Βιοπληροφορική

Εισαγωγή

΄Υλη Βιοπληροφορικής

Προτεινόμενα συγγράμματα

- Ελληνικά συγγράμματα:
  - Andreas D. Baxevanis & B.F. Francis Quellette. Βιοπληροφορική: Ένας πρακτικός οδηγός για την ανάλυση γονιδίων και πρωτεΐνών.
  - Σοφία Κοσσίδα. Βιοπληροφορική - Δυνατότητες & Προοπτικές.

- Αγγλικά συγγράμματα:
  - Jin Xiong. Essential Bioinformatics. (Σύντομο, περιεκτικό και απλά γραμμένο σύγγραμα).
  - David W. Mount. Bioinformatics. Sequence and genome analysis. (Εκτενές και πολύ αναλυτικό σύγγραμα)

Baxevanis & Quellette

Κεφάλαια:

- 3 Β.Δ. Genbank.
- 8 Στοίχιση ακολουθιών και αναζήτηση σε Β.Δ.
- 9 Πολλαπλές στοιχίσεις.
- 14 Φυλογένεση

Essential bioinformatics

- Προτείνεται να διαβαστούν και κάποια κεφάλαια από το αγγλικό σύγγραμμα του Jin Xiong, και ειδικότερα τα κεφάλαια 1-6, 10,11.

Βιοπληροφορική: τι είναι

- Η ανάπτυξη και χρήση τεχνικών και εργαλείων πληροφορικής/μαθηματικών/ στατιστικής για την ανάλυση βιολογικών δεδομένων (κυρίως μοριακής βιολογίας)

- Σήμερα γίνεται διάκριση μεταξύ της βιοπληροφορικής και της υπολογιστικής βιολογίας
  - Βιοπληροφορική: Η ανάπτυξη μεθόδων και προγραμμάτων.
  - Υπολογιστική Βιολογία:Η χρήση των παραπάνω μεθόδων και προγραμμάτων για την ανάλυση βιολογικών δεδομένων.

- Συχνά συμβαίνουν και τα δύο ταυτόχρονα και τα σύνορα δεν είναι πάντα ευδιάκριτα

- Πολλές και συμπληρωματικές μεταξύ τους ειδικότητες (από Βιολογία, Βιοχημεία, χημεία, χημική μηχανική, μηχανική, υπολογιστές, μαθηματικά, στατιστική κ.α.) συνεργάζονται σήμερα στο χώρο της βιοπληροφορικής

Βιοπληροφορική: βασικοί τομείς

- Βάσεις δεδομένων (Databases)
  - Οργάνωση, αποθήκευση, αναζήτηση των δεδομένων.

- Ανάλυση ακολουθιών DNA, RNA, πρωτεϊνών. (Sequence analysis)
  - Στοίχιση ακολουθιών:Σύγκριση των αντίστοιχων περιοχών, μεταξύ δύο ή περισσοτέρων ακολουθειών.
  - Φυλογενετική ανάλυση: Οι εξελικτικές σχέσεις μεταξύ ομοειδών αντικειμένων (γονίδια, πρωτεΐνες, οργανισμοί).

- Γονιδιακή ρύθμιση/έκφραση (Gene expression)
  Ανάλυση δεδομένων από μικροσυστοιχίες.

- Δομή RNA/πρωτεϊνών (structural biology):
  Πρόβλεψη δευτεροταγούς και τριτοταγούς δομής. Ανάλυση πρωτεϊνικών επιφανειών που αλληλεπιδρούν μεταξύ τους.

- Εξόρυξη δεδομένων από βιβλιογραφία (text mining).

- Βιολογικά δίκτυα/μονοπάτια, Βιολογία Συστημάτων.

- Οντολογίες (Ontologies)
  Η χρήση ενός ελεγχόμενου λεξιλογίου (με ιεραρχική δόμηση), για την περιγραφή των ιδιοτήτων και των λειτουργιών ομοειδών αντικειμένων (π.χ πρωτεϊνών).

Ιστορική αναδρομή

- 1965: Η πρώτη έκδοση του Atlas of protein sequence and structure (Margaret Dayhoff), πρόδρομος της βάσης δεδομένων πρωτεϊνικών ακολουθιών PIR (protein information resource).
  - Ακολουθούν και άλλες βάσεις δεδομένων. 1986:Swissprot, Geneva

- 1970: Αλγόριθμος Needleman-Wunsch για την σύγκριση ακολουθιών
- 1990: Blast

- 1990s: Αρχή του Human genome project, που ‘ολοκληρώθηκε’ το 2001. Κινητήριος δύναμη για την αλματώδη ανάπτυξη της βιοπληροφορικής.

Παρόν/μέλλον

- Μέχρι το 2000, βιοπληροφορική σήμαινε κυρίως ανάλυση ακολουθιών.
- Η γενωμική αποτέλεσε το ερέθισμα για την ανάπτυξη τεχνολογιών που κάνουν μετρήσεις ευρείας κλίμακας.
- Από το 2000 και μετά, η βιοπληροφορική καλείται επίσεις να διαχειριστεί και να αναλύσει μεγάλα και πολύπλοκα δεδομένα από το χώρο της γενωμικής, της γονιδιακής έκφρασης, της πρωτεομικής κ.α.
- Πλέον ο όρος ‘βιοπληροφορική’ είναι τόσο εξειδικευμένος/γενικός, όσο και ο όρος ‘μοριακή βιολογία’!
- Βρισκόμαστε σε μια μεταβατική περίοδο για τις βιολογικές επιστήμες, όπως η φυσική πριν πολλά χρόνια. Βέβαιη η εισδοχή περισσότερων μαθηματικών, στατιστικής και πληροφορικής (προγραμματισμός) μεσοπρόθεσμα στο πρόγραμμα σπουδών.

Bioinformatics Market - Advanced Technologies, Global Forecast and Winning Imperatives (2009 - 2014)

- Απόσπασμα από:
  - http://www.marketsandmarkets.com/Market-Reports/bioinformatics-39.html

- The market for bioinformatics platforms is growing at a significant pace with the increasing demand from U.S. and Europe.
- This trend is supported by the increasing demand for sequencing platforms with increasing life science research using techniques such as gene expression analysis, sequence analysis, and protein expression analysis.
- The global bioinformatics market is expected to reach $8.3 billion by 2014 at a high CAGR of 24.8% from 2009-2014. While knowledge management formed the largest submarket is 2009 at $1.3 billion, the bioinformatics platforms market is expected to have greatest market share in 2014 at an estimated $3.9 billion, due to rising demand from the U.S. and Europe.

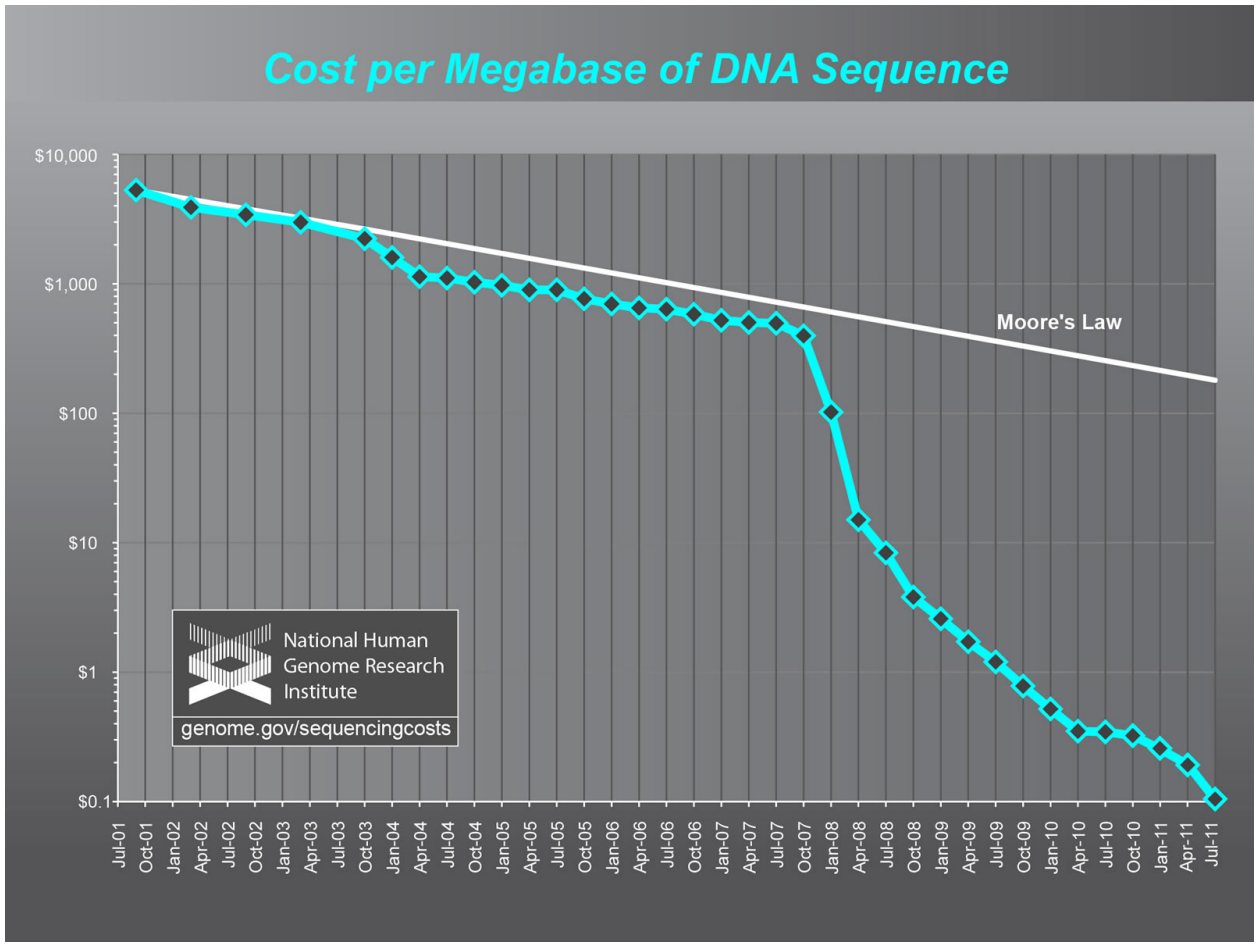
- Συμβουλευτική (δουλειά από το σπίτι)?

Χαμηλό κόστος γενωμικών τεχνολογιών θα οδηγήσει σε καθημερινές εφαρμογές.

- Κόστος αλληλούχισης πέφτει διαρκώς.
  - Illumina -> 1 lane: 19GBp, ~ €3000, 10 βακτηριακά γενώματα.
- Τα δείγματα αποστέλλονται σε κέντρα με μεγάλες εγκαταστάσεις και χαμηλό κόστος λειτουργίας (οικονομία κλίμακας). Η ανάλυση των δεδομένων όμως δεν υπόκειται σε όρους οικονομίας κλίμακας.
- Πλέον, ένα σημαντικό μέρος του ολικού κόστους είναι η βιοπληροφορική ανάλυση.
- Μηχανήματα αλληλούχισης ακριβά (Illumina ~ €600.000) - service φτηνό.
- Μισθός ακριβός (ίσως ένα νέο μοντέλο συμβουλευτικής?)
- Υπολογιστής φτηνός (€3-5.000), εφόσον πρόκειται για μικρά γονιδιώματα (de novo assembly), ή για re-sequencing.

