Programas para Alineamientos Múltiples

http://en.wikipedia.org/wiki/Sequence_alignment_software

Multiple sequence alignment [ed

Name	Description	Sequence Type*	Alignment Type**	Link	Author	Year
ABA	A-Bruijn alignment	Protein	Global	download	B.Raphael et al.	2004
DNA Baser	Multi alignment/Batch alignment	Nucleotides	Local or Global	download	C.Mathias et al.	(latest version July 2008)
ALE	manual alignment ; some software assistance	Nucleotides	Local	download	J. Blandy and K. Fogel	1994 (latest version 2007)
AMAP	Sequence annealing	Both	Global	server	A. Schwartz and L. Pachter	2006
BAli-Phy	Tree+Multi alignment ; Probabilistic/Bayesian ; Joint Estimation	Both	Global	WWW+download	BD Redelings and MA Suchard	2005 (latest version 2007)
CHAOS/DIALIGN	Iterative alignment	Both	Local (preferred)	server	M. Brudno and B. Morgenstern	2003
ClustalW	Progressive alignment	Both	Local or Global	download EBI DDBJ PBIL EMBNet GenomeNet	Thompson et al.	1994
CodonCode Aligner	Multi alignment; ClustalW & Phrap support	Nucleotides	Local or Global	download	P. Richterich et al.	2003 (latest version 2007)
DIALIGN-TX and DIALIGN-T	Segment-based method	Both	Local (preferred) or Global	download and server	A.R.Subramanian	2005 (latest version 2008)
DNA Alignment	Segment-based method for intraspecific alignments	Both	Local (preferred) or Global	server	A.Roehl	2005 (latest version 2008)
Ed'Nimbus	Seeded filtration	Nucleotides	Local	server	P. Peterlongo et al.	2006
FSA	Sequence annealing	Both	Global	download and server	R. Bradley et al.	2008

