrthoDB • Hierarchical catalog of eukaryotic orthologs • [more]

Tools

- EPD • collection of eukaryotic promoters • [more]
- smirnaDB • miRNA expression profiles analysis • [more]

- Association Viewer • SNPs display in a genetic context • [more]
- BayeScan • identify natural selection • [more]
- BLAST • sequence similarity search • [more]
- boxshade • MSA pretty printer • [more]
- ChIP-Seq • ChIP-Seq data analysis tools • [more]
- CLIPZ • binding sites of RNA-binding proteins • [more]
- Codon Suite • codon-based sequence analysis • [more]
- Dotlet • sequence similarity plots • [more]
- EIMMo • miRNA target predictions • [more]
- EMBnet services • bioinformatics tools and databases • [more]
- ESTscan • coding region detection • [more]

Visual Guidance

Categories

- **proteomics**
  - protein sequences and identification
  - mass spectrometry and 2-DE data
  - protein characterisation and function
  - families, patterns and profiles
  - post-translational modification
  - protein structure
  - protein-protein interaction
  - similarity search/alignment
- genomics
- structural bioinformatics
- systems biology
- phylogeny/evolution
- population genetics
- transcriptomics
- biophysics
- imaging

Databases

- UniProtKB • functional information on proteins • [more]
- UniProtKB/Swiss-Prot • protein sequence database • [more]
- STRING • protein-protein interactions • [more]
- SWISS-MODEL Repository • protein structure homology models • [more]
- neXtProt • human proteins • [more]
- PROSITE • protein domains and families • [more]
- ViralZone • portal to viral UniProtKB entries • [more]

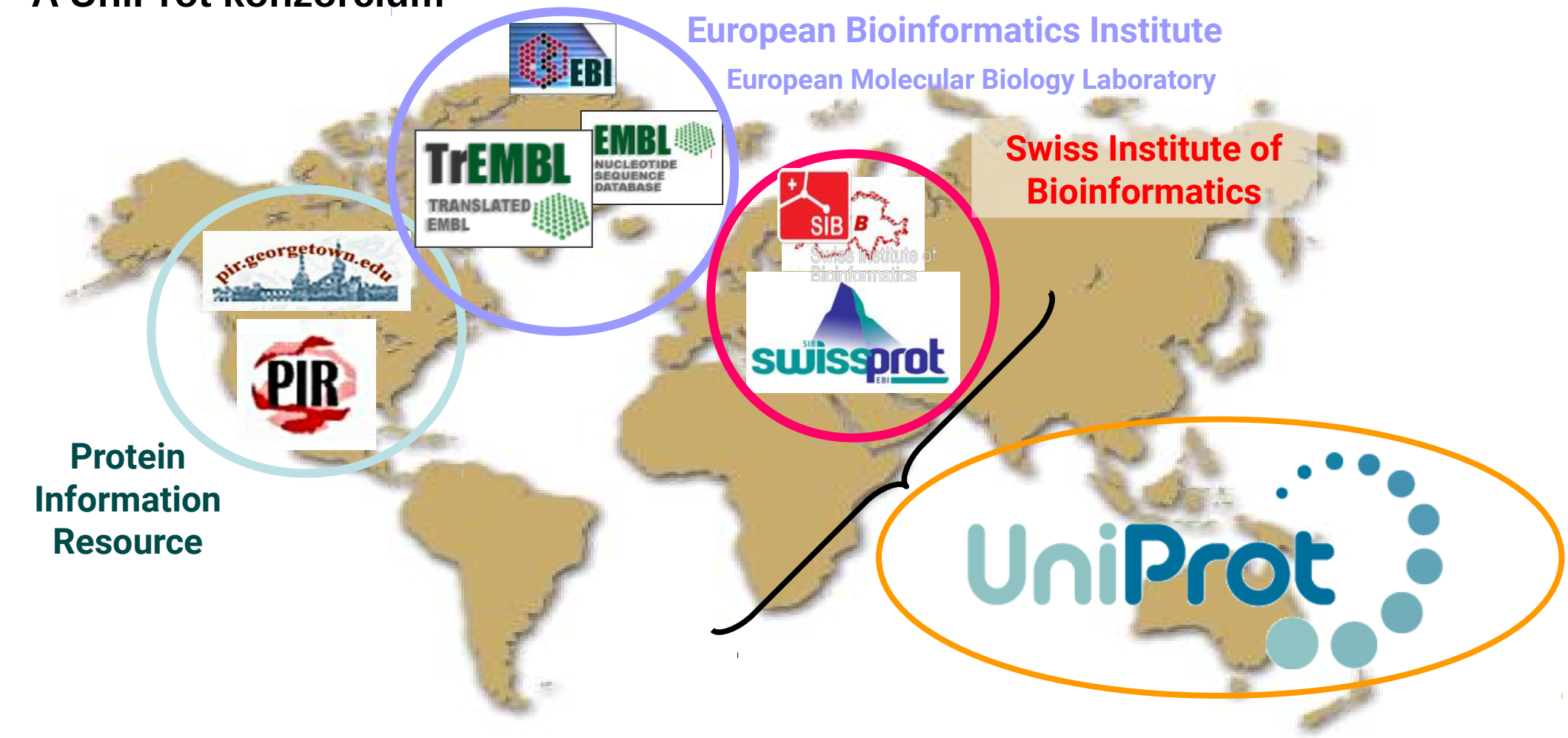
- ENZYME • nomenclature of enzymes • [more]
- GlycoSuiteDB • glycan database • [more]
- GPSDB • gene and protein synonyms • [more]
- HAMAP • manual protein annotation • [more]
- MIAPEGelDB • MIAPE document edition • [more]
- MyHits • protein domains database and tools • [more]
- PaxDb • protein abundance database • [more]
- Prolune • Popular science articles (in French) • [more]
- Protein Model Portal • structural information for a protein • [more]
- Protein Spotlight • Informally written reviews on proteins • [more]

Tools

- SWISS-MODEL Workspace • structure homology-modeling • [more]
- SwissDock • protein ligand docking server • [more]