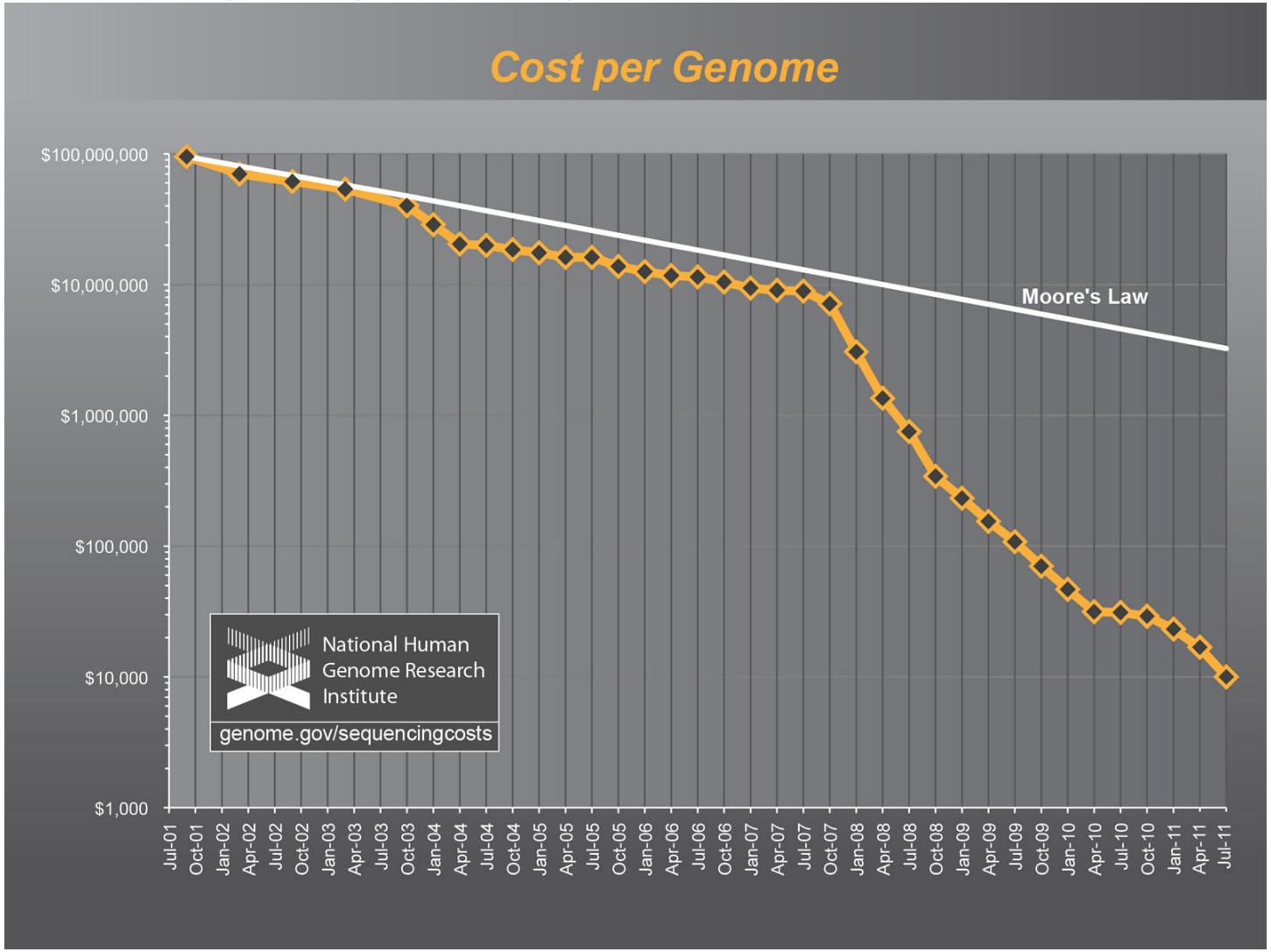
Χαμηλό κόστος γενωμικών τεχνολογιών θα οδηγήσει σε καθημερινές εφαρμογές

- Κόστος αλληλούχισης
  - http://www.genome.gov/sequencingcosts/
- Ο νόμος του Moore προβλέπει διπλασιασμό της υπολογιστικής ισχύς κάθε δύο χρόνια.



Χαμηλό κόστος γενωμικών τεχνολογιών θα οδηγήσει σε καθημερινές εφαρμογές

- Κόστος αλληλούχισης
  - http://www.genome.gov/sequencingcosts/



Εφαρμογές

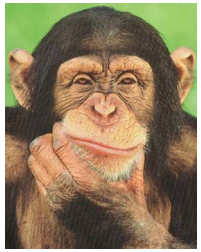
‘Ελεγχος εξελικτικών υποθέσεων -

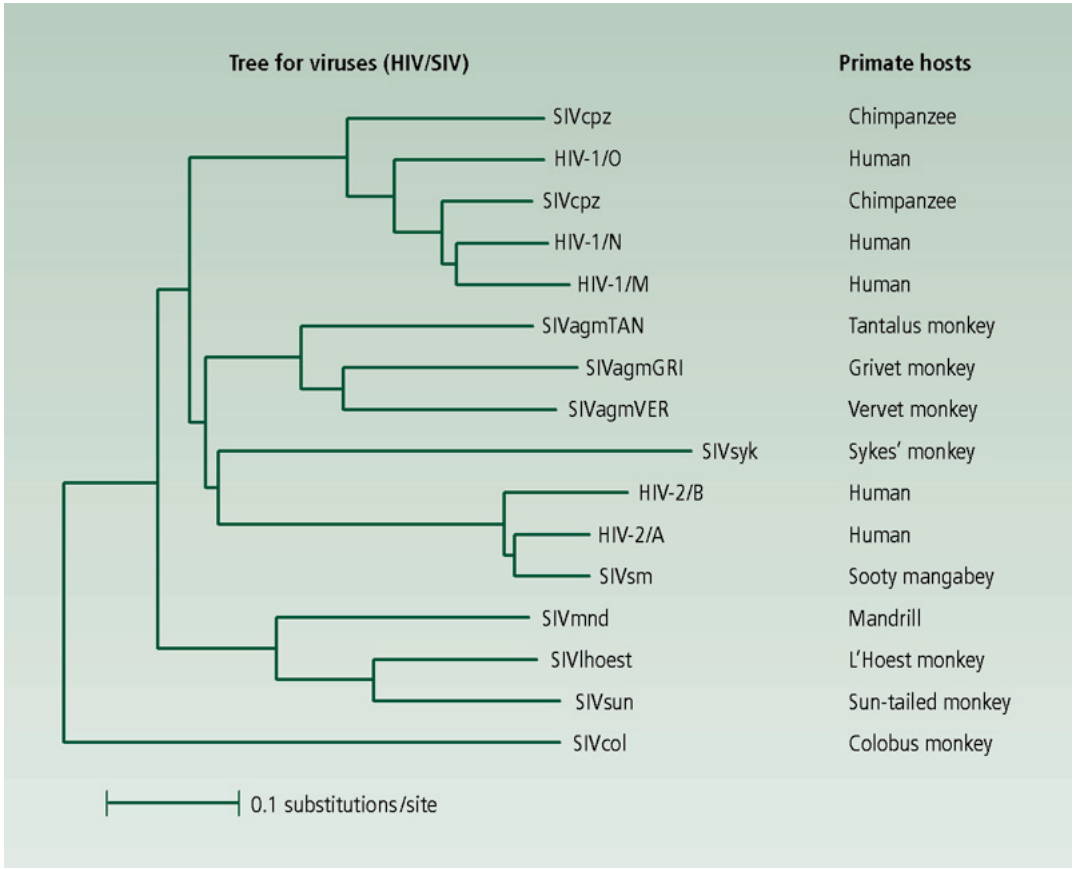
Προέλευση -

Επιδημιολογία

Έλεγχος εξελικτικών υποθέσεων

Από που προήλθε ο ιός HIV;





Πρωτοεμφανίστηκε μυστηριωδώς στις αρχές της δεκαετίας του 1980.

Ο τύπος HIV-1 εισήλθε στους ανθρώπους, ίσως περισσότερες από μια φορές, από τον χιμπατζή.

Ο τύπος HIV-2 εισήλθε στους ανθρώπους, από τους sooty mangabees

Έλεγχος εξελικτικών υποθέσεων

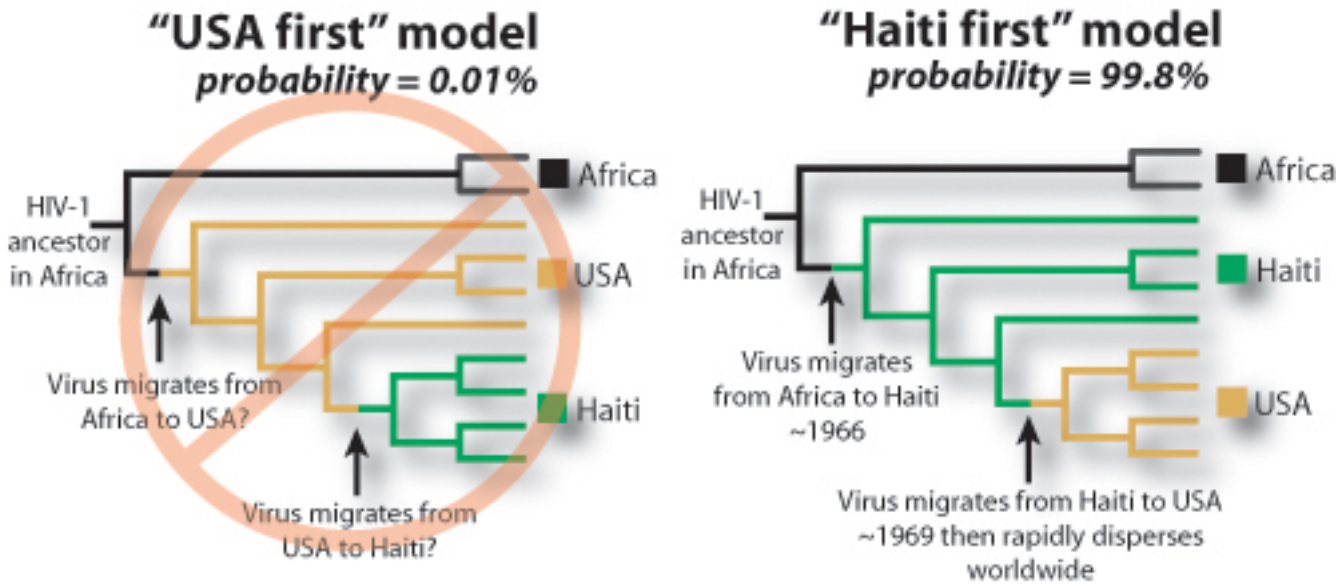
Από που προήλθε ο ιός HIV-1 subtype M; Προέλευση στην Κεντρική Αφρική.

Όταν πρωτοεντοπίστηκε, αρκετοί ασθενείς στην Αμερική ήταν πρόσφατοι Αϊτινοί μετανάστες.

Κάποιοι ισχυρίζονταν ότι πήγε από την Αμερική στην Αϊτή στα μέσα των 70s, λόγω σεξοτουρισμού.

Από την Αϊτή στην Αμερική ή το αντίθετο;

Ο Worobey χρησιμοποίησε ακολουθίες HIV από συντηρημένα δείγματα Αϊτινών ασθενών (1983)



Επιδημία χολέρας στην Αϊτή 2010

•Μετά τον σεισμό στην Αϊτή (Ιανουαριος 2010), ξέσπασε επιδημία χολέρας (Οκτώβριος 2010).
•Το βακτήριο Vibrio cholerae ελευθερώνει μια τοξίνη που προκαλεί έντονες διάρροιες και αφυδάτωση, έως και θάνατο, εντός ολίγων ωρών, αν δεν αντιμετωπιστεί!
•Η μετάδοση γίνεται όταν τα κόπρανα ενός μολυσμένου ατόμου έρθουν σε επαφή με πόσιμο νερό ή τροφή.
•Τα άτομα που δεν παράγουν αρκετό γαστρικό υγρό στο στομάχι τους, ή τα άτομα με ομάδα αίματος Ο είναι πιο ευάλωτα.
•Το Vibrio cholerae υπάρχει σε υδάτινα περιβάλλοντα ανά την υφήλιο και εάν οι συνθήκες είναι ευνοϊκές, μπορεί να ξεσπάσει επιδημία.
•Η χολέρα είναι διαδεδομένη στην Ασία.
•Τα πρώτα κρούσματα παρατηρήθηκαν σε κεντρικές περιοχές του νησιού, στην κοιλάδα Artibonite, μια εβδομάδα μετά την έλευση Νεπαλέζων κυανόκρανων, κοντά στο στρατόπεδό τους.
•Λύμματα από το στρατόπεδο κατέληγαν σε γειτονικό ποταμό.
•Οι κάτοικοι κατηγόρησαν τον ΟΗΕ ότι
    •οι κυανόκρανοι που ήρθαν να βοηθήσουν ευθύνονται για το ξέσπασμα της επιδημίας.
    •ότι ο ΟΗΕ προσπάθησε να αποκρύψει το γεγονός και να μην αναλάβει τις ευθύνες του

Ξέσπασαν ταραχές.

Εφαρμογές

Ανίχνευση οργανισμών
-
Μεταγενωμική

Μεταγενωμική

- Παράλληλη ανίχνευση όλων των οργανισμών (μικροβιακών) που απαρτίζουν την υπό μελέτη οικολογική κοινότητα.

- Υπάρχει προοπτική να χρησιμοποιηθεί για περιβαλλοντικές μελέτες/ αναλύσεις/παρακολούθηση (σε βάση ρουτίνας), όταν το κόστος αλληλούχισης (ή μικροσυστοιχιών) μειωθεί περισσότερο.

- Πλεονέκτημα: Δεν χρειάζεται να καλλιεργηθούν
  - Κλινικά δείγματα
  - Περιβαλλοντικά δείγματα

Genome assembly

Key steps in de novo assembly

1. Find reads that overlap by a specified number of bases (the k-mer size)

2. Merge overlapping, "good" reads into longer contigs

3. Link contigs to form scaffolds using paired-end information

Diagrams from S. Batzoglou, Stanford

In vitro διαγνωστικά τεστ που βασίζονται σε μικροσυστοιχίες

FDA: In Vitro Diagnostic Multivariate Index Assays (IVDMIAs)

- FDA's **In Vitro Diagnostic Product Database**
- http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfivd/index.cfm

- http://www.ivdtechnology.com/article/exploring-fda-approved-ivdmias

- Some IVDMIAs are laboratory-developed tests (LDTs). LDTs are tests that are developed by a single clinical laboratory for use only in that laboratory.

- http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm079148.htm

- IVDMIAs raise significant issues of safety and effectiveness. These types of tests are developed based on observed correlations between multivariate data and clinical outcome, such that the clinical validity of the claims is not transparent to patients, laboratorians, and clinicians who order these tests. Additionally, IVDMIAs frequently have a high risk intended use. FDA is concerned that patients are relying upon IVDMIAs with high risk intended uses to make critical healthcare decisions when FDA has not ensured that the IVDMIA has been clinically validated and the healthcare practitioners are unable to clinically validate the test themselves. Therefore, there is a need for FDA to regulate these devices to ensure that the IVDMIA is safe and effective for its intended use.