Alindamientos múltiples de proteínas y comparaciones DB

- Los análisis filogenéticos basados en alineamientos múltiples son muy importantes para entender las relaciones entre secuencias de interés.
  - Como podemos saber cómo nuestras secuencias se relacionan con otras y como es esa relación?
  - Como podemos asociar nuestras secuencias a otras ya estudiadas para establecer su función?
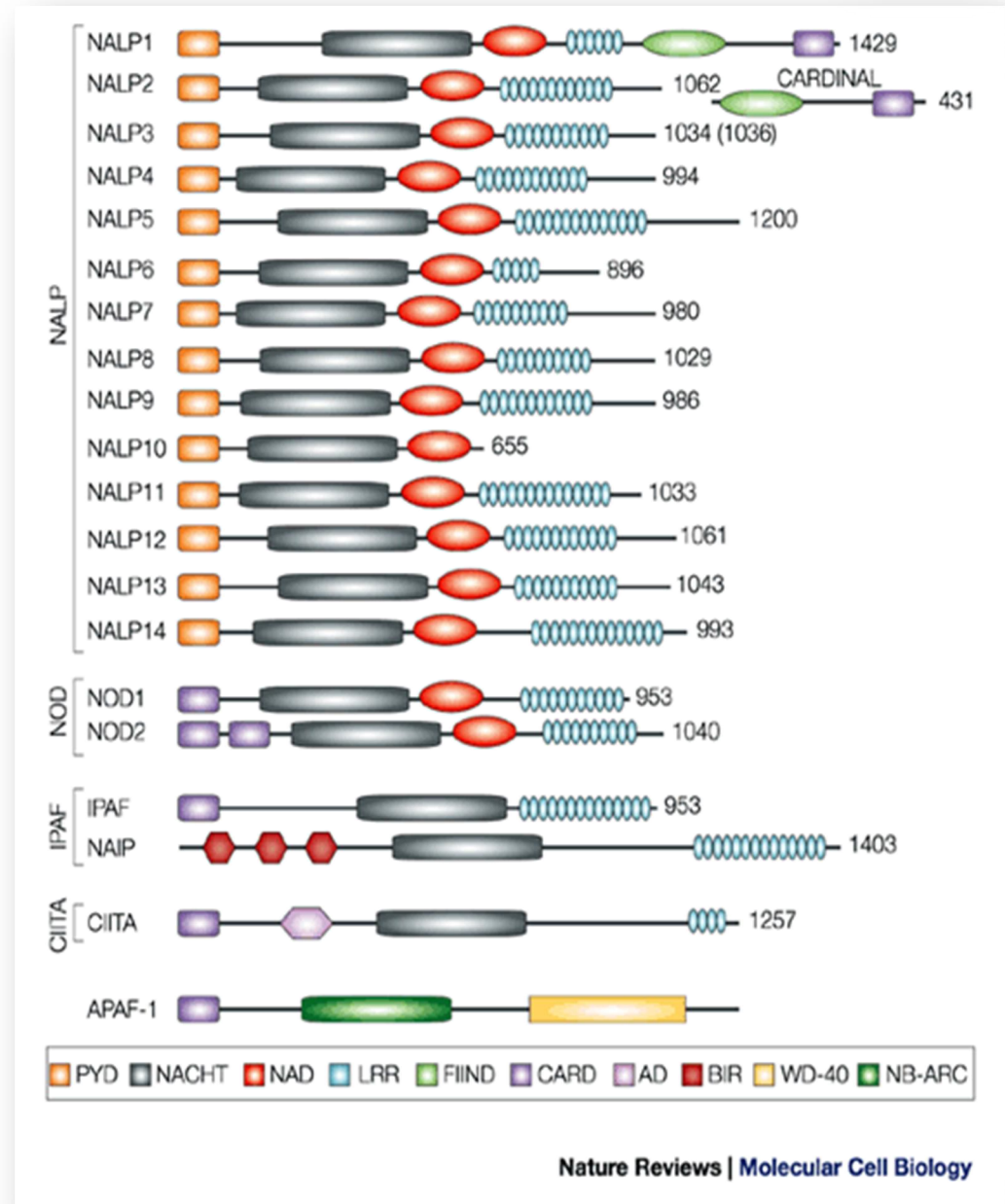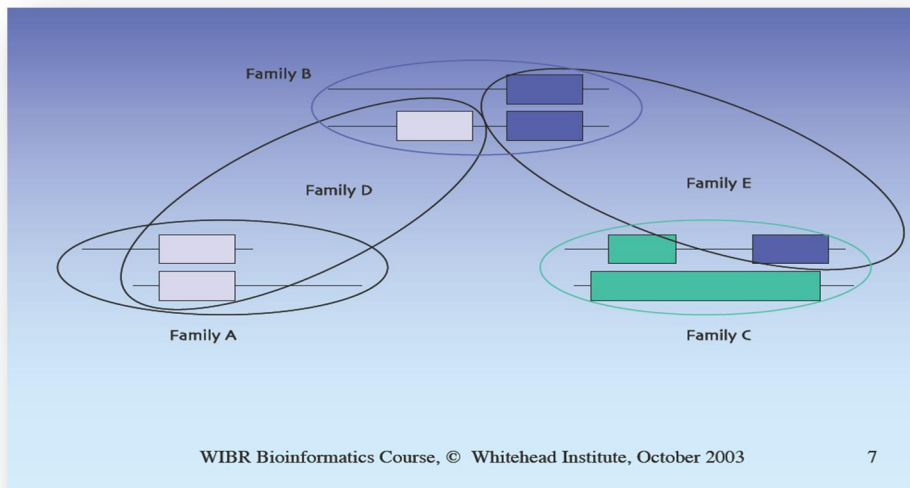
Familias de proteínas

- Las proteínas se pueden estudiar por "módulos" a lo largo de su secuencia de aa:
- Dominios de Proteínas
- Bases de datos de Familias de Proteínas (donde se pueden comparar una seq problema contra DB) usando las homologías :
- Búsqueda de Patrones (Patscan)
- Búsquedas por Perfiles (PSI-BLAST/HMMER2)