- AACompSim • amino acid composition comparison • [more]
- AllAll • protein sequences comparisons • [more]
- BLAST • sequence similarity search • [more]
- boxshade • MSA pretty printer • [more]
- Compute pI/MW • theoretical pI and Mw computation • [more]
- Dotlet • sequence similarity plots • [more]
- EMBnet services • bioinformatics tools and databases • [more]
- FindMod • protein post-translational modifications • [more]
- FindPept • peptide identification from unspecific cleavage • [more]
- GlycanMass • oligosaccharide structure mass calculation • [more]
- GlycoMod • oligosaccharide structure prediction • [more]
- HAMAP • manual protein annotation • [more]
- HamapScan • scan sequences against HAMAP • [more]
- HCD/CID spectra merger • combine HCD and CID MS/MS spectra • [more]
- ImageMaster / Melanie • software for 2-D PAGE analysis • [more]

A UniProt konzorcium



UniProt (Universal Protein Resource): a világon a legteljesebb fehérjekatalógus
http://www.uniprot.org

- UniProt Knowledgebase
- UniRef clusters (100/90/50% azonosság)
- UniParc (UniProt Archive)
- UniMES (Metagenomic and Environmental Sequences)

UniProt adatbázisok

UniProt Knowledgebase

UniProtKB::

UniProtKB/TrEMBL
Automatikusan annotált fehérjeszekvenciák (>50 millió)

UniProtKB/Swiss-Prot
Kutatók által annotált fehérjeszekvenciák (kb. 500 ezer)

UniRef100
UniRef90
UniRef50

· Egy **UniRef1**00 bejegyzés **= Minden azonos szekve**cia (fragmensek is).

· Egy **UniRef90** bejegyzés = Legalább **90%-os azonosságot mutató szekvenciák**

·

Fajtól függetlenül

UniParc

UniProt Archives:

Archivált nyers fehérjeszekvenciák, publikus adatbázisokból

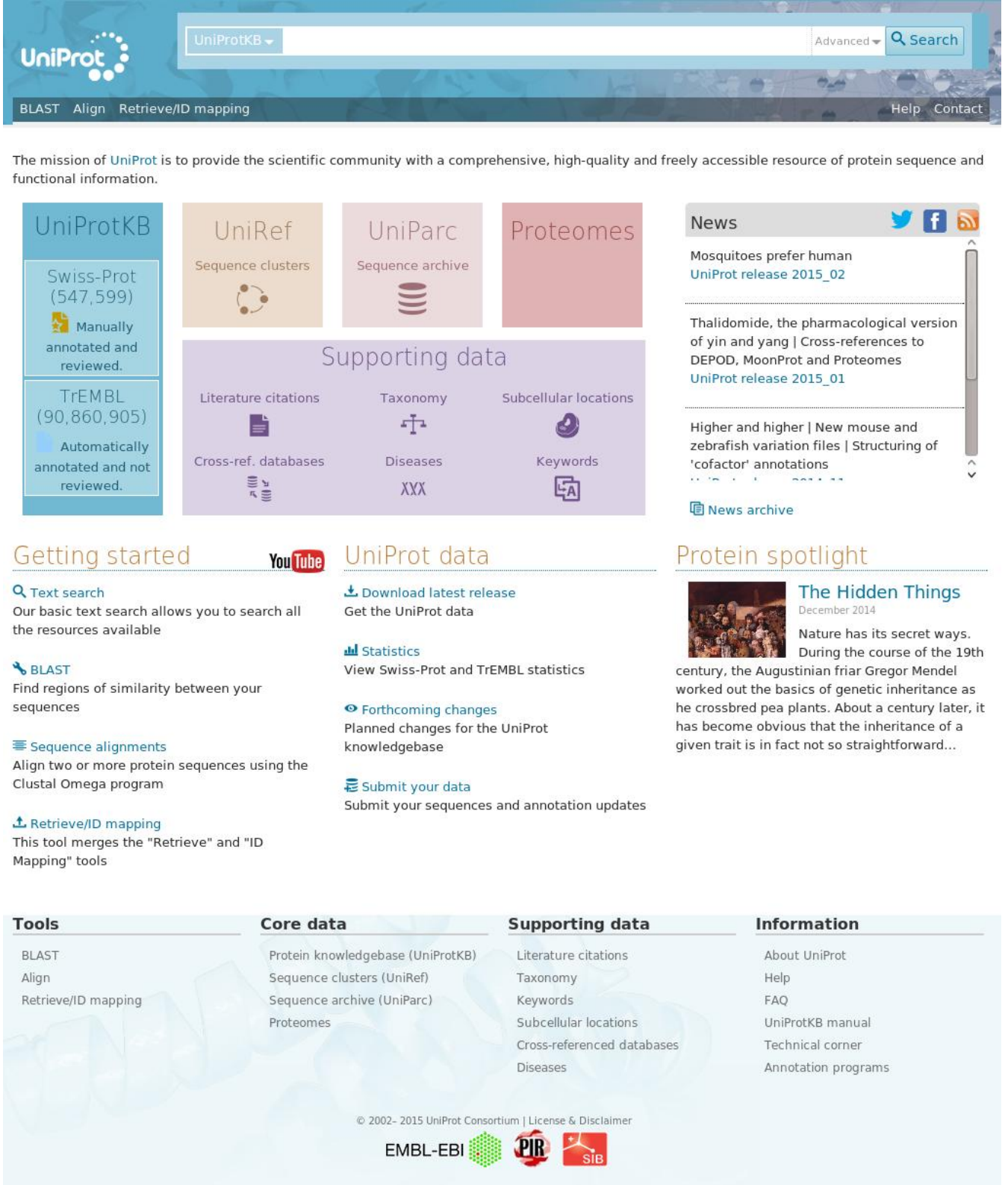
Swiss-Prot, TrEMBL, PIR, EMBL, Ensembl, IPI, PDB, RefSeq, FlyBase, WormBase, Patent Offices.