Mammaprint - Tissue of origin

- http://www.ivdtechnology.com/article/exploring-fda-approved-ivdmias
- **MammaPrint.**
  The first IVDMIA, the MammaPrint system, made by Agendia Inc., is a qualitative IVD test service performed in a single lab outside the United States using a 70-gene expression profile of fresh frozen breast cancer tissue samples to assess a breast cancer patient¬s risk for distant metastasis. FDA approved MammaPrint in February 2007 under de novo classification procedures.

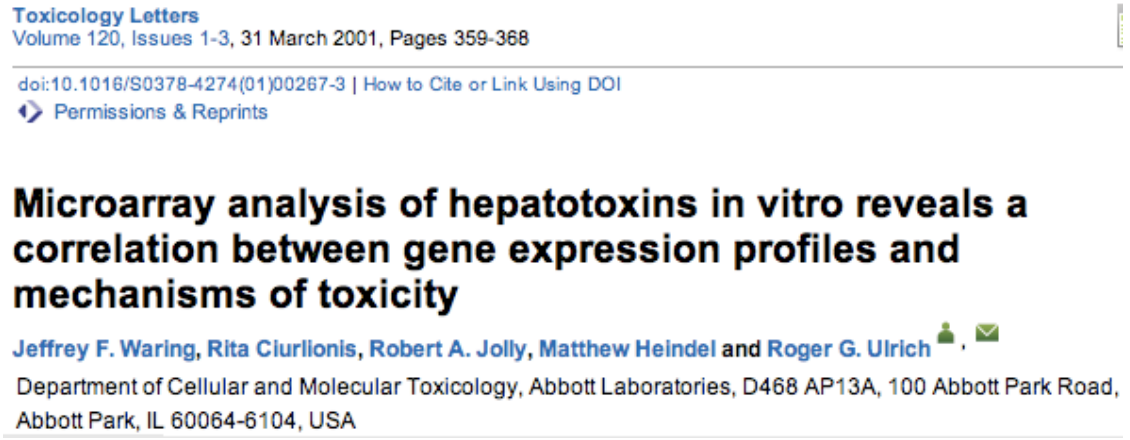
- **Tissue of Origin Test**
  In July 2008, the Tissue of Origin Test, made by Pathwork Diagnostics, was cleared. This microarray RNA profiling test is to be used on clinical, formalin-fixed, paraffin-embedded (FFPE) biopsy tissue to aid in the classification of the origin of the tumor tissue. In June 2010 a second clearance introduced a different specimen and specimen-preparation method, and the algorithm for analysis of the expression data to create a diagnostics report and interpretation. The test uses microarray technology by Affymetrix Inc. and advanced analytics to measure the gene-expression patterns of challenging tumors, including metastatic, poorly differentiated, and undifferentiated cancer. It is intended to measure the degree of similarity between the RNA expression patterns in a patient’s tumor tissue with the RNA expression patterns in a database of fifteen known tumor types.

Εφαρμογές στην Τοξικολογία

Εφαρμογές στην τοξικολογία/ τοξικογενωμική

- Μέτρηση της γονιδιακής έκφρασης μετά από έκθεση σε τοξικό παράγοντα μπορεί να δείξει τον μοριακό μηχανισμό δράσης του παράγοντα.
- Μπορεί να αποτελέσει μοναδική μοριακή υπογραφή του συγκεκριμένου τοξικού παράγοντα, για μελλοντική ανίχνευσή του.
  - Ομαδοποίηση τοξικών παραγόντων με κοινή δράση, με βάση την ομοιότητα των μοριακών προφιλ τους

Μοριακό προφίλ τοξικότητας



Toxicology Letters
Volume 120, Issues 1-3, 31 March 2001, Pages 359-368

doi:10.1016/S0378-4274(01)00267-3 | How to Cite or Link Using DOI
Permissions & Reprints

Microarray analysis of hepatotoxins in vitro reveals a correlation between gene expression profiles and mechanisms of toxicity

Jeffrey F. Waring, Rita Ciurlionis, Robert A. Jolly, Matthew Heindel and Roger G. Ulrich

Department of Cellular and Molecular Toxicology, Abbott Laboratories, D468 AP13A, 100 Abbott Park Road, Abbott Park, IL 60064-6104, USA

- Ηπατοκύτταρα αρουραίων εκτέθηκαν σε 15 γνωστές ηπατο-τοξίνες.
- Για την κάθε μία, δημιουργήθηκε το μοριακό προφίλ γονιδιακής έκφρασης (ποιά γονίδια υπερ/υπο-εκφράστηκαν).
- Τοξίνες με παρόμοιο μηχανισμό δράσης είχαν παρόμοια (όχι όμως ακριβώς ίδια) προφίλ γονιδιακής έκφρασης και ομαδοποιούνταν.

Μοριακό προφίλ τοξικότητας

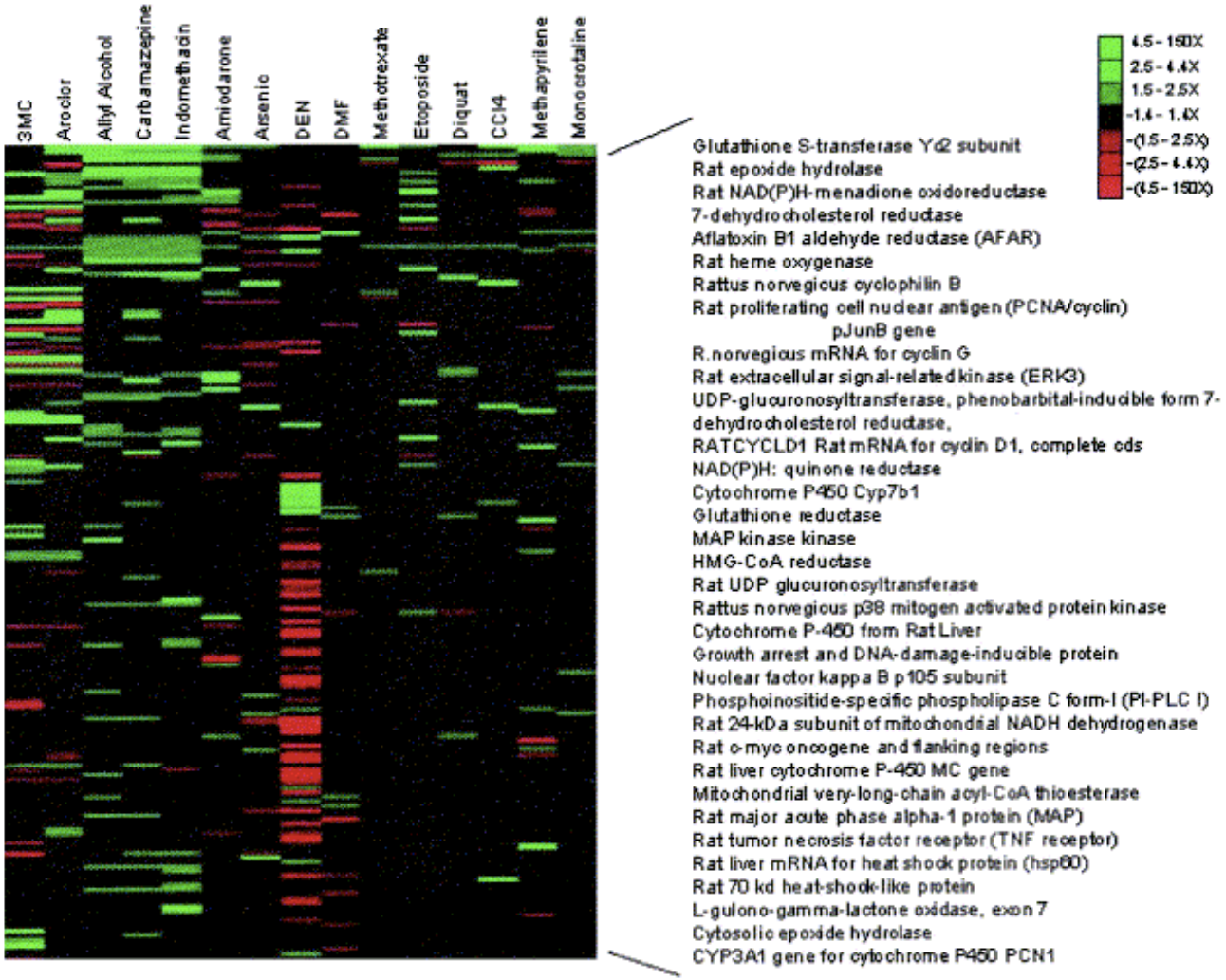


Fig. 2. Graph showing the gene changes occurring in livers from rats treated with the 15 known hepatotoxins. A total of 179 genes were shown to be regulated at least two-fold by at least one compound. Some of these genes are shown to the right of the figure.

Μοριακό προφίλ τοξικότητας

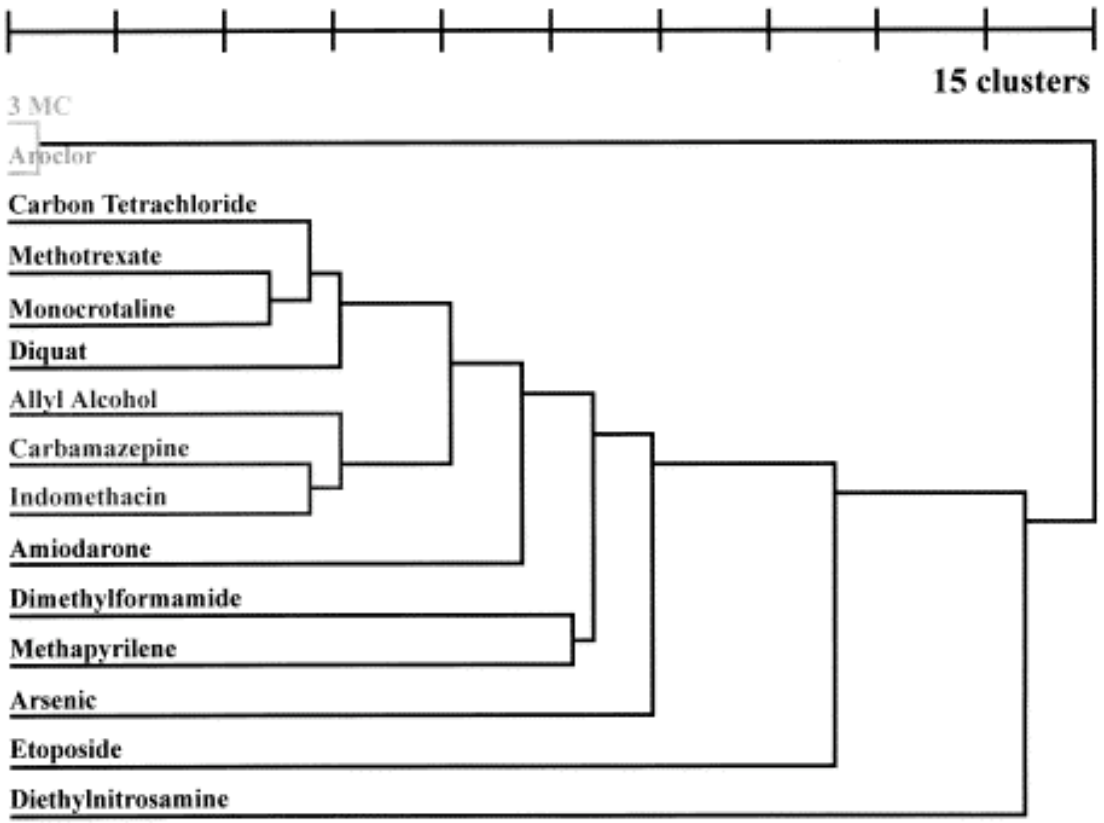


Fig. 3. Dendrogram showing the clustering of the hepatotoxins based on gene regulation. The clustering was hierarchical using correlation as the distance (see Section 2).

Hierarchical cluster analysis showed a close association in gene expressional responses between aroclor 1254 and 3-methylcholanthrene.

Βάσεις Δεδομένων

Βάσεις Δεδομένων: Εισαγωγή

Χρησιμοποιούνται για:

- Οργάνωση
- Αποθήκευση
- Επεξεργασία
- Αναζήτηση/επαναπόκτηση

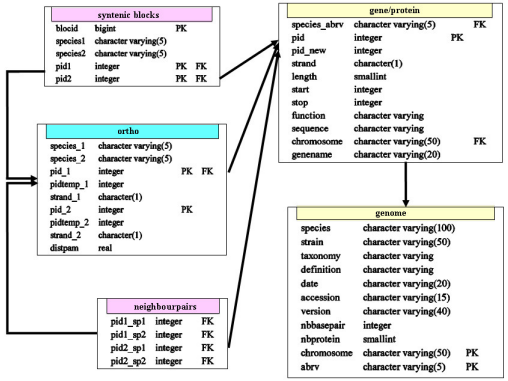
της βιολογικής πληροφορίας

```
LOCUS       name of locus, length and type of sequence,
            classification of organism, data of entry
DEFINITION  description of entry
ACCESSION   accession numbers of original source
KEYWORDS    key words for cross referencing this entry
SOURCE      source organism of DNA
ORGANISM    description of organism
REFERENCE
COMMENT     biological function or database information
FEATURES    information about sequence by base position or range of positions
       source          range of sequence, source organism
       misc_signal     range of sequence, type of function or signal
       mRNA            range of sequence, mRNA
       CDS             range of sequence, protein coding region
       intron          range of sequence, position of intron
       mutation        sequence position, change in sequence for mutation
BASE COUNT  count of A, C, G, T and other symbols
ORIGIN      text indicating start of sequence
     1 gaattcgata aatctctggt ttattgtgca gtttatggtt ccaaaatcgc
    51 atatactcac agcataactg tatatacacc cagggggcgg aatgaaagcg
//              database symbol for end of sequence
```

Figure 2.5. GenBank DNA sequence entry.