La evolución va mezclando estos "módulos", originando un inmenso repertorio de proteínas, que al final comparten características globales o locales.
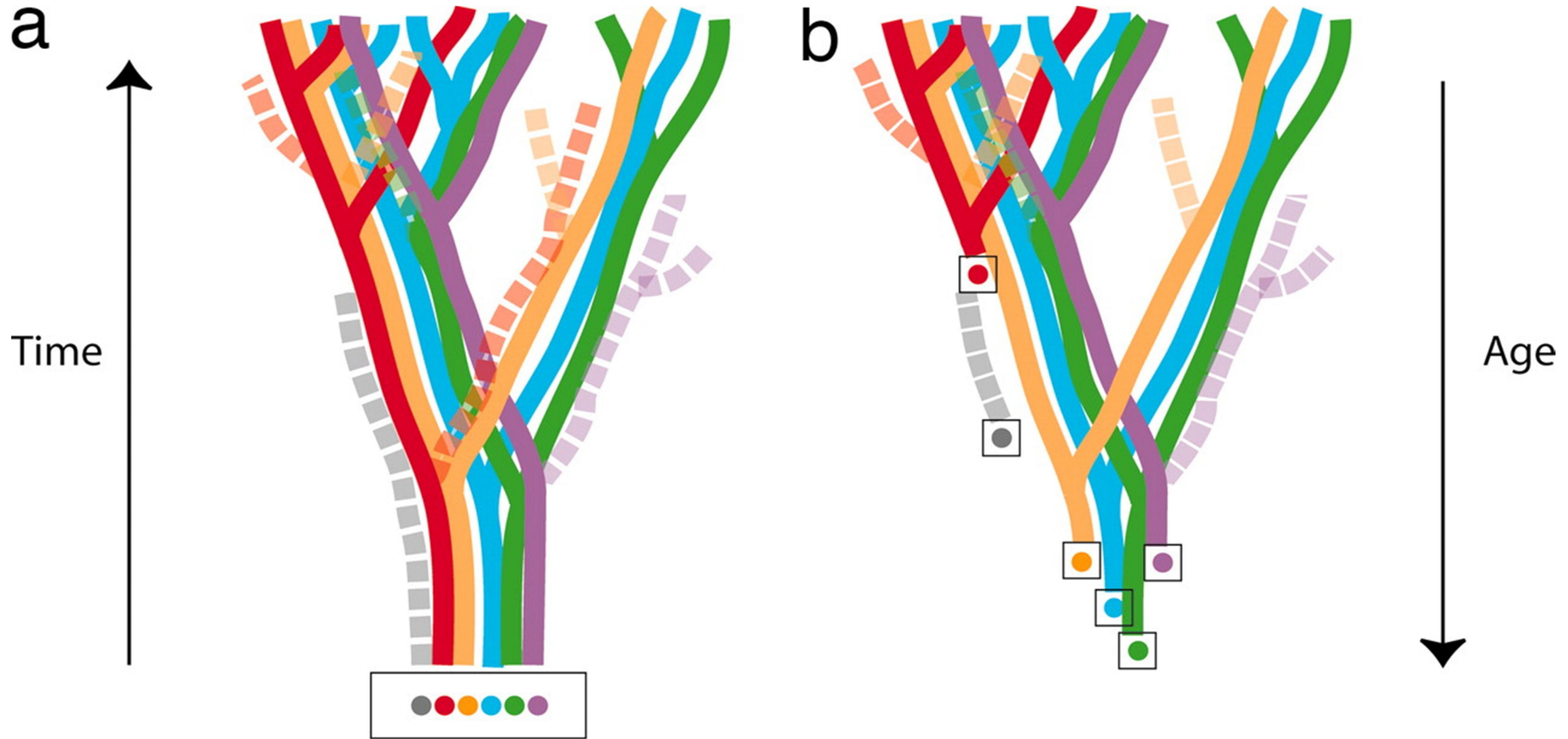
Familias de Proteínas

- Estas comparten una función o estructura en común, que potencialmente comparte un ancestro común.