Körültekintés szükséges: pszeudogének, inkorrekt CDS predikciók stb

+ UniMES metagenomikai és környezeti minták

A UniProt website

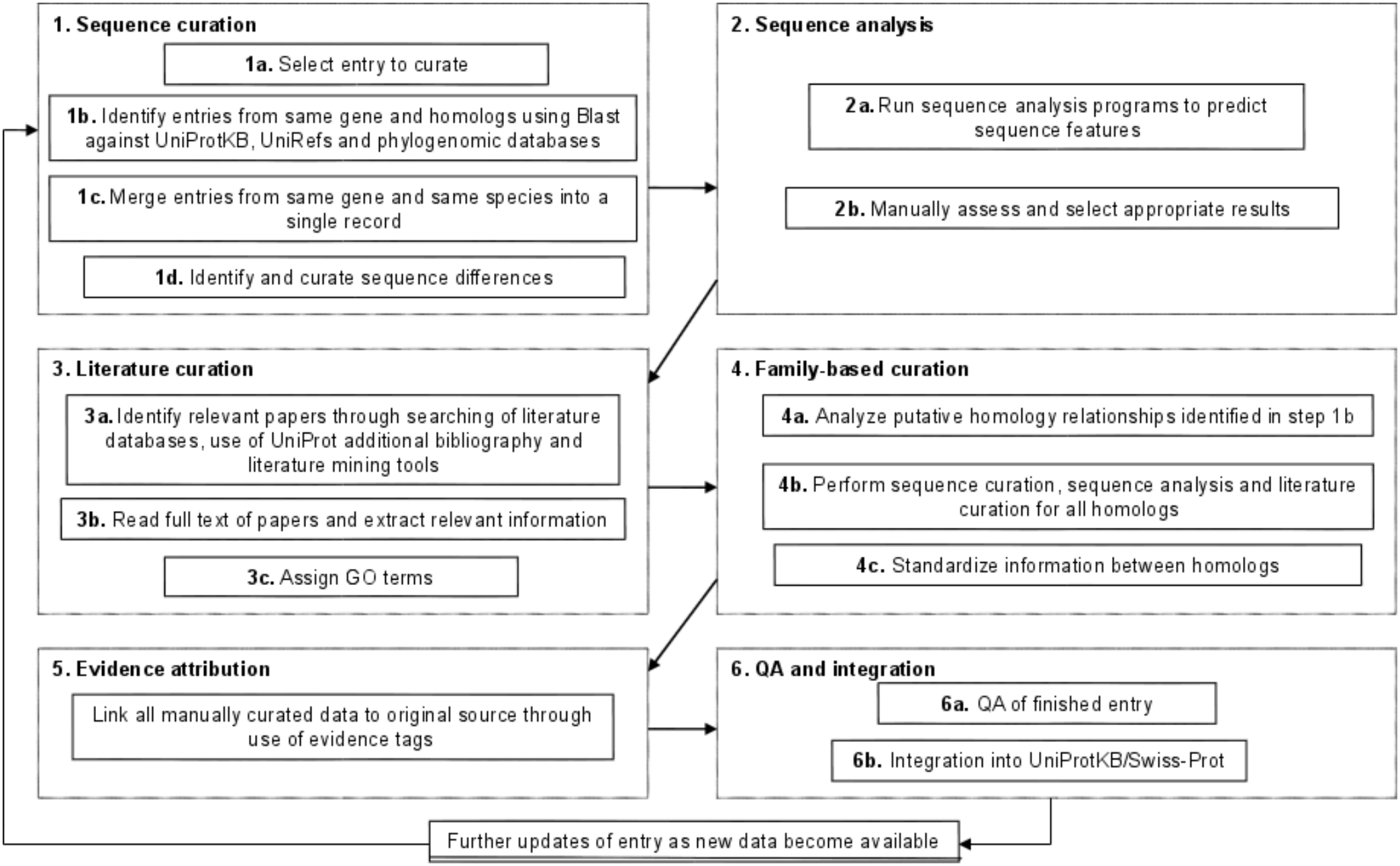
- www.uniprot.org
- Teljes adatkészletek letöltéséhez: ftp.uniprot.org



UniProtKB/Swiss-Prot annotáció

- Egy adott fehérjéről való kurrens tudásunk
- Ún. 'controlled vocabulary"', azaz adott kifejezések, (kulcs)szavak következetes használata
  Keywords, Tissues, Post-translational modifications, Strains, Species, Subcellular location, Extracellular domains, Journals…
- Megbízható annotációt ad, ami felhasználható:
  - Nem jellemzett genomokból származó szekvenciákra az annotáció átviteléhez (de legyünk óvatosak az ortológia/paralógia kapcsán!)
  - Programok, eszközök tesztelése (adatbányászat, predikciók stb.)
- UniProtKB/Swiss-Prot adatforrások:
  - publikációk (Pubmed)
  - adatbázisok
  - Nevezéktani bizottságok
  - predikciók
  - Szerzőkkel való kapcsolatfelvétel
- A manuális annotációra kiválasztott fehérjék:
  - Új, nagy hatású publikációkban leírt fehérjék (pl. cereblon, Q96SW2, http://www.expasy.org/spotlight/back_issues/sptlt117.shtml)
  - Adott metabolikus vagy jelátviteli útvonal (pl. az ubiquitin-szerű konjugációs útvonal)
  - Felhasználói kérések
  - 3D szerkezettel rendelkező fehérjék

UniProtKB/Swiss-Prot annotációs folyamatábra



Az annotáció eredetét/megbízhatóságát jellemző azonosítók

Coiled coil

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Coiled coil [i]	842 – 1939	1098	Sequence analysis			Add BLAST

Domain [i]

The rodlike tail sequence is highly repetitive, showing cycles of a 28-residue repeat pattern composed of 4 heptapeptides, characteristic for alpha-helical coiled coils.
Each myosin heavy chain can be split into 1 light meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be split further into 2 globular subfragments (S1) and 1 rod-shaped subfragment (S2).

Sequence similarities [i]

Belongs to the TRAFAC class myosin-kinesin ATPase superfamily. Myosin family. Curated

Contains 1 IQ domain. PROSITE-ProRule annotation

Contains 1 myosin motor domain. Curated

Keywords - Domain [i]

Coiled coil

Manual assertion inferred by curator

Manual assertion inferred by rules

Manual assertion based on experiment

Natural variant

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Natural variant [i]	56 – 56	1	G → R. 1 Publication Corresponds to variant rs28711516 [dbSNP | Ensembl].		VAR_063550	
Natural variant [i]	88 – 88	1	E → Q. 2 Publications Corresponds to variant rs442275 [dbSNP | Ensembl].		VAR_030203	
Natural variant [i]	275 – 275	1	I → N. 1 Publication		VAR_063551	
Natural variant [i]	721 – 721	1	R → W in SSS3; rare variant predisposing to sick sinus syndrome. 1 Publication		VAR_065561	