Κύρια είδη:

Επίπεδης οργάνωσης (Flat-files:) Το ποιό απλό είδος. Ουσιαστικά είναι κατάλογοι



Σχεσιακές βάσεις. Πιο περίπλοκες και πλέον πολύ διαδεδομένες . Π.χ., SQL. Η πληροφορία οργανώνεται σε πίνακες που σχετίζονται μεταξύ τους. Έτσι αποφεύγεται η επανάληψη και συσσώρευση δεδομένων

Αντικειμενοστρεφείς βάσεις κ.α.

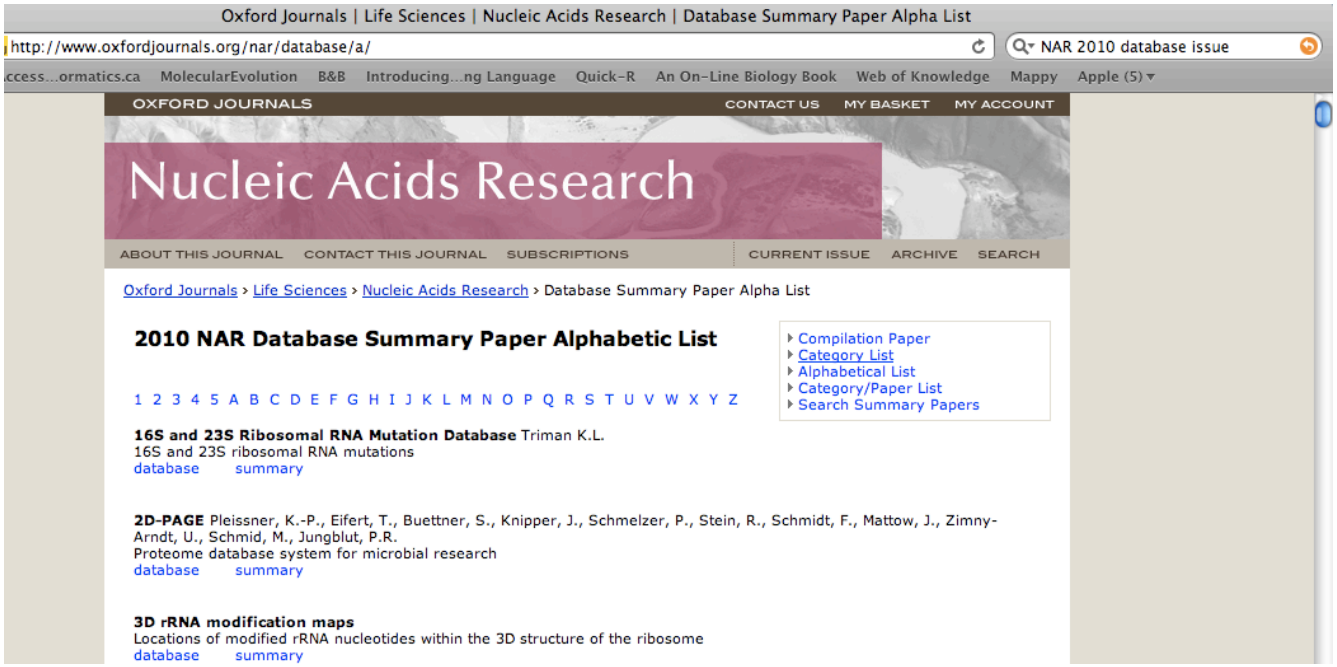
Διακρίνονται κυρίως σε αρχειακές/πρωτεύοντες και δευτερεύοντες

Στις αρχειακές γίνεται κατάθεση δεδομένων ενώ στις δευτερεύοντες τα δεδομένα είναι περαιτέρω επεξεργασμένα/σχολιασμένα/ αλληλοσυνδεδεμένα

Ετήσιος κατάλογος Β.Δ.

http://www.oxfordjournals.org/nar/database/a/

- Κάθε Ιανουάριο στο Nucleic Acids Research (Special database issue)
- 2010: 58 νέες και 73 ανανεωμένες
- Σύνολο: 1230
- 5% ετήσια ανάπτυξη
- Επίσης υπάρχει το περιοδικό Database: the journal of biological databases and curation



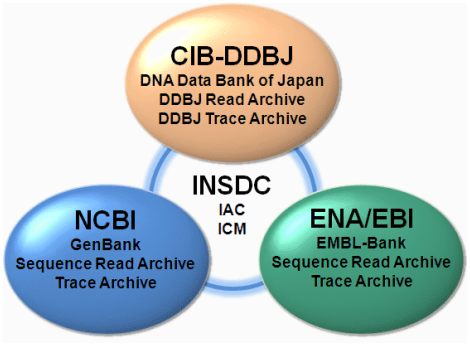
Βάσεις νουκλεοτιδικών δεδομένων (Ι)

- Αρχειακές ΒΔ για νουκλεοτιδικές αλληλουχίες:
  - EMBL-BANK. European Nucleotide Archive (ENA), EBI. Hinxton, UK.
  - GENBANK. NCBI, NIH. Bethesda, USA
  - DNA databank of Japan (DDBJ). National institute of Genetics,.Mishima, JP

- Η ακολουθία κατατίθεται σε μία απο τις ΒΔ, η οποία έχει και την δυνατότητα να την αναθεωρήσει (μόνο αυτή, για αποτροπή 'συγκρούσεων')

- Και οι 3 ΒΔ ανήκουν στο International nucleotide sequence database collection (INSDC). Κάθε μέρα ανταλλάσουν δεδομένα. Η ίδια ακολουθία Χ3. Νέα έκδοση ανά δίμινο.

- Από το 2009, το INSDC ξεκίνησε να καταχωρεί και αμορφοποίητα δεδομένα από μεγάλης κλίμακας αλληλουχίσεις (Sequencing projects), είτε αυτά προέρχονται από κλασσικές μεθόδους αλληλούχισης (Trace archive) (capillary sequencing), είτε από μεθόδους αλληλούχισης 2ης γενιάς (Read Archive) (454, Solexa, Solid, Helicos)



EMBL bank help page

http://www.ebi.ac.uk/embl/Documentation/User_manual/usrman.html

Note that each line begins with a two-character line code, which indicates the type of information contained in the line. The currently used line types, along with their respective line codes, are listed below:

```
ID - identification             (begins each entry; 1 per entry)
AC - accession number           (>=1 per entry)
PR - project identifier         (0 or 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)
```

Note that some entries will not contain all of the line types, and some line types occur many times in a single entry. As indicated, each entry begins with an identification line (ID) and ends with a terminator line (//). The various line types appear in entries in the order in which they are listed above (except for XX lines which may appear anywhere between the ID and SQ lines). A detailed description of each line type is given in the following sections.

Βάσεις νουκλεοτιδικών δεδομένων. EMBL format (i)

```
ID   X56734; SV 1; linear; mRNA; STD; PLN; 1859 BP.
XX
AC   X56734; S46826;
XX
DT   12-SEP-1991 (Rel. 29, Created)
DT   25-NOV-2005 (Rel. 85, Last updated, Version 11)
XX
DE   Trifolium repens mRNA for non-cyanogenic beta-glucosidase
XX
KW   beta-glucosidase.
XX
OS   Trifolium repens (white clover)
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Trifolium.
XX
RN   [5]
RP   1-1859
RX   PUBMED; 1907511.
RA   Oxtoby E., Dunn M.A., Pancoro A., Hughes M.A.;
RT   "Nucleotide and derived amino acid sequence of the cyanogenic
RT   beta-glucosidase (linamarase) from white clover (Trifolium repens L.)";
RL   Plant Mol. Biol. 17(2):209-219(1991).
XX
RN   [6]
RP   1-1859
RA   Hughes M.A.;
RT   ;
RL   Submitted (19-NOV-1990) to the EMBL/GenBank/DDBJ databases.
RL   Hughes M.A., University of Newcastle Upon Tyne, Medical School, Newcastle
RL   Upon Tyne, NE2 4HH, UK
XX
```

Βάσεις νουκλεοτιδικών δεδομένων. EMBL format (ii)

```
FH   Key             Location/Qualifiers
FH
FT   source          1..1859
FT                   /organism="Trifolium repens"
FT                   /mol_type="mRNA"
FT                   /clone_lib="lambda gt10"
FT                   /clone="TRE361"
FT                   /tissue_type="leaves"
FT                   /db_xref="taxon:3899"
FT   CDS             14..1495
FT                   /product="beta-glucosidase"
FT                   /EC_number="3.2.1.21"
FT                   /note="non-cyanogenic"
FT                   /db_xref="GOA:P26204"
FT                   /db_xref="HSSP:P26205"
FT                   /db_xref="InterPro:IPR001360"
FT                   /db_xref="UniProtKB/Swiss-Prot:P26204"
FT                   /protein_id="CAA40058.1"
FT                   /translation="MDFIVAIFALFVISSFTITSTNAVEASTLLDIGNLSRSSFPRGFI
FT                   FGAGSSAYQFEGAVNEGGRGPSIWDTFTHKYPEKIRDGSNADITVDQYHRYKEDVGIMK
FT                   DQNMDSYRFSISWPRILPKGKLSGGINHEGIKYYNNLINELLANGIQPFVTLFHWDLPQ
FT                   VLEDEYGGFLNSGVINDFRDYTDLCFKEFGDRVRYWSTLNEPWVFSNSGYALGTNAPGR
FT                   CSASNVAKPGDSGTGPYIVTHNQILAHAEAVHVYKTKYQAYQKGKIGITLVSNWLMPLD
FT                   DNSIPDIKAAERSLDFQFGLFMEQLTTGDYSKSMRRIVKNRLPKFSKFESSLVNGSFDF
FT                   IGINYYSSSYISNAPSHGNAKPSYSTNPMTNISFEKHGIPLGPRAASIWIYVYPYMFIQ
FT                   EDFEIFCYILKINITILQFSITENGMNEFNDATLPVEEALLNTYRIDYYYRHLYYIRSA
FT                   IRAGSNVKGFYAWSFLDCNEWFAGFTVRFGLNFVD"
FT   mRNA            1..1859
FT                   /experiment="experimental evidence, no additional details
FT                   recorded"
XX
```

Βάσεις νουκλεοτιδικών δεδομένων. EMBL format (ii)

```
SQ   Sequence 1859 BP; 609 A; 314 C; 355 G; 581 T; 0 other;
aaacaaacca aatatggatt ttattgtagc catatttgct ctgtttgtta ttagctcatt        60
cacaattact tccacaaatg cagttgaagc ttctactctt cttgacatag gtaacctgag       120
tcggagcagt tttcctcgtg gcttcatctt tggtgctgga tcttcagcat accaatttga       180
aggtgcagta aacgaaggcg gtagaggacc aagtatttgg gataccttca cccataaata       240
tccagaaaaa ataagggatg gaagcaatgc agacatcacg gttgaccaat atcaccgcta       300
caaggaagat gttgggatta tgaaggatca aaatatggat tcgtatagat tctcaatctc       360
ttggccaaga atactcccaa agggaaagtt gagcggaggc ataaatcacg aaggaatcaa       420
atattacaac aaccttatca acgaactatt ggctaacggt atacaaccat ttgtaactct       480
ttttcattgg gatcttcccc aagtcttaga agatgagtat ggtggtttct taaactccgg       540
tgtaataaat gattttcgag actatacgga tctttgcttc aaggaatttg gagatagagt       600
gaggtattgg agtactctaa atgagccatg ggtgtttagc aattctggat atgcactagg       660
aacaaatgca ccaggtcgat gttcggcctc caacgtggcc aagcctggtg attctggaac       720
aggaccttat atagttacac acaatcaaat tcttgctcat gcagaagctg tacatgtgta       780
taagactaaa taccaggcat atcaaaaggg aaagataggc ataacgttgg tatctaactg       840
gttaatgcca cttgatgata atagcatacc agatataaag gctgccgaga gatcacttga       900
cttccaattt ggattgttta tggaacaatt aacaacagga gattattcta agagcatgcg       960
gcgtatagtt aaaaaccgat tacctaagtt ctcaaaattc gaatcaagcc tagtgaatgg      1020
ttcatttgat tttattggta taaactatta ctcttctagt tatattagca atgcccttc       1080
acatggcaat gccaaaccca gttactcaac aaatcctatg accaatattt catttgaaaa      1140
acatgggata cccttaggtc caagggctgc ttcaatttgg atatatgttt atccatatat      1200
gtttatccaa gaggacttcg agatcttttg ttacatatta aaaataaata taacaatcct      1260
gcaattttca atcactgaaa atggtatgaa tgaattcaac gatgcaacac ttccagtaga      1320
agaagctctt ttgaatactt acagaattga ttactattac cgtcacttat actacattcg      1380
ttctgcaatc agggctggct caaatgtgaa gggtttttac gcatggtcat ttttggactg      1440
taatgaatgg tttgcaggct ttactgttcg ttttggatta aactttgtag attagaaaga      1500
tggattaaaa aggtacccta agctttctgc ccaatggtac aagaactttc tcaaaagaaa      1560
ctagctagta ttattaaaag aactttgtag tagattacag tacatcgttt gaagttgagt      1620
tggtgcacct aattaaataa aagaggttac tcttaacata tttttaggcc attcgttgtg      1680
aagttgttag gctgttattt ctattatact atgttgtagt aataagtgca ttgttgtacc      1740
agaagctatg atcataacta taggttgatc cttcatgtat cagtttgatg ttgagaatac      1800
tttgaattaa aagtcttttt ttatttttt aaaaaaaaaa aaaaaaaaaa aaaaaaaaa        1859
//
```