- Siendo los motivos o dominios comunes a una familia.

http://www.russelllab.org/

WIBR Bioinformatics Course, © Whitehead Institute, October 2003   7

Nature Reviews | Molecular Cell Biology

EVOLUCIÖN EM PROTEINA

A) *Single birth model of protein families / b) Multiple birth model of protein families (Evolution of protein structural classes and protein sequence families. PNAS September 19, 2006 vol. 103 no. 38 14056-14061.*

Bases de datos de Proteínas

Curated Databases

- Proteins are placed into families with which they share a specific sequence pattern.
- **Pfam:** http://pfam.wustl.edu/hmmsearch.shtml/
- **Prosite**: http://www.expasy.ch/tools/scanprosite/
- **PRINTS:** http://www.bioinf.man.ac.uk/fingerPRINTScan/

- **Clustering Databases**
- Sequence similarity-based without the prior knowledge of a specific patterns

- **ProDom:** http://prodes.toulouse.inra.fr/prodom/doc/prodom.html
- **Protomap** : http://protomap.cornell.edu/

- **Derived Databases**
- Pool other databases into one central resource
- **Blocks:** http://blocks.fhcrc.org/blocks/
- **Proclass**: http://pir.georgetown.edu/gfserver/proclass.html/
- **MEME** http://meme.sdsc.edu/meme/website/intro.html

- ***Mas en***
- http://biology.unm.edu/biology/maggieww/Public_Html/ANALYSIS.HTM
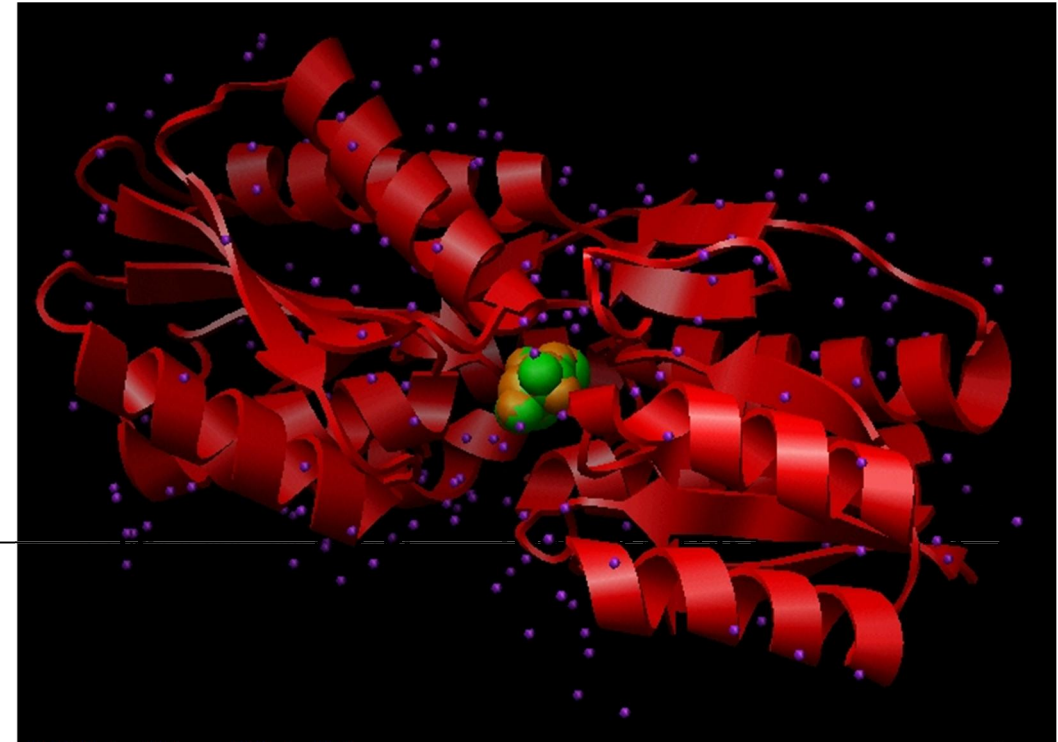
Bases de Datos de estructuras de Proteínas