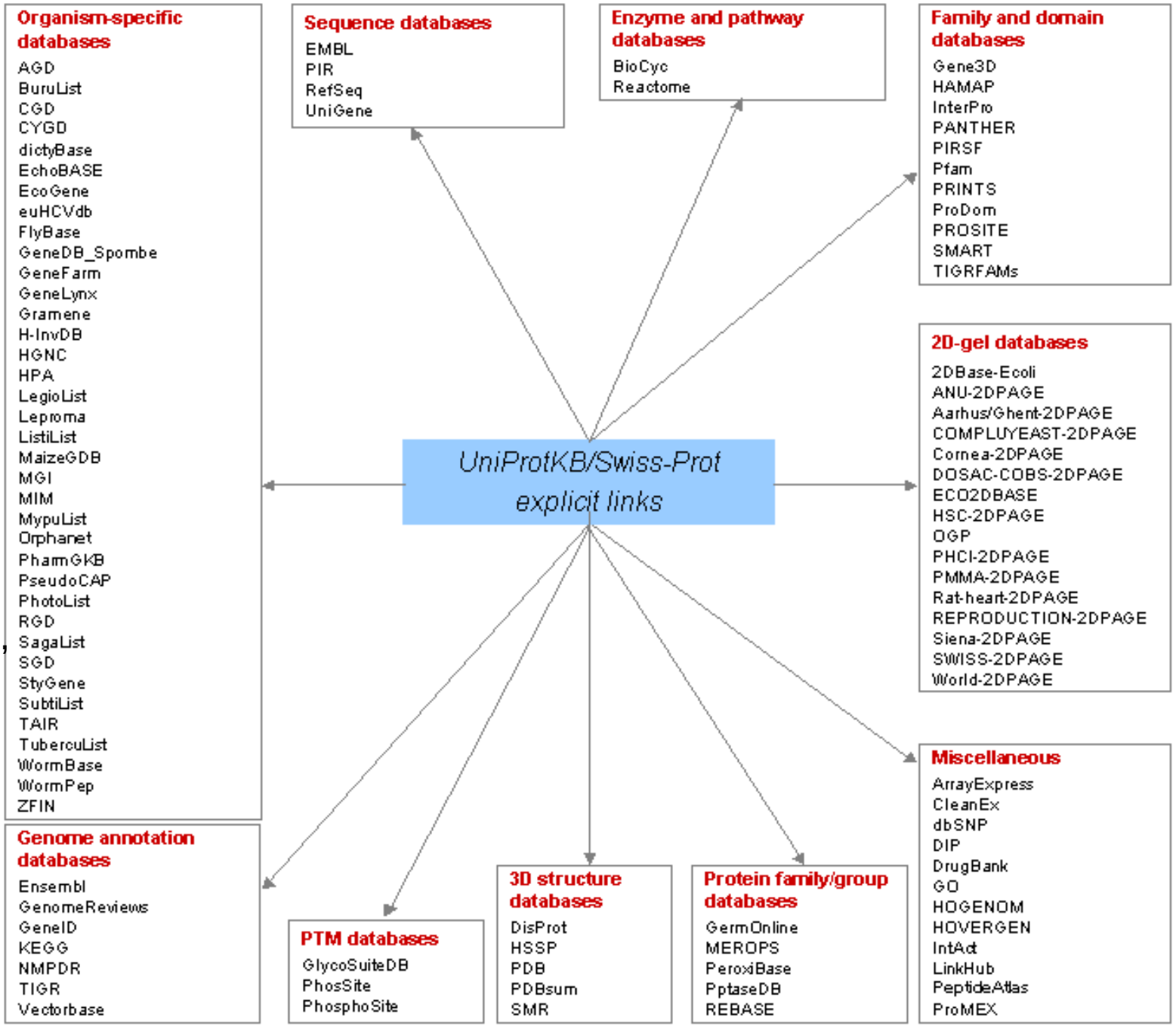
Fehérje izoformák a SwissProt adatbázisban

- Egy reprezentatív izoforma az „alapértelmezett”
- A többi izoforma variánsként listázva, letölthetőek
- Teljes adatkészletek az összes variánssal letölthetőek (számos proteom)

- Információk további variánsokról, módosításokról

SwissProt: kereszthivatkozások

- 125 adatbázishoz:
- EMBL/ GenBank/DDBJ, RefSeq,
- PDB
- InterPro,
- PROSITE, Pfam, Prints, etc.
- Organizmus-specifikusak: MIM, MGI, FlyBase, SGD, GenoList,
- SWISS-2DPAGE
- PubMed



SwissProt és TrEMBL

- Definíciók:
  - Egy TrEMBL bejegyzés egy automatikusan generált bejegyzés egy EMBL CDS sor alapján, automatikus annotációval ('*Unreviewed*')
  - Egy Swiss-Prot bejegyzés egy manuálisan annotált rekord (*Reviewed*)

- A SwissProtba való integráció során az ID megváltozik, de az AC (accession number) megmarad