Βάσεις νουκλεοτιδικών δεδομένων. FASTA format

```
>ENA|X56734|X56734.1 Trifolium repens mRNA for non-cyanogenic beta-glucosidase
aaacaaaccaaatatggattttattgtagccatatttgctctgtttgttattagctcatt
cacaattacttccacaaatgcagttgaagcttctactcttcttgacataggtaacctgag
tcggagcagttttcctcgtggcttcatctttggtgctggatcttcagcataccaatttga
aggtgcagtaaacgaaggcggtagaggaccaagtatttgggataccttcacccataaata
tccagaaaaaataagggatggaagcaatgcagacatcacggttgaccaatatcaccgcta
caaggaagatgttgggattatgaaggatcaaaatatggattcgtatagattctcaatctc
ttggccaagaatactcccaaagggaaagttgagcggaggcataaatcacgaaggaatcaa
atattacaacaaccttatcaacgaactattggctaacggtatacaaccatttgtaactct
ttttcattgggatcttccccaagtcttagaagatgagtatggtggtttcttaaactccgg
tgtaataaatgattttcgagactatacggatctttgcttcaaggaatttggagatagagt
gaggtattggagtactctaaatgagccatgggtgtttagcaattctggatatgcactagg
aacaaatgcaccaggtcgatgttcggcctccaacgtggccaagcctggtgattctggaac
aggaccttatatagttacacacaatcaaattcttgctcatgcagaagctgtacatgtgta
taagactaaataccaggcatatcaaaagggaaagataggcataacgttggtatctaactg
gttaatgccacttgatgataatagcataccagatataaaggctgccgagagatcacttga
cttccaatttggattgtttatggaacaattaacaacaggagattattctaagagcatgcg
gcgtatagttaaaaaccgattacctaagttctcaaaattcgaatcaagcctagtgaatgg
ttcatttgattttattggtataaactattactcttctagttatattagcaatgccccttc
acatggcaatgccaaacccagttactcaacaaatcctatgaccaatatttcatttgaaaa
acatgggataccccttaggtccaagggctgcttcaatttggatatatgtttatccatatat
gtttatccaagaggacttcgagatcttttgttacatattaaaaataaatataacaatcct
gcaattttcaatcactgaaaatggtatgaatgaattcaacgatgcaacacttccagtaga
agaagctcttttgaatacttacagaattgattactattaccgtcacttatactacattcg
ttctgcaatcagggctggctcaaatgtgaagggtttttacgcatggtcatttttggactg
taatgaatggtttgcaggctttactgttcgttttggattaaactttgtagattagaaaga
tggattaaaaaggtaccctaagctttctgcccaatggtacaagaactttctcaaaagaaa
ctagctagtattattaaaagaactttgtagtagattacagtacatcgtttgaagttgagt
tggtgcacctaattaaataaaagaggttactcttaacatatttttaggccattcgttgtg
aagttgttaggctgttatttctattatactatgttgtagtaataagtgcattgttgtacc
agaagctatgatcataactataggttgatccttcatgtatcagtttgatgttgagaatac
tttgaattaaaagtctttttttatttttttaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
```

EMBL-EBI X03635 Find Help | Feedback

Databases | Tools | Research | Training | Industry | About Us | Help | Site Index

All results *(12)*
Genomes *(2)*
Nucleotide Sequences *(7)*
Protein Sequences *(2)*
Gene Expression *(1)*

ADVANCED SEARCH

QUERY SUGGESTIONS

EBI > Search for *X03635* in *All results*

Nucleotide Sequences / EMBL Release (Normal Divisions)

X03635
Homo sapiens mRNA for oestrogen receptor

View: in ENA | in EMBL format | in SRS | in EMBL-SVA | Launch NCBI BLAST | Launch FASTA
References: Taxonomy | InterPro | Ensembl Gene | UniProtKB | EMBL-Bank (Coding Sequence) | PDBe | HGNC | Medline

View all 4 results...

Genomes / HGNC

HGNC:3467
Approved Symbol: ESR1 Discover more about this gene...
Approved Name: estrogen receptor 1
Status: Approved
Aliases: NR3A1 Era
Locus Type: gene with protein product
Chromosome: 6q24-q27
References: Medline | UniProtKB | Ensembl Gene | EMBL-Bank

Genomes / Ensembl Gene

ENSG00000091831 Discover more about this gene...
estrogen receptor 1 [Source:HGNC Symbol;Acc:3467]
Species: Homo sapiens
References: Taxonomy | Ensembl Genomes Gene | UniProtKB | Ensembl | OMIM | GO | PDBe | HGNC | EMBL-Bank

Nucleotide Sequences / ASTD

TRAN00000039246
Homo sapiens transcript, product of gene ENSG00000091831 with CDS and translation.
References: Ensembl | Taxonomy | EMBL-Bank

Nucleotide Sequences / EMBL-Bank (Coding Sequence)

CAD97416
Homo sapiens (human) hypothetical protein

View: in ENA | in EMBL format | in SRS | Launch NCBI BLAST | Launch FASTA
References: EMBL-Bank (Coding Sequence) | UniProtKB | EMBL-Bank

EMBL-EBI Enter Text Here Find Help | Feedback

Databases | Tools | Research | Training | Industry | About Us | Help | Site Index

ENA

- ENA Home
- Search & Browse
- Submit & Update
- About ENA
- Contact

Text search | Sequence search

Enter or paste text or ENA accession number: Search

Upload file of accessions: Choose File no file selected Search

EMBL-Bank: X03635.1 : Homo sapiens mRNA for oestrogen receptor

View: TEXT FASTA XML **Download:** TEXT FASTA XML

Overview Source Feature(s) Other Features Assembly References Comments Sequence Send Feedback

Organism	Molecule type	Topology	Data class	Taxonomic Division
Homo sapiens	mRNA	linear	STD	HUM

Sequence length	Sequence Version	First public	Last updated
6,450	1	18-NOV-1986	07-OCT-2008

Keywords
estrogen receptor, receptor, steroid hormone receptor.

Secondary Accession(s)
M11457.

Lineage
Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi, Mammalia, Eutheria, Euarchontoglires, Primates, Haplorrhini, Catarrhini, Hominidae, Homo

Navigation

Βάσεις πρωτεϊνικών δεδομένων

- Swissprot. 1987, Uni Geneva + SIB. Σχολιασμός των εγγραφών/ πρωτεϊνών από επιστήμονες.

- TrEMBL. 1996. SIB + EBI. Αυτόματη μετάφραση των ακολουθιών που βρίσκονται στην EMBL. Δεδομένα στην ίδια μορφή με την Swissprot. Μπορεί να είναι υποθετικές ή ο σχολιασμός να μην είναι εκτενής, όπως στην Swissprot.

- PIR. 1984, USA

- UniProt. 2002. Ενώθηκαν οι παραπάνω βάσεις.

- UniMes: για μεταγενωμικά δεδομένα, όπου δεν γνωρίζουμε από ποιά είδη προέρχονται οι ακολουθίες.

Swissprot (ι)

– Από την εγγραφή του προηγούμενου παραδείγματος, ακολουθήστε τον σύνδεσμο (link) προς την Β.Δ. UniprotKB/Swissprot, με κωδικό εγγραφής P03372

CDS	**361..2148**
product	oestrogen receptor
translation	MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYGQTGLPY GPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLENEPSGYTVREAGPPAFYRPNSDNRRQG GRERLASTNDKGSMAMESAKETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQAC RLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRDDGEGRGEVGSAGDMRAANLWPSPLMIKRSKKNSLALSLTADQMVSALL DAEPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPV KLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLD KITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAHRLHAPTSRGGASV EETDQSHLATAGSTSSHSLQKYYITGEAEGFPATV
⬇ EMBL-Bank CDS:	CAA27284
➡ GOA	P03372
➡ HGNC	3467
➡ InterPro	IPR000536, IPR001292, IPR001628, IPR001723, IPR008946, IPR013088, IPR024178
➡ PDB	1A52, 1AKF, 1ERE, 1ERR, 1G50, 1GWQ, 1GWR, 1HCP, 1HCQ, 1L2I, 1PCG, 1QKT, 1QKU, 1R5K, 1SJ0, 1UOM, 1X7E, 1X7R, 1XP1, 1XP6, 1XP9, 1XPC, 1XQC, 1YIM, 1YIN, 1ZKY, 2AYR, 2B1V, 2B1Z, 2B23, 2BJ4, 2FAI, 2G44, 2G5O, 2I0J, 2IOG, 2IOK, 2JF9, 2JFA, 2OCF, 2OUZ, 2P15, 2POG, 2Q6J, 2Q70, 2QA6, 2QA8, 2QAB, 2QE4, 2QGT, 2QGW, 2QH6, 2QR9, 2QSE, 2QXM, 2QXS, 2QZO, 2R6W, 2R6Y, 2YAT, 3CBM, 3CBO, 3CBP, 3DT3, 3ERD, 3ERT, 3HLV, 3HM1, 3L03, 3OS8, 3OS9, 3OSA
➡ UniProtKB/Swiss-Prot	P03372

P03372 (ESR1_HUMAN) Reviewed, UniProtKB/Swiss-Prot

Last modified September 21, 2011. Version 176. This entry in the past...

Clusters with 100%, 90%, 50% identity | Documents (6) | Third-party data
text xml rdf/xml gff fasta

Names · Attributes · General annotation · Ontologies · Interactions · Alt products · Sequence annotation · Sequences · References · Web links · Cross-refs · Entry info · Documents · Customize order

Names and origin

Protein names	*Recommended name:* **Estrogen receptor** Short name=ER *Alternative name(s):* ER-alpha Estradiol receptor Nuclear receptor subfamily 3 group A member 1
Gene names	Name: **ESR1** Synonyms:ESR, NR3A1
Organism	**Homo sapiens (Human)**
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo

Protein attributes

Sequence length	595 AA.
Sequence status	Complete.
Protein existence	Evidence at protein level

ΒΔ πρωτεϊνικών επικρατειών

- Πρωτεϊνική επικράτεια: Μια περιοχή της πρωτεΐνης με συγκεκριμένη λειτουργία/ δομή και καλά συντηρημένη.

- Διάφορες βάσεις δεδομένων, όπως:
  - PROSITE
  - Pfam
  - PRINTS
  - ProDom
  - SMART
  - TIGRFAMs
  - PIR superfamily
  - Superfamily

- Έχουν ενσωματωθεί στο INTERPRO.

- Το INTERPRO περιέχει πρωτεϊνικές επικράτειες. Το πρόγραμμα INTERPROscan ανιχνεύει αυτές τις επικράτειες στις πρωτεΐνες.

Family and domain databases

InterPro	IPR008946. Nucl_hormone_rcpt_ligand-bd. IPR000536. Nucl_hrmn_rcpt_lig-bd_core. IPR001292. Oestr_rcpt. IPR024178. Oestrogen_rcpt-rel. IPR001723. Str_hrmn_rcpt. IPR001628. Znf_hrmn_rcpt. IPR013088. Znf_NHR/GATA. [Graphical view]
Gene3D	G3DSA:1.10.565.10. Nucl_hrmn_rcpt_lig_bd. 1 hit. G3DSA:3.30.50.10. Znf_NHR/GATA. 1 hit.
Pfam	PF00104. Hormone_recep. 1 hit. PF02159. Oest_recep. 1 hit. PF00105. zf-C4. 1 hit. [Graphical view]
PIRSF	PIRSF002527. ER-like_NR. 1 hit.
PRINTS	PR00543. OESTROGENR. PR00398. STRDHORMONER. PR00047. STROIDFINGER.
SMART	SM00430. HOLI. 1 hit. SM00399. ZnF_C4. 1 hit. [Graphical view]
SUPFAM	SSF48508. Str_ncl_receptor. 1 hit.
PROSITE	PS00031. NUCLEAR_REC_DBD_1. 1 hit. PS51030. NUCLEAR_REC_DBD_2. 1 hit. [Graphical view]
ProtoNet	Search...