Formatos:

MMDB "Molecular Modeling DataBank" Format ASN.1 standard data description language (explicit bond information)

mmCIF "Chemical Interchange Format" (relational db format)

PDB "Protein DataBank" Format Column oriented, "flexible format" (chemistry rules)

PDB Database


```
SHEET
SHEET
SHEET
SHEET
SHEET
SHEET
SHEET
SHEET
SHEET
SHEET
SHEET
SSBOND
SSBOND    2  CYS A  217        CYS A  247
SSBOND    3  CYS A  263        CYS A  275
SSBOND    4  CYS A  320        CYS A  418
SSBOND    5  CYS A  337        CYS A  409
SSBOND    6  CYS E  149        CYS E  302
SSBOND    7  CYS E  217        CYS E  247
SSBOND    8  CYS E  320        CYS E  418
SSBOND    9  CYS E  337        CYS E  409
CISPEP    1  GLU A  187        PRO A  188          0        -2.06
CISPEP    2  SER A  191        PRO A  192          0        -5.95
CISPEP    3  GLU E  187        PRO E  188          0        -2.03
CISPEP    4  SER E  191        PRO E  192          0        -3.85
CRYST1    54.124    54.124   214.144  90.00  90.00 120.00 P 31         6
ORIGX1      1.000000  0.000000  0.000000        0.00000
ORIGX2      0.000000  1.000000  0.000000        0.00000
ORIGX3      0.000000  0.000000  1.000000        0.00000
SCALE1      0.018476  0.010667  0.000000        0.00000
SCALE2      0.000000  0.021334  0.000000        0.00000
SCALE3      0.000000  0.000000  0.004670        0.00000
ATOM       1   N   ASN A 108     -3.300  43.171  49.886  1.00 33.67        N
ATOM       2   CA  ASN A 108     -2.100  42.301  50.061  1.00 33.38        C
ATOM       3   C   ASN A 108     -1.168  42.317  48.852  1.00 33.08        C
ATOM       4   O   ASN A 108     -1.499  41.752  47.813  1.00 32.57        O
ATOM       5   CB  ASN A 108     -2.507  40.845  50.376  1.00 33.34        C
ATOM       6   CG  ASN A 108     -1.301  39.962  50.768  1.00 34.09        C
```

Bases de datos de estructuras

SCOP: Structural Classification Of Proteins based on a definition of structural similarities. Hierarchical levels to reflect evolutionary and structural relationships
http://scop.mrc-lmb.cam.ac.uk/scop

CATH: Classification by Class, Architecture, Topology, and Homology classified first into hierarchical levels like SCOP
http://www.biochem.ucl.ac.uk/bsm/cath/

FSSP: Fold classification based on Structure-structure alignment of proteins based on structural alignment of all pair-wise combinations of proteins in PDB by DALI (used to id common folds and place into groups)
http://www2.embl-ebi.ac.uk/dali/fssp/fssp.html

MMDB: Aligns 3D structures based on similar arrangements of secondary structural elements (VAST)
http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml\

SARF: categorized on the basis of structural similarity, categories are similar to other DBs
http://123d.ncifcrf.gov/

Modelando estructuras

Realizan de manera "automática" reconstrucciones de modelos

– SWISS-MODEL: Compare sequence to ExPdb to find a template (homology). Define your own templates (from threading)

http://www.expasy.ch/swissmod/SWISS-MODEL.html

– GENO3D: PSI-BLAST to identify homologs possessing structures to be used as templates

http://geno3d-pbil.ibcp.fr

Visualización

RasMol (Chime is the Firefox plug-in)

http://www.umass.edu/microbio/rasmol/index2.html

Cn3D MMDB viewer (See in 3D) with explicit bonding

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

SwissPDB Viewer

http://www.expasy.ch/spdbv/mainpage.html

iMol
http://www.pirx.com/iMol