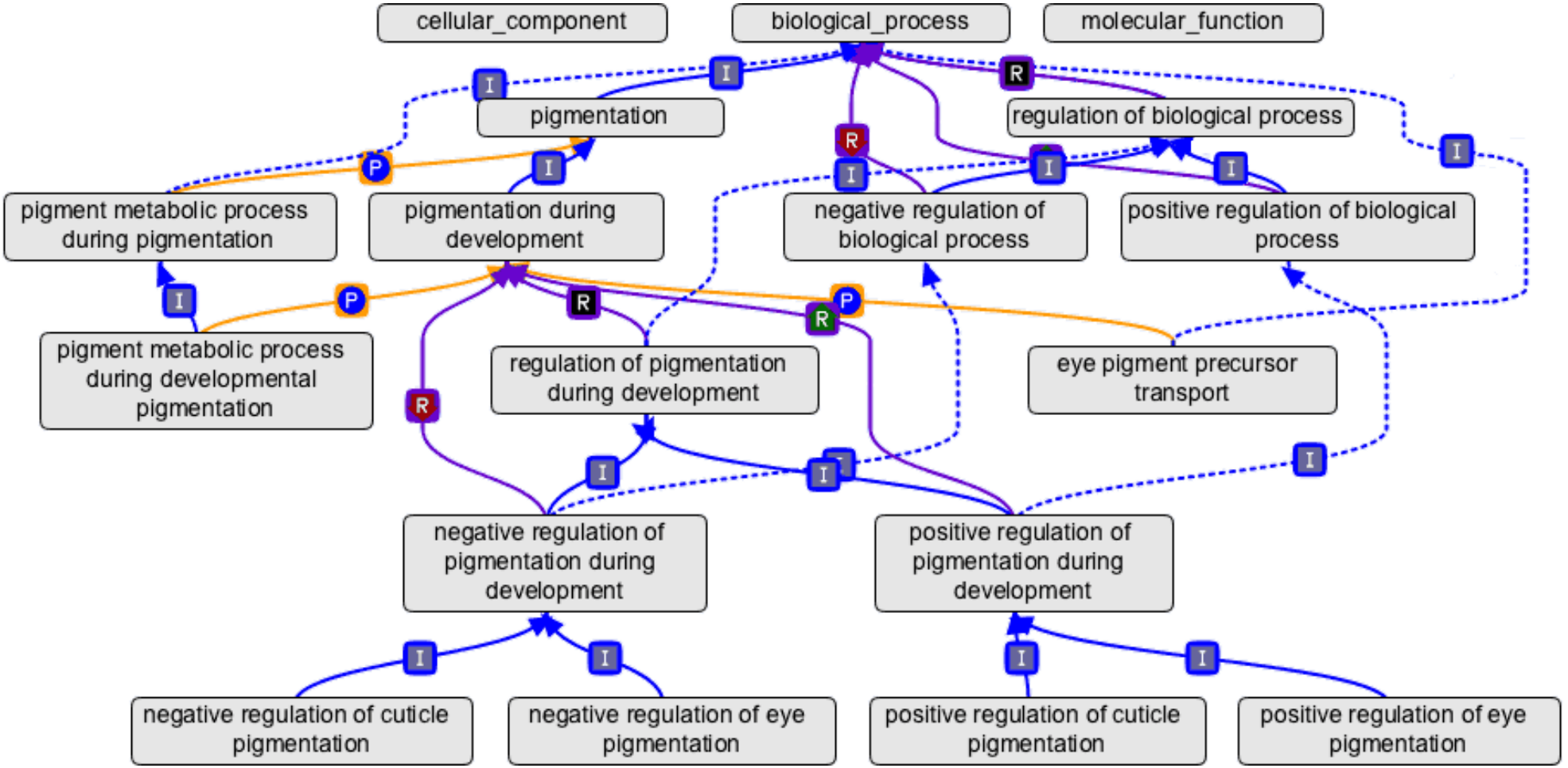
Serine/threonine protein phosphatase 2A 55 kDa regulatory subunit B beta isoform *Oryza sativa*

```
ID   A2X2K3_ORYSI              Unreviewed;       524 AA.
AC   A2X2K3;
DT   20-MAR-2007, integrated into UniProtKB/TrEMBL.
DT   20-MAR-2007, sequence version 1.
DT   24-JUL-2007, entry version 5.

ID   2ABB_ORYSI                Reviewed;         525 AA.
AC   A2X2K3; O82774; Q6Z8B7;
DT   11-SEP-2007, integrated into UniProtKB/Swiss-Prot.
DT   11-SEP-2007, sequence version 2.
DT   11-SEP-2007, entry version 6.
```

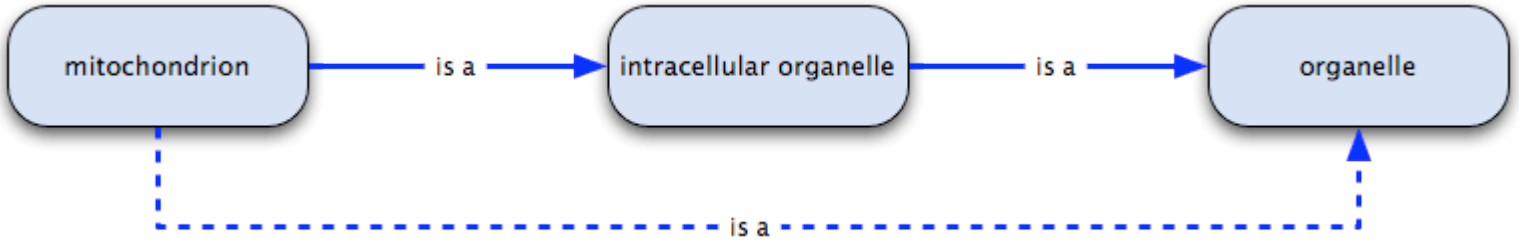
A GO (Gene Ontology) osztályozás

- www.geneontology.org
- Cél: egységes, jól követhető biológiai jellemzést adni génekről/fehérjékről
- Három különböző aspektus:
  - Cellular component (hol lokalizált a fehérje)
  - Biological process (milyen tágabb folyamatban vesz részt)
  - Molecular function (mi a szűkebb [molekuláris szintű] feladata)
- Az egyes kifejezések, kulcsszavak között relációk vannak