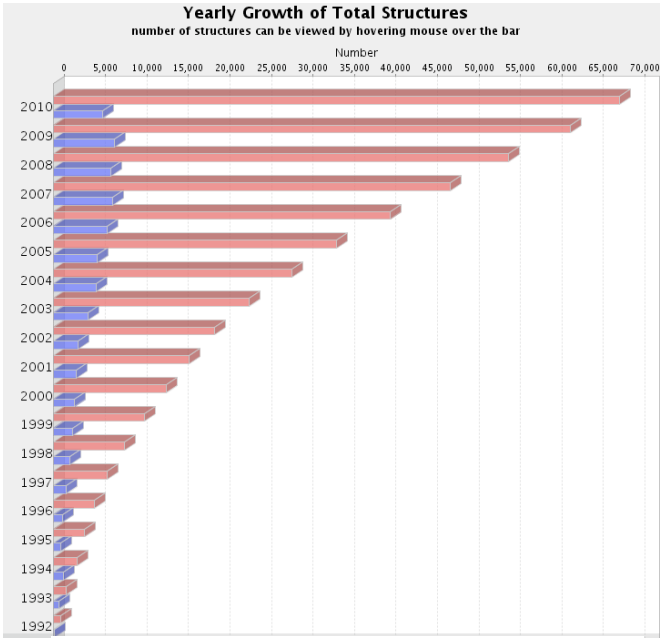
ΒΔ τρισδιάστατων δομών PDB

- Protein Data Bank (PDB)
  - Πρωτεΐνες
  - Νουκλεϊκά οξέα
  - Σύμπλοκα των παραπάνω

- Μέθοδοι
  - X-ray (~59000)
  - NMR (~8500)
  - Κρύο-ηλεκτρονική μικροσκοπία (~300)

- Οι παραπάνω μέθοδοι βρίσκουν τις συντεταγμένες (3D) των ατόμων του βιολογικού μορίου.

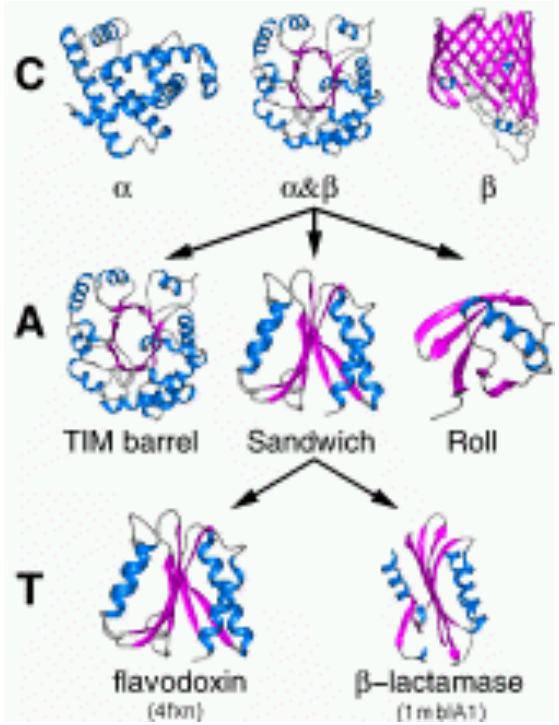
- Τα αρχεία με τις συντεταγμένες διαβάζονται από ειδικά προγράμματα (π.χ Rasmol) που απεικονίζουν τη δομή στο χώρο



Β.Δ. τρισδιάστατων δομών

- CATH: κατηγοριοποιεί τις τρισδιάστατες δομές των πρωτεϊνικών επικρατειών ιεραρχικά, σε 4 βασικά επίπεδα.

- Η κατηγοριοποίηση γίνεται με ένα συνδυασμό αυτόματων μεθόδων και ανθρώπινης κρίσης.



What do the letters "C.A.T.H.S.O.L.I.D" mean?

CATH is a tree-like, hierarchical classification that starts off at the tree "trunk" by clustering protein domains into broad categories (e.g. C, or class, where domains are clustered solely based on their general secondary structure content). As the hierarchy moves away from the "trunk" to the "branches", more stringent clustering criteria are applied to provide clusters of domains with finer granularity of similarity.

Depth	Letter	Name	Clustering criteria
1	C	Class	Secondary structure content
2	A	Architecture	General spatial arrangement of secondary structures
3	T	Topology	Spatial arrangement and connectivity of secondary structures (fold)
4	H	Homologous Superfamily	Manual curation of evidence of evolutionary relationship (at least two criteria from sequence/structure/function must be observed)
5	S	Sequence Family (S35)	>= 35% sequence similarity
6	O	Orthologous Family (S60) *	>= 60% sequence similarity
7	L	"Like" domain (S95) *	>= 95% sequence similarity
8	I	Identical domain (S100)	100% sequence similarity
9	D	Domain counter	Unique domains

Βάσεις τρισδιάστατων δομών

CATH Domain: 1cukA01 XML

PDB 1cuk, Chain A, Domain 1

CATH Code	Level Description	Links
2	Mainly Beta	
2.40	Beta Barrel	
2.40.50	OB fold (Dihydrolipoamide Acetyltransferase, E2P)	
2.40.50.140	Nucleic acid-binding proteins	Gene3D
2.40.50.140.47		
2.40.50.140.47.1		
2.40.50.140.47.1.1		
2.40.50.140.47.1.1.1		
2.40.50.140.47.1.1.1.1		

What do the letters "C.A.T.H.S.O.L.I.D" mean?

CATH is a tree-like, hierarchical classification that starts off at the tree "trunk" by clustering protein domains into broad categories (e.g. C, or class, where domains are clustered solely based on their general secondary structure content). As the hierarchy moves away from the "trunk" to the "branches", more stringent clustering criteria are applied to provide clusters of domains with finer granularity of similarity.

Depth	Letter	Name	Clustering criteria
1		Class	Secondary structure content
2		Architecture	General spatial arrangement of secondary structures
3		Topology	Spatial arrangement and connectivity of secondary structures (fold)
4		Homologous Superfamily	Manual curation of evidence of evolutionary relationship (at least two criteria from sequence/structure/function must be observed)
5		Sequence Family (S35)	>= 35% sequence similarity
6		Orthologous Family (S60) *	>= 60% sequence similarity
7		"Like" domain (S95) *	>= 95% sequence similarity
8		Identical domain (S100)	100% sequence similarity
9	D	Domain counter	Unique domains

Μεταβολικά μονοπάτια

- Metabolic and Signaling Pathways
  - Enzymes and enzyme nomenclature
  - Metabolic pathways
    - BioCarta
    - BioCyc
    - Bionemo
    - BioSilico
    - BRITE - Biomolecular Relations in Information Transmission and Expression
    - BSD - Biodegradative Strain Database
    - HMDB
    - HMDB - The Human Metabolome Database
    - KEGG - Kyoto Encyclopedia of Genes and Genomes
    - Klotho
    - LIGAND
    - MedicCyc
    - MetaCrop
    - MetaCyc
    - Metagrowth
    - MMCD
    - MODOMICS
    - NMPDR - National Microbial Pathogen Data Resource
    - Pathguide
    - PMAP
    - PUMA2
    - SYSTOMONAS
    - UM-BBD
  - Protein-protein interactions
  - Signalling pathways

KEGG pathways

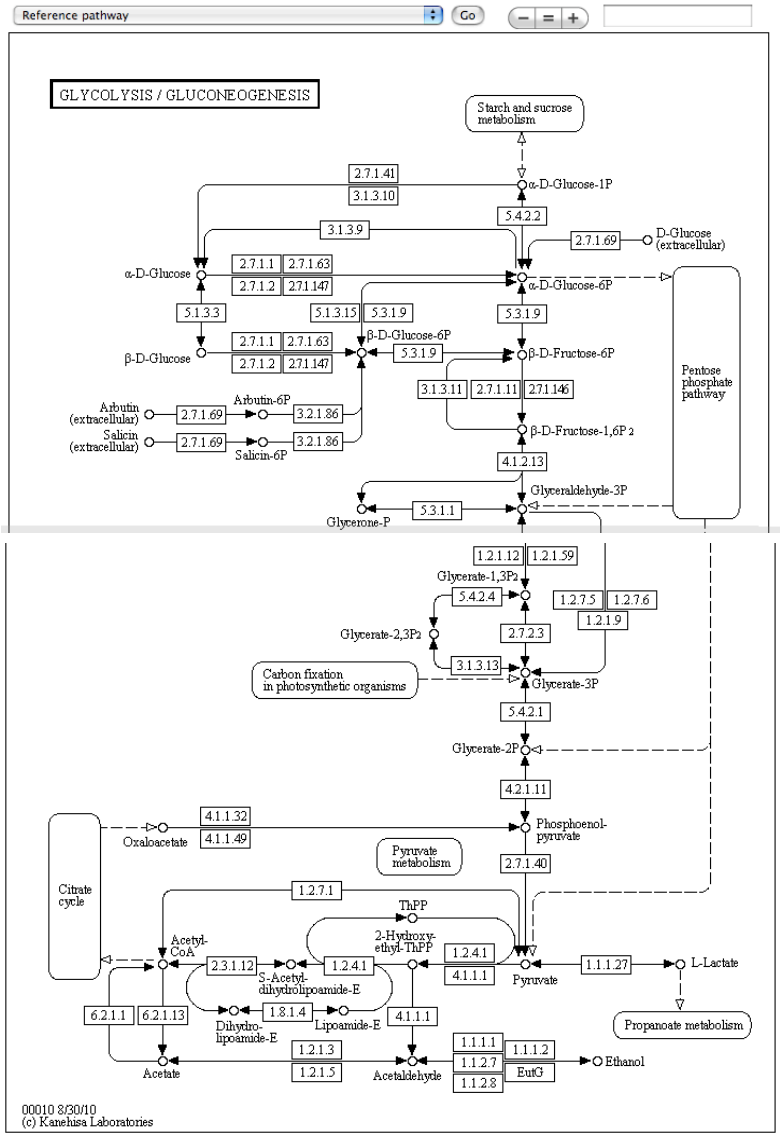
- Kyoto encyclopedia of genes and genomes.
- 2010: 374 μεταβολικά μονοπάτια.

KEGG pathways

Glycolysis / Gluconeogenesis - Reference pathway

[Pathway menu | Pathway entry | Hide description]

Glycolysis is the process of converting glucose into pyruvate and generating small amounts of ATP (energy) and NADH (reducing power). It is a central pathway that produces important precursor metabolites: six-carbon compounds of glucose-6P and fructose-6P and three-carbon compounds of glycerone-P, glyceraldehyde-3P, glycerate-3P, phosphoenolpyruvate, and pyruvate [MD:M00001]. Acetyl-CoA, another important precursor metabolite, is produced by oxidative decarboxylation of pyruvate [MD:M00679]. When the enzyme genes of this pathway are examined in completely sequenced genomes, the reaction steps of three-carbon compounds from glycerone-P to pyruvate form a conserved core module [MD:M00002], which is found in almost all organisms and which often corresponds to operon structures in bacterial genomes. Gluconeogenesis is a synthesis pathway of glucose from noncarbohydrate precursors. It is essentially a reversal of glycolysis with minor variations of alternative paths [MD:M00003].



ENZYME: 3.1.3.9

Help

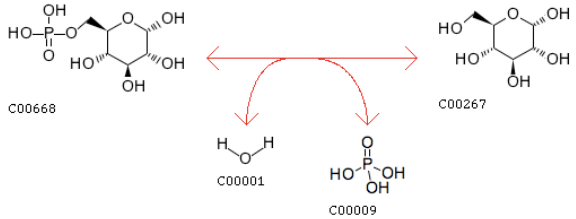
Entry	EC 3.1.3.9 Enzyme
Name	glucose-6-phosphatase; glucose 6-phosphate phosphatase
Class	Hydrolases; Acting on ester bonds; Phosphoric-monoester hydrolases BRITE hierarchy
Sysname	D-glucose-6-phosphate phosphohydrolase
Reaction(IUBMB)	D-glucose 6-phosphate + H2O = D-glucose + phosphate [RN:R00303]
Reaction(KEGG)	R00303 > R01788 Show all
Substrate	D-glucose 6-phosphate [CPD:C00092]; H2O [CPD:C00001]
Product	D-glucose [CPD:C00031]; phosphate [CPD:C00009]
Comment	Wide distribution in animal tissues. Also catalyses potent transphosphorylations from carbamoyl phosphate, hexose phosphates, diphosphate, phosphoenolpyruvate and nucleoside di- and triphosphates, to D-glucose, D-mannose, 3-methyl-D-glucose or 2-deoxy-D-glucose [cf. EC 2.7.1.62 (phosphoramidate---hexose phosphotransferase), EC 2.7.1.79 (diphosphate---glycerol phosphotransferase) and EC 3.9.1.1 (phosphoamidase)].
Pathway	ec00010 Glycolysis / Gluconeogenesis ec00052 Galactose metabolism ec00500 Starch and sucrose metabolism ec01100 Metabolic pathways
Orthology	K01084 glucose-6-phosphatase
Genes	HSA: 2538(G6PC) 57818(G6PC2) PTR: 741431(G6PC2) MCC: 712053 MMU: 14377(G6pc) RNO: 25634(G6pc) CFA: 403492(G6PC) BTA: 538710(G6PC) SSC: 100134959(G6PC)