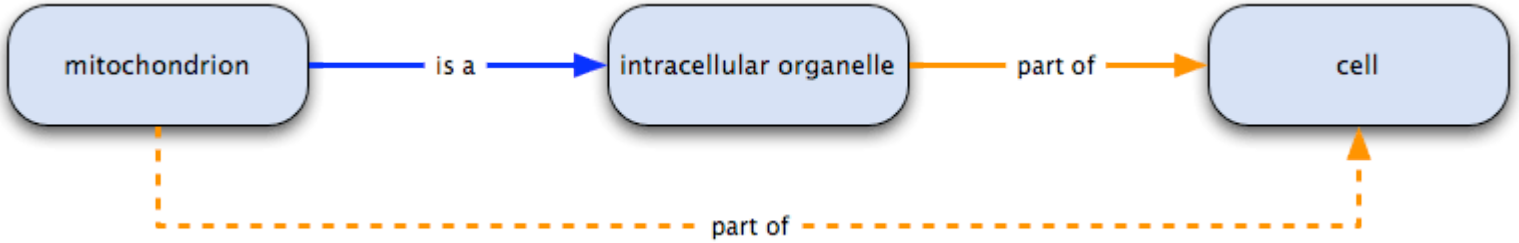


A GO (Gene Ontology) osztályozás

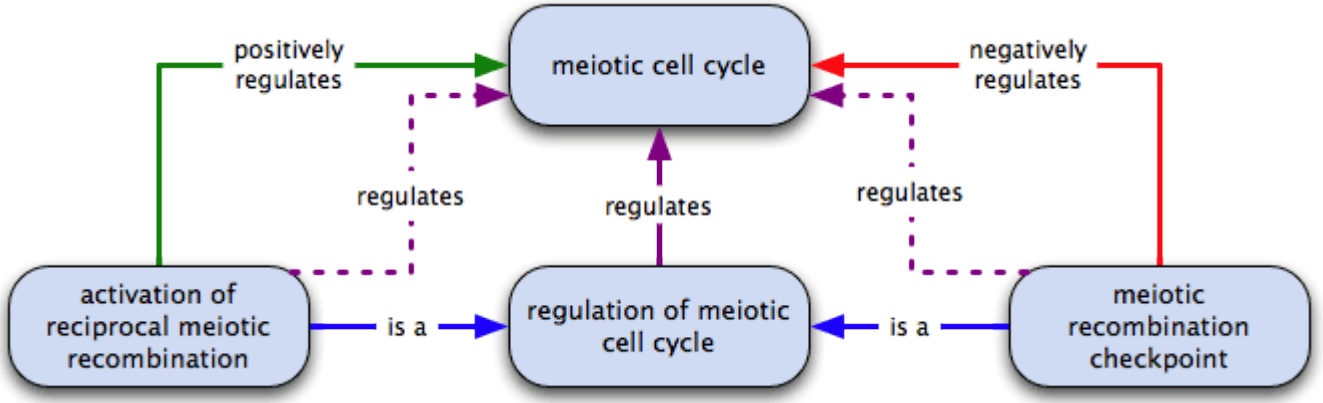
- Relációk
  - Vannak alesetek (lásd a regulációnál)
  - Egyes relációk implikálnak másokat (logikai összefüggések)



„is a”
= típusba sorolás



„part of”
= rész-egész viszony



„regulates”
= szabályozás

A GO (Gene Ontology) osztályozás

- Jelentőség proteomikai vizsgálatokban:
  - Adott fehérjecsoportok összevetése: van-e köztük funkcionális eltérés (pl. azonosítunk n darab fehérjét, amik egy betegségben fontosak stb. - van-e stattisztikailaig kimutatható különbség a GO kulcsszavak eloszlásában)

BIOINFORMATICS REVIEW

Vol. 23 no. 4 2007, pages 401–407
doi:10.1093/bioinformatics/btl633

Databases and ontologies

Enrichment or depletion of a GO category within a class of genes: which test?

Isabelle Rivals*, Léon Personnaz, Lieng Taing[1] and Marie-Claude Potier[1]

Équipe de Statistique Appliquée and [1]Laboratoire de Neurobiologie et Diversité Cellulaire, École Supérieure de Physique et de Chimie Industrielles (ESPCI), 10 rue Vauquelin, 75005 Paris, France

Received on June 20, 2006; revised and accepted December 11, 2006
Advance Access publication December 20, 2006
Associate Editor: Jonathan Wren

Table 2. Classification of the genes expressed in a microarray experiment

	Category 1 ($\in$GO category)	Category 2 ($\notin$GO category)	Total
Class 1 (DE)	n_{11}	n_{12}	n_{1+}
Class 2 (not DE)	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	n

Motívum és domén adatbázisok

- Másodlagos adatbázisok
- Motívum: egy szekvenciális szegmens, ami különböző fehérjékben előfordul, és funkcionális jelentősége van
- Domén: szerkezeti/feltekeredési/funkcionális/evolúciós egység (lásd még később)
- A motívumok és a domének szekvenciális hasonlóság alapján detektálhatók

SBASE — Doménygyűjtemény, cél a ritka variánsok felismerése/lefedése

Pfam — Fehérjedomének/családok illesztésekkel és HMM-ekkel reprezentálva

ProDom — Illesztett fehérjeszekvenciák és domének

BLOCKS — Illesztések konzervált „blokkjai”

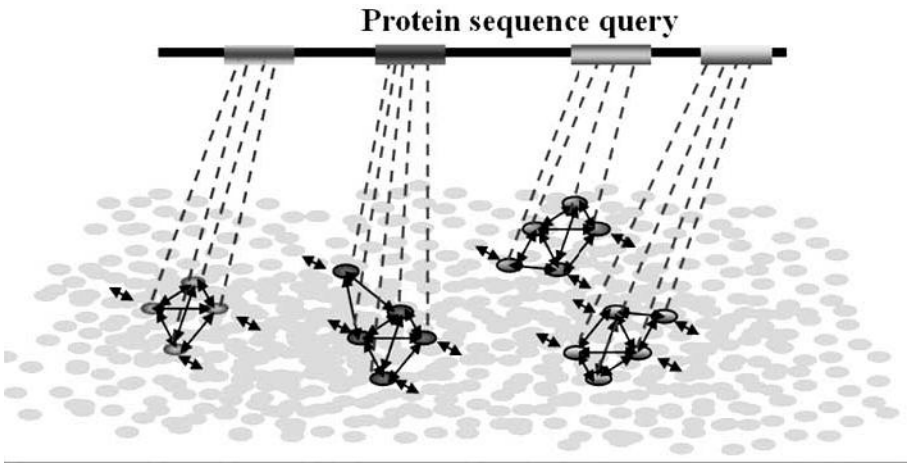
SMART — Néhány száz domén manuálisan karbantartott modellje

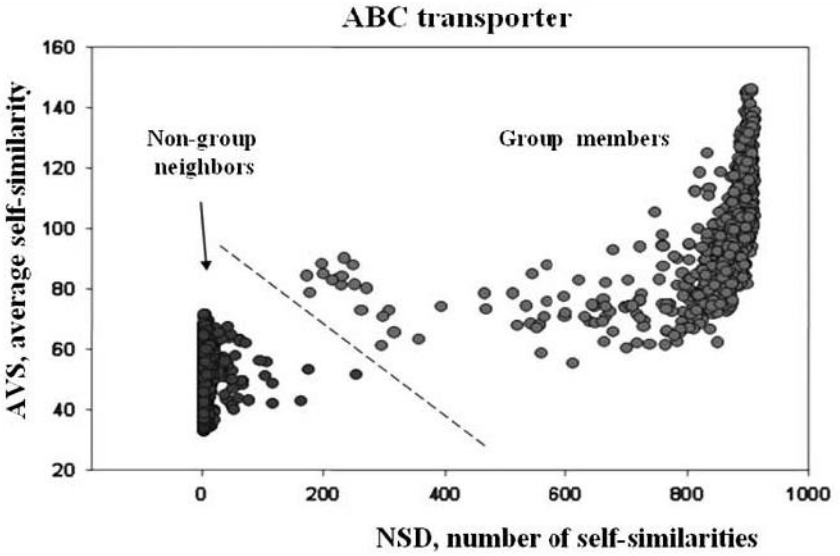
Superfamily 1.75 — HMM library and genome assignments server — Minden fehérje és genom funkcionális annotációja