All links

Ontology (5)
KEGG BRITE (5)
Pathway (89)
KEGG PATHWAY (89)
Disease (1)
OMIM (1)
Chemical substance (6)
KEGG COMPOUND (6)
Chemical reaction (12)
KEGG ENZYME (4)
KEGG REACTION (2)
KEGG RPAIR (5)
KEGG RCLASS (1)
Genome (2)
KEGG GENOME (2)
Gene (58)
KEGG ORTHOLOGY (1)
KEGG GENES (20)
KEGG DGENES (8)
KEGG EGENES (29)
Protein sequence (64)
UniProt (25)
PRF (3)
RefSeq(pep) (25)
PMD (11)
DNA sequence (50)
RefSeq(nuc) (30)
GenBank (10)
EMBL (10)
Protein domain (2)
InterPro (1)
Pfam (1)
Literature (3)
PubMed (3)
Enzyme (4)
BRENDA (1)
EXPASY-ENZYME (1)
EXPLORENZ (1)
IUBMB (1)
All databases (296)

REACTION: R01788

Help

Entry	R01788 Reaction
Name	alpha-D-Glucose 6-phosphate phosphohydrolase
Definition	alpha-D-Glucose 6-phosphate + H2O <=> alpha-D-Glucose + Orthophosphate
Equation	C00668 + C00001 <=> C00267 + C00009
	
RPair	RP00216 C00267_C00668 main RP05676 C00001_C00009 leave RP06709 C00009_C00668 leave
Enzyme	3.1.3.9
Pathway	rn00010 Glycolysis / Gluconeogenesis rn00052 Galactose metabolism rn00500 Starch and sucrose metabolism rn01100 Metabolic pathways
Orthology	K01084 glucose-6-phosphatase [EC:3.1.3.9]

All links

Ontology (2)
KEGG BRITE (2)
Pathway (8)
KEGG PATHWAY (8)
Chemical substance (4)
KEGG COMPOUND (4)
Chemical reaction (5)
KEGG ENZYME (1)
KEGG RPAIR (3)
KEGG RCLASS (1)
Gene (1)
KEGG ORTHOLOGY (1)
All databases (20)

Pubmed

- ΒΔ του NCBI. Ξεκίνησε τον Ιανουάριο του 1996.
- Καταχωρεί όλες τις δημοσιευμένες εργασίες που προέρχονται από τον ευρύτερο χώρο της βιοϊατρικής
- ~20 εκατομύρια εργασίες καταχωρημένες (Ιούλιος 2010)
- Όταν μια εργασία γίνεται δεκτή από το περιοδικό, κατατίθεται και στην Pubmed
- Η Pubmed δίνει ένα μοναδικό κωδικό εγγραφής (PMID) και λέξεις κλειδιά που χαρακτηρίζουν το περιεχόμενο της εργασίας (MeSH terms).
- Από το 2007, το NIH απαιτεί όποιες ερευνητικές εργασίες έχουν χρηματοδοτηθεί από αυτό, τα αποτελέσματά τους να γίνονται προσβάσιμα σε όλους, μέσω του Pubmed Central (εντός 12 μηνών από την ημερομηνία δημοσίευσης). (~ 1 εκατομύριο εργασίες)

Pubmed

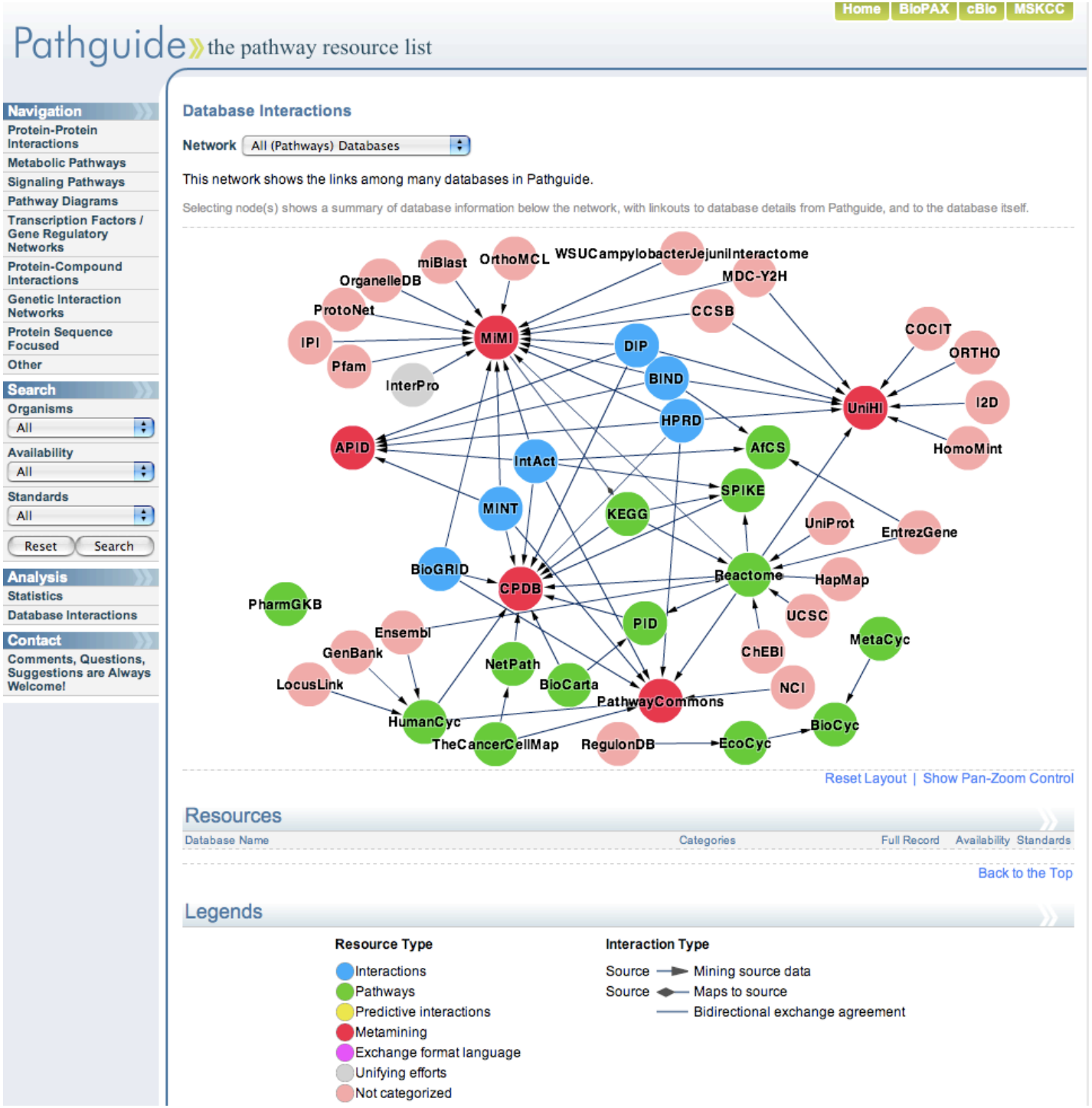

NCBI
Resources
How To
My NCBI Sign In
PubMed.gov
U.S. National Library of Medicine
National Institutes of Health
Search: PubMed
Limits
Advanced search
Help
Search
Clear
Display Settings: Abstract
Send to:
Science
Science. 1996 Oct 25;274(5287):546, 563-7.
Life with 6000 genes.
Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG.
Université Catholique de Louvain, Unité de Biochimie Physiologique, Place Croix du Sud, 2/20, 1348 Louvain-la-Neuve, Belgium.
Comment in:
Science. 1997 Feb 21;275(5303):1051-2.
Abstract
The genome of the yeast Saccharomyces cerevisiae has been completely sequenced through a worldwide collaboration. The sequence of 12,068 kilobases defines 5885 potential protein-encoding genes, approximately 140 genes specifying ribosomal RNA, 40 genes for small nuclear RNA molecules, and 275 transfer RNA genes. In addition, the complete sequence provides information about the higher order organization of yeast's 16 chromosomes and allows some insight into their evolutionary history. The genome shows a considerable amount of apparent genetic redundancy, and one of the major problems to be tackled during the next stage of the yeast genome project is to elucidate the biological functions of all of these genes.
PMID: 8849441 [PubMed - indexed for MEDLINE]
Publication Types, MeSH Terms, Substances, Grant Support
LinkOut - more resources
Related citations
Sequence analysis of a near-subtelomeric 35.4 kb DNA segment on the right arm o [Yeast. 1997]
Complete nucleotide sequence of Saccharomyces cerevisiae chror [Science. 1994]
The sequence of a 36 kb segment on the left arm of yeast chromosome X identifies 2 [Yeast. 1994]
Review Sequencing the yeast genome: an international achievement. [Yeast. 1994]
Review Complete nucleotide sequence of Saccharomyces cerevisiae chror [EMBO J. 1996]
See reviews...
See all...
Cited by over 100 PubMed Central articles
Species concepts in Calonectria (Cylindrocladium). [Stud Mycol. 2010]
Genome sequence of the necrotrophic plant pathogen Pythium ultimum [Genome Biol. 2010]
Reconstruction and validation of RefRec: a global model for the yeast mol [PLoS One. 2010]
See all...

Pubmed

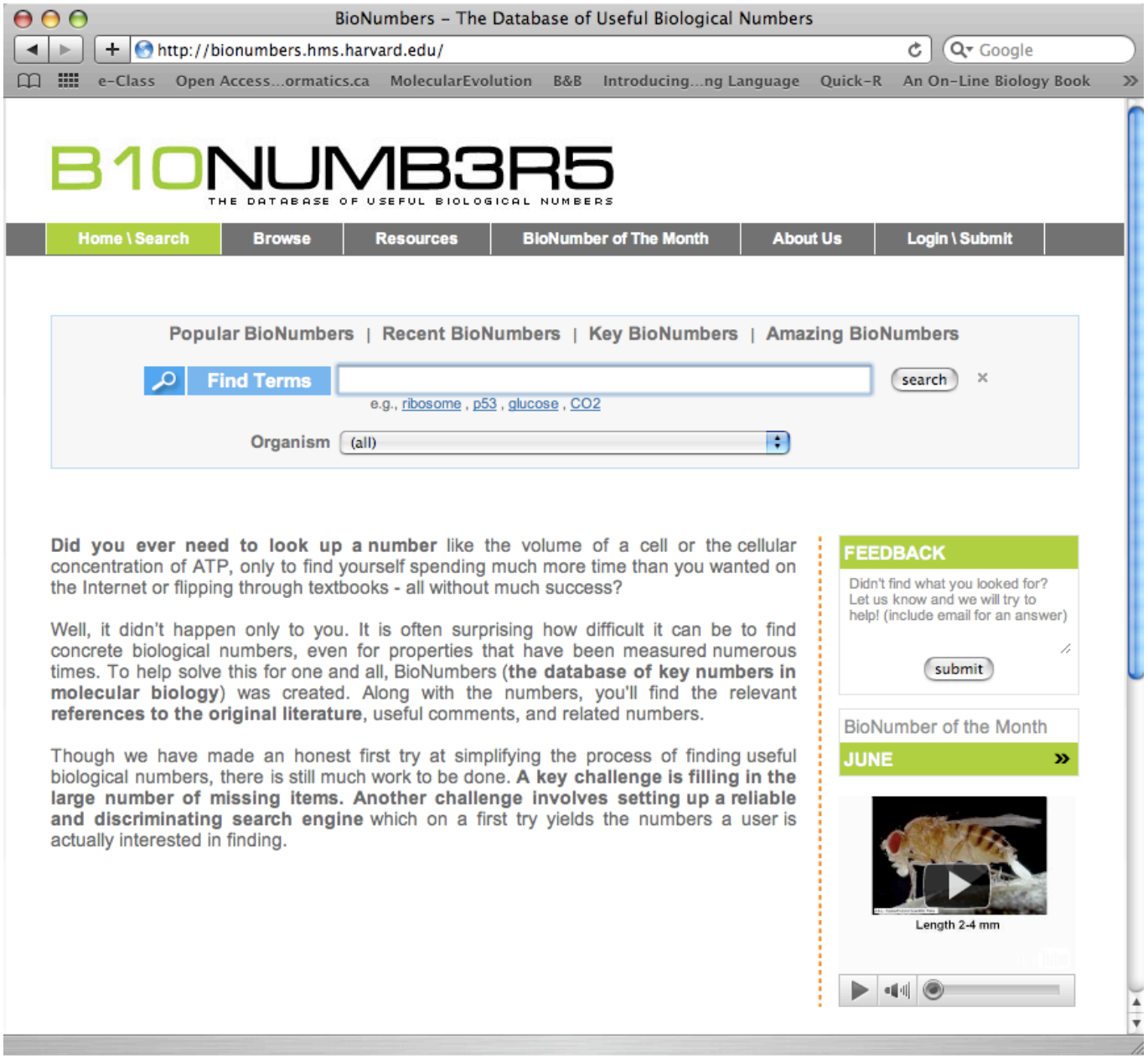
```
PMID- 8849441
OWN - NLM
STAT- MEDLINE
DA  - 19961122
DCOM- 19961122
LR  - 20090929
IS  - 0036-8075 (Print)
IS  - 0036-8075 (Linking)
VI  - 274
IP  - 5287
DP  - 1996 Oct 25
TI  - Life with 6000 genes.
PG  - 546, 563-7
AB  - The genome of the yeast Saccharomyces cerevisiae has been completely sequenced
      through a worldwide collaboration. The sequence of 12,068 kilobases defines 5885
      potential protein-encoding genes, approximately 140 genes specifying ribosomal
      RNA, 40 genes for small nuclear RNA molecules, and 275 transfer RNA genes. In
      addition, the complete sequence provides information about the higher order
      organization of yeast's 16 chromosomes and allows some insight into their
      evolutionary history. The genome shows a considerable amount of apparent genetic
      redundancy, and one of the major problems to be tackled during the next stage of
      the yeast genome project is to elucidate the biological functions of all of these
      genes.
AD  - Universite Catholique de Louvain, Unite de Biochimie Physiologique, Place Croix
      du Sud, 2/20, 1348 Louvain-la-Neuve, Belgium.
FAU - Goffeau, A
AU  - Goffeau A
FAU - Oliver, S G
AU  - Oliver SG
LA  - eng
GR  - Wellcome Trust/United Kingdom
PT  - Journal Article
PT  - Review
PL  - UNITED STATES
TA  - Science
JT  - Science (New York, N.Y.)
JID - 0404511
RN  - 0 (DNA, Fungal)
RN  - 0 (Fungal Proteins)
RN  - 0 (RNA, Fungal)
SB  - IM
CIN - Science. 1997 Feb 21;275(5303):1051-2. PMID: 9054002
MH  - Amino Acid Sequence
MH  - Base Sequence
MH  - *Chromosome Mapping
MH  - Chromosomes, Fungal/genetics
MH  - Computer Communication Networks
MH  - DNA, Fungal/genetics
MH  - Evolution, Molecular
MH  - Fungal Proteins/chemistry/genetics/physiology
MH  - Gene Library
MH  - *Genes, Fungal
MH  - *Genome, Fungal
MH  - International Cooperation
MH  - Multigene Family
MH  - Open Reading Frames
MH  - RNA, Fungal/genetics
MH  - Saccharomyces cerevisiae/*genetics
MH  - Sequence Analysis, DNA
RF  - 86
EDAT- 1996/10/25
MHDA- 1996/10/25 00:01
CRDT- 1996/10/25 00:00
PST - ppublish
SO  - Science. 1996 Oct 25;274(5287):546, 563-7.
```

Κατάλογος με ΒΔ: Pathguide

- http://www.pathguide.org/

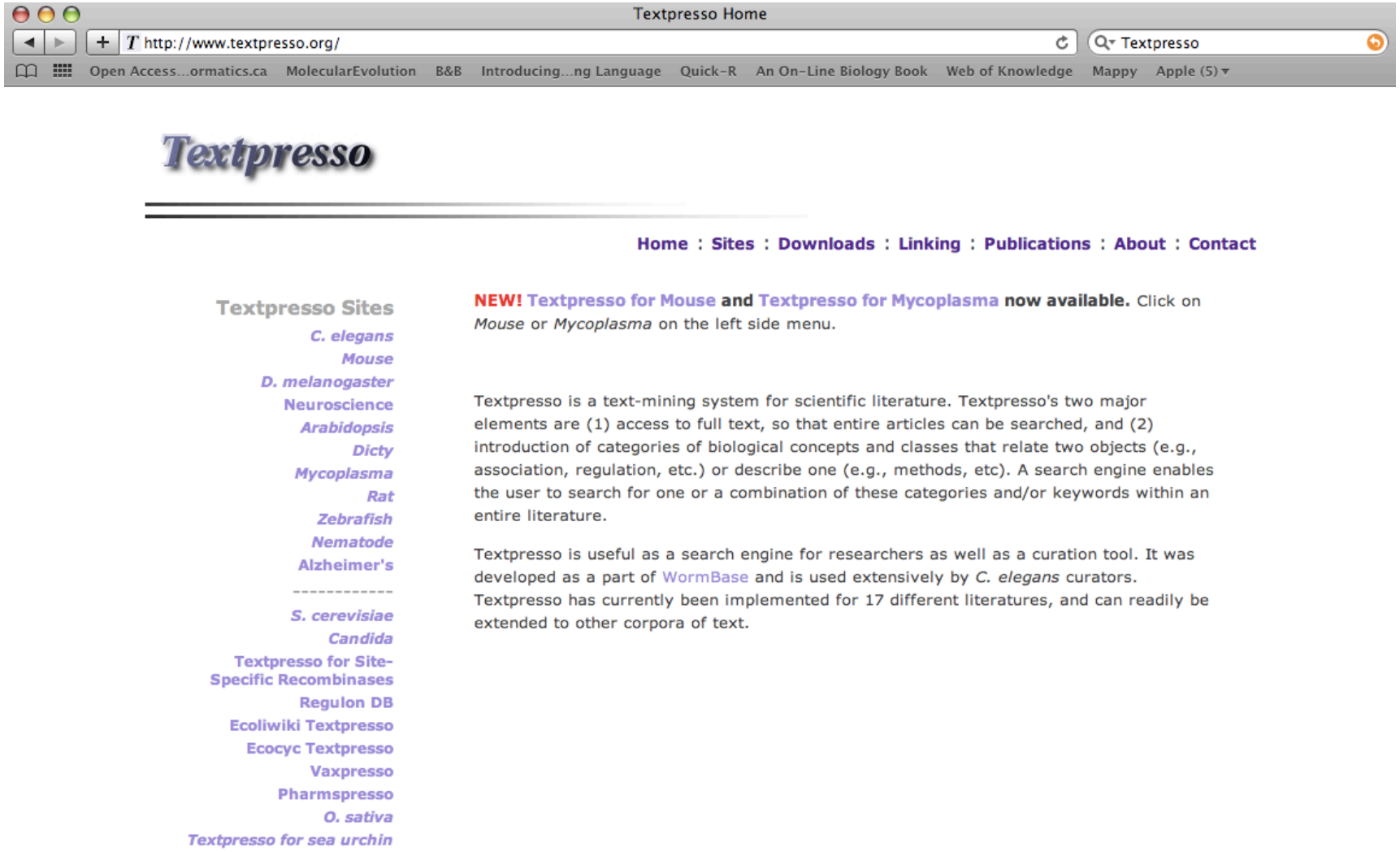


Bionumbers


BioNumbers – The Database of Useful Biological Numbers
http://bionumbers.hms.harvard.edu/
Google
e-Class
Open Access...ormatics.ca
MolecularEvolution
B&B
Introducing...ng Language
Quick-R
An On-Line Biology Book
B10NUMB3R5
THE DATABASE OF USEFUL BIOLOGICAL NUMBERS
Home \ Search
Browse
Resources
BioNumber of The Month
About Us
Login \ Submit
Popular BioNumbers | Recent BioNumbers | Key BioNumbers | Amazing BioNumbers
Find Terms
search
e.g., ribosome , p53 , glucose , CO2
Organism
(all)
Did you ever need to look up a number like the volume of a cell or the cellular concentration of ATP, only to find yourself spending much more time than you wanted on the Internet or flipping through textbooks - all without much success?
Well, it didn't happen only to you. It is often surprising how difficult it can be to find concrete biological numbers, even for properties that have been measured numerous times. To help solve this for one and all, BioNumbers (the database of key numbers in molecular biology) was created. Along with the numbers, you'll find the relevant references to the original literature, useful comments, and related numbers.
Though we have made an honest first try at simplifying the process of finding useful biological numbers, there is still much work to be done. A key challenge is filling in the large number of missing items. Another challenge involves setting up a reliable and discriminating search engine which on a first try yields the numbers a user is actually interested in finding.
FEEDBACK
Didn't find what you looked for? Let us know and we will try to help! (include email for an answer)
submit
BioNumber of the Month
JUNE
Length 2-4 mm

Textpresso

- Μηχανή αναζήτησης που ελέγχει ολόκληρο το κείμενο μιας εργασίας (full text).



Clinical tests webpages

- http://labtestsonline.org/
- Ποιά test για ποιές ασθένειες

- http://informeddna.com/index.php/
- Informed Medical Decisions, Inc. is the only nationwide network of independent genetic counselors.

Σημερινή άσκηση

- Δείτε τον ετήσιο κατάλογο των Β.Δ.
- http://www.oxfordjournals.org/nar/database/a/

Σημερινή άσκηση

- EMBL-bank
- Ακολουθείστε το link:
- http://www.ebi.ac.uk/embl/
- X03635 : Estrogen receptor alpha, Human
- Αναζητήστε την εγγραφή του Estrogen receptor alpha, στον άνθρωπο, χρησιμοποιώντας το accession number του (X03635).
- Δείτε το Nucleotide Sequence του mRNA σε μορφή ENA και σε μορφή EMBL format.
- Στην μορφή ENA, δείτε την ακολουθία ως FASTA format.

Σημερινή άσκηση

Swissprot

- Από την εγγραφή του προηγούμενου παραδείγματος, ακολουθήστε τον σύνδεσμο (link) προς την Β.Δ. UniprotKB/Swissprot, με κωδικό εγγραφής P03372.
  - Δείτε
    - το όνομα και τα συνώνυμα της ακολουθίας
    - Την ταξινόμιση του οργανισμού. Η ταξινόμιση μπορεί επίσεις να βρεθεί και στην ιστοσελίδα του NCBI taxonomy http://www.ncbi.nlm.nih.gov/
    - Λειτουργίες της πρωτεΐνης (και στο τμήμα των Ontologies)
    - Την ακολουθία σε FASTA format
    - Ακολουθείστε το σύνδεσμο (Hs.208124) προς την Β.Δ. Unigene και από εκεί δείτε το προφίλ γονιδιακής έκφρασης μέσω του link 'EST profile'
    - Από την προηγούμενη ιστοσελίδα του Uniprot, ακολουθείστε το σύνδεσμο P03372 προς την Β.Δ. Intact (στο τμήμα protein-protein interaction databases) για να δείτε πόσες πρωτεϊνικές αλληλεπιδράσεις έχει το estrogen receptor alpha.

Σημερινή άσκηση

PFAM

- Για την ακολουθία του Estrogen receptor alpha, από τη Uniprot ακολουθείστε τη σύνδεση για την Β.Δ. πρωτεϊνικών επικρατειών (domains) Pfam (graphical view).
- Δείτε την αρχιτεκτονική της πρωτεΐνης.
- Ποιά είναι τα βασικά domains;
- Δείτε λεπτομερέστερα την εγγραφή για το Hormone receptor / ligand binding domain.
- Δείτε σε ποιά είδη έχει βρεθεί αυτή η επικράτεια (σύνδεσμος ‘species’ στα αριστερά της ιστοσελίδας) (Tree).

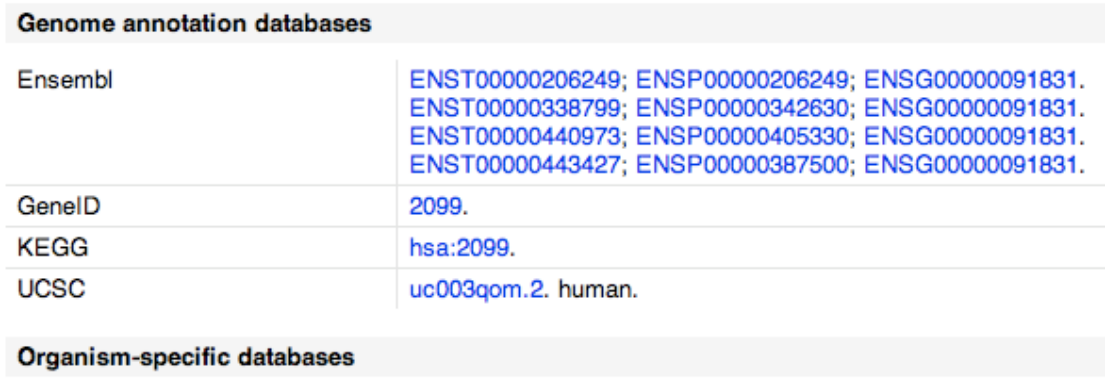
Σημερινή άσκηση

PDB

- Από την προηγούμενη ιστοσελίδα του Uniprot για την εγγραφή estrogen receptor alpha, στο τμήμα 3D structure databases, επιλέξτε RCSB PDB και ακολουθείστε το σύνδεσμο για την 1A52 (είναι ο κωδικός εγγραφής στην PDB). Είναι η κρυσταλλική δομή της επικράτειας σε σύμπλεγμα με την οιστραδιόλη.
- Στην δεξιά πλευρά της ιστοσελίδας μπορείτε να δείτε την τρισδιάστατη δομή μέσω του συνδέσμου ‘view in Jmol’.

KEGG

- Από την ιστοσελίδα του Uniprot για το Estrogen receptor alpha, ακολουθείστε το σύνδεσμο hsa:2099 προς τη Β.Δ. KEGG.



Genome annotation databases	
Ensembl	ENST00000206249; ENSP00000206249; ENSG00000091831. ENST00000338799; ENSP00000342630; ENSG00000091831. ENST00000440973; ENSP00000405330; ENSG00000091831. ENST00000443427; ENSP00000387500; ENSG00000091831.
GeneID	2099.
KEGG	hsa:2099.
UCSC	uc003qom.2. human.
Organism-specific databases	

•Δεξιά της νέας ιστοσελίδας (στο KEGG), ακολουθείστε το σύνδεσμο KEGG disease και στη συνέχεια το σύνδεσμο H00026 για endometrial cancer.

•Στη νέα ιστοσελίδα, στο τμήμα ‘markers’ δείτε ποιά γονίδια χρησιμοποιούνται ως μοριακοί δείκτες της ασθένειας.

•Στα δεξιά της ιστοσελίδας ακολουθείστε το σύνδεσμο KEGG pathways, για να δείτε το μοριακό μονοπάτι του καρκίνου του ενδομητρίου (link: hsa05213).