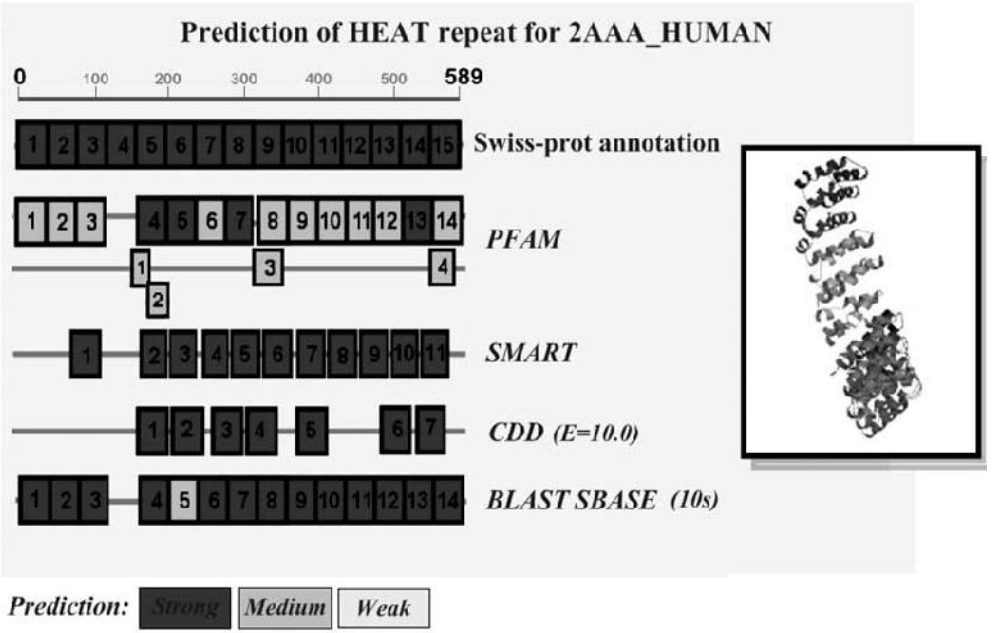
InterPro — Protein sequence analysis & classification — 18, főleg másodlagos szervert integráló ún. metaszerver

Példa: SBASE áttekintés

- Egyszerű szekvenciális hasonlóságokat keres
- Saját/nem saját megkülönböztetése hasonlósági csoportok alapján
- Képes ritka doménvariánsokat is detektálni

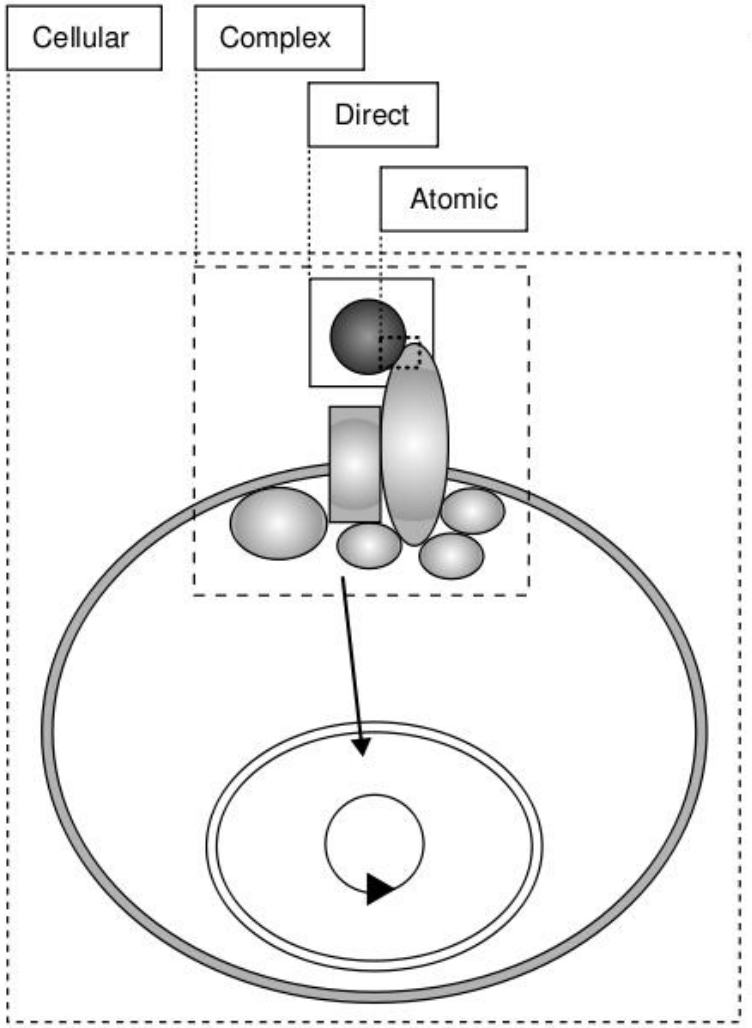


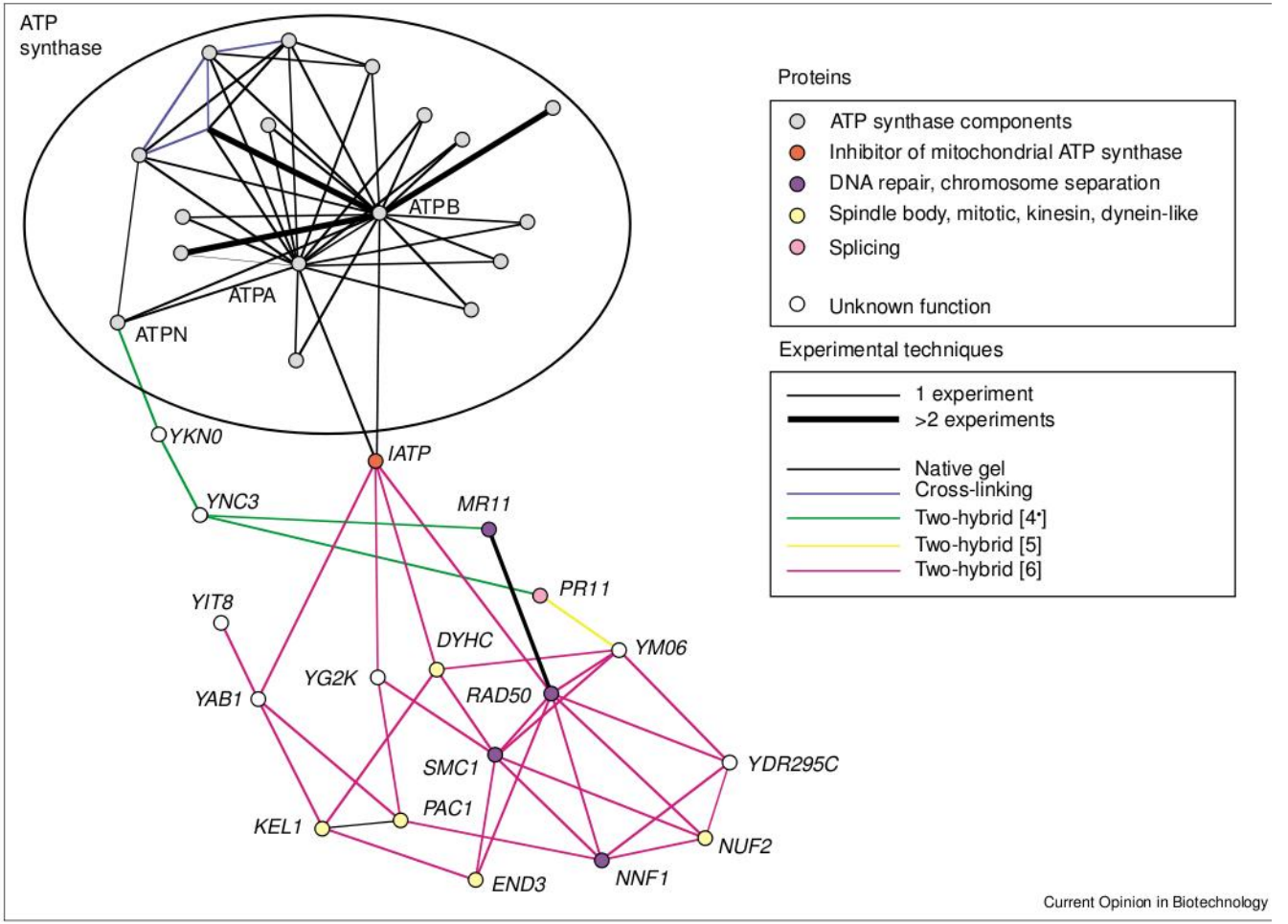




Fehérje-fehérje kölcsönhatási adatbázisok

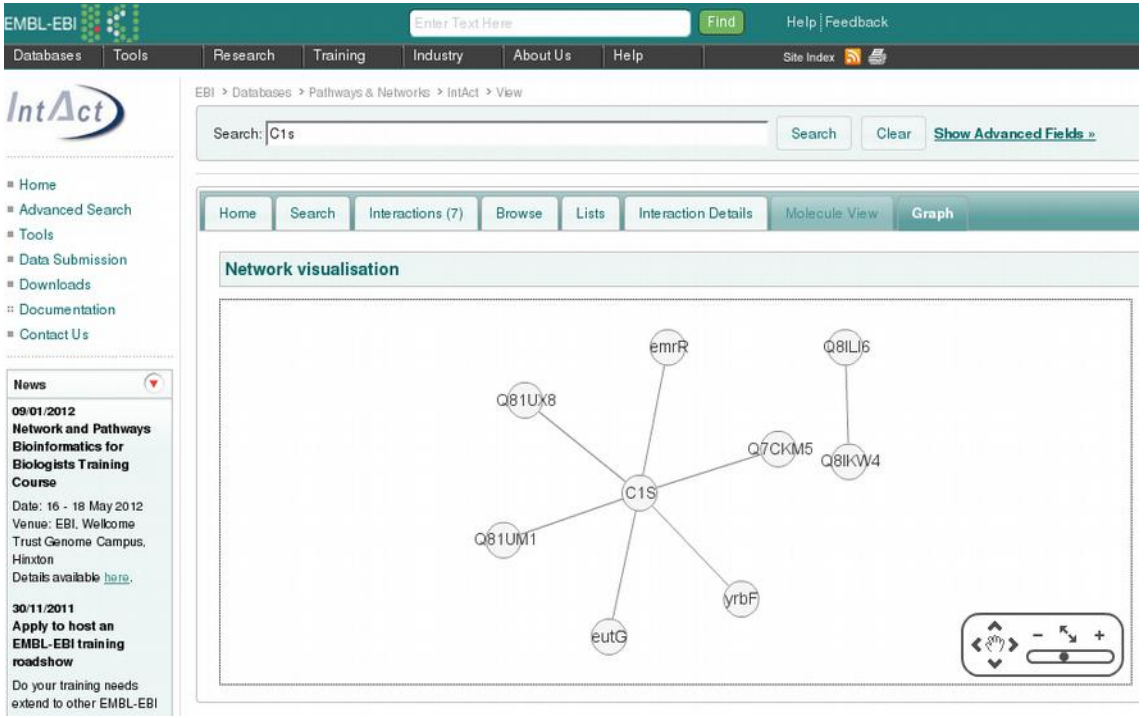
- Probléma: honnan származik a kölcsönhatásról az információ
  - Rengeteg különféle kísérleti módszer, eltérő felbontással és megbízhatósággal
  - Egyedi vizsgálatok vs. nagy adatkészleten végzett kutatások (large scale)
  - A kölcsönhatások jellemző paraméterei is több nagyságrendet ölelnek fel
  - De szeretjük igen-nem alapon megjeleníteni a kölcsönhatásokat
- Sokféle adatbázis



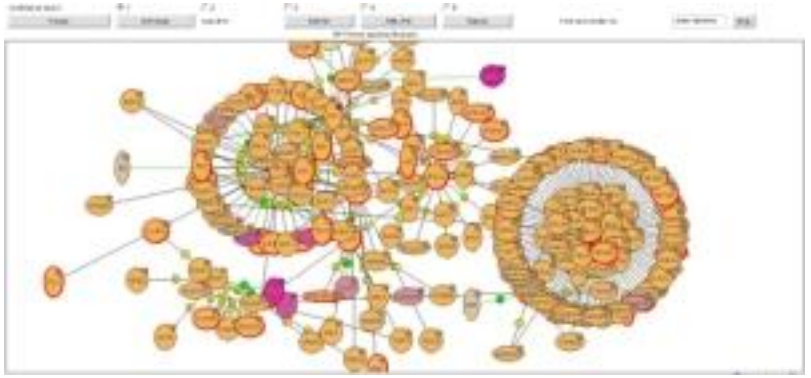


Fehérje-fehérje kölcsönhatási adatbázisok

IntAct



MINT



STRING