



Βιοπληροφορική

Ενότητα 4:

Βάσεις Δεδομένων (3/3), 2 ΔΩ

Τμήμα: **Βιοτεχνολογίας**

Όνομα καθηγητή: **Τ. Θηραίου**



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης





Μαθησιακοί Στόχοι

- Παρουσίαση βασικών εκπροσώπων των:
 - δευτερογενών βιολογικών ΒΔ.
 - εξειδικευμένων ΒΔ.
 - βιβλιογραφικών ΒΔ.
- Εξοικείωση με την αναζήτηση στις βιολογικές ΒΔ.



Λέξεις Κλειδιά

- Λέξεις κλειδιά: Δευτερογενείς ΒΔ, Εξειδικευμένες ΒΔ, Βιβλιογραφικές ΒΔ, Ολοκληρωμένα συστήματα ανάκτησης πληροφοριών.
- Key words: Secondary biological databases, Specialized biological databases, Bibliographic databases, Data retrieval systems.



Κατηγορίες ΒΔ

- Αρχειακές / Πρωτογενείς.
 - κατάθεση δεδομένων.
- Δευτερογενείς.
 - επεξεργασία δεδομένων.
- Εξειδικευμένες.
- Βιβλιογραφικές.



Δευτερογενείς ΒΔ 1/3

● ΒΔ Οικογενειών πρωτεϊνών και domains:

- domains: υπομονάδες με καλά συντηρημένη ακολουθία και συγκεκριμένη δομή / λειτουργία.
- ανάλυση των ακολουθιών με διάφορους τρόπους:
 - Patterns (regular expression).
 - Profiles (weighted matrices).
 - Hidden Markov Model (HMM).

```
HPT1_HUMAN : NLTTGATLINEQWLLTAKNA
ACRO_RABIT : YHACGGVLLNAHWVLTAAHCS
KLKE_HUMAN : RFLCGGALLSGQWVITATHCL
MCT3_SHEEP : SYICGGFLVREDFVLTAAHCP
TRB2_HUMAN : MHFCGGS LIHPQWVLTAAHCE
PRTC_HUMAN : KLACGAVLIHPSWVLTAAHCA
EL2_MOUSE : RHNCGGSLVANNWVLTAAHCH
HPT_CANFA : NLTSGATLINEQWLMTAKNV
VSP3_TRIFL : GALCGGT LINQEWVLTASHCL
TMS3_HUMAN : YHLCGGSVITPLWIIITAAHCA
TRY2_RAT : YHFCGGS LINDQWVVSAAHCF
MCT2_RAT : RVICGGELISROFVLTAAHCF
HPT_MUSSA : GLTTGATLISDQWLLTAKNN
TRY4_LUCCU : SHSCGGSVYNSRIIVTAAHCY
PLMN_MACMU : MHFCGGT LISPEWVLTAGHCN
```



Δευτερογενείς ΒΔ 2/3

● ΒΔ Οικογενειών πρωτεϊνών και domains:

– InterPro <http://www.ebi.ac.uk/interpro/index.html>.

● ολοκλήρωση επιμέρους ΒΔ π.χ.

– PROSITE <http://prosite.expasy.org/>.

– PFAM <http://pfam.sanger.ac.uk/>.

● InterProScan: αναζήτηση των πρωτεϊνικών μοτίβων σε μια ακολουθία.

– εύρεση απομακρυσμένων σχέσεων σε ακολουθίες.

– εξαγωγή συμπερασμάτων για την πρωτεϊνική λειτουργία.



Δευτερογενείς ΒΔ 3/3

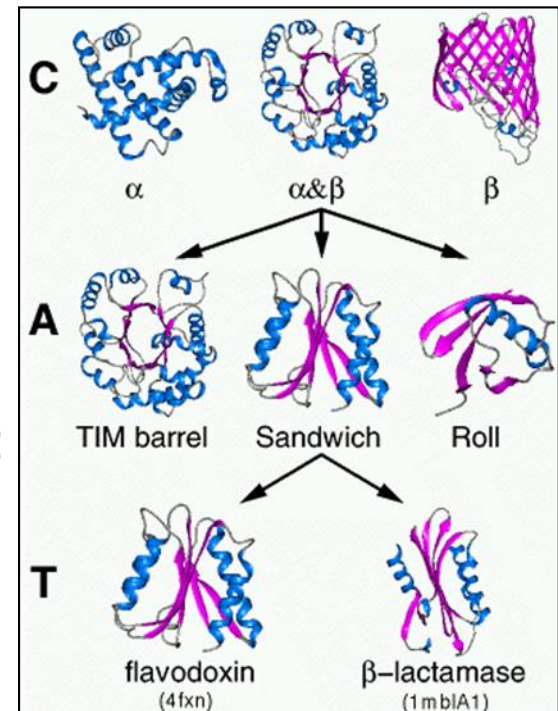
● ΒΔ Ταξινόμησης πρωτεϊνικών δομών:

– CATH <http://www.cathdb.info/>:

- Class.
- Architecture.
- Topology (fold family).
- Homologous superfamily.

– SCOP <http://scop.mrc-lmb.cam.ac.uk/scop/>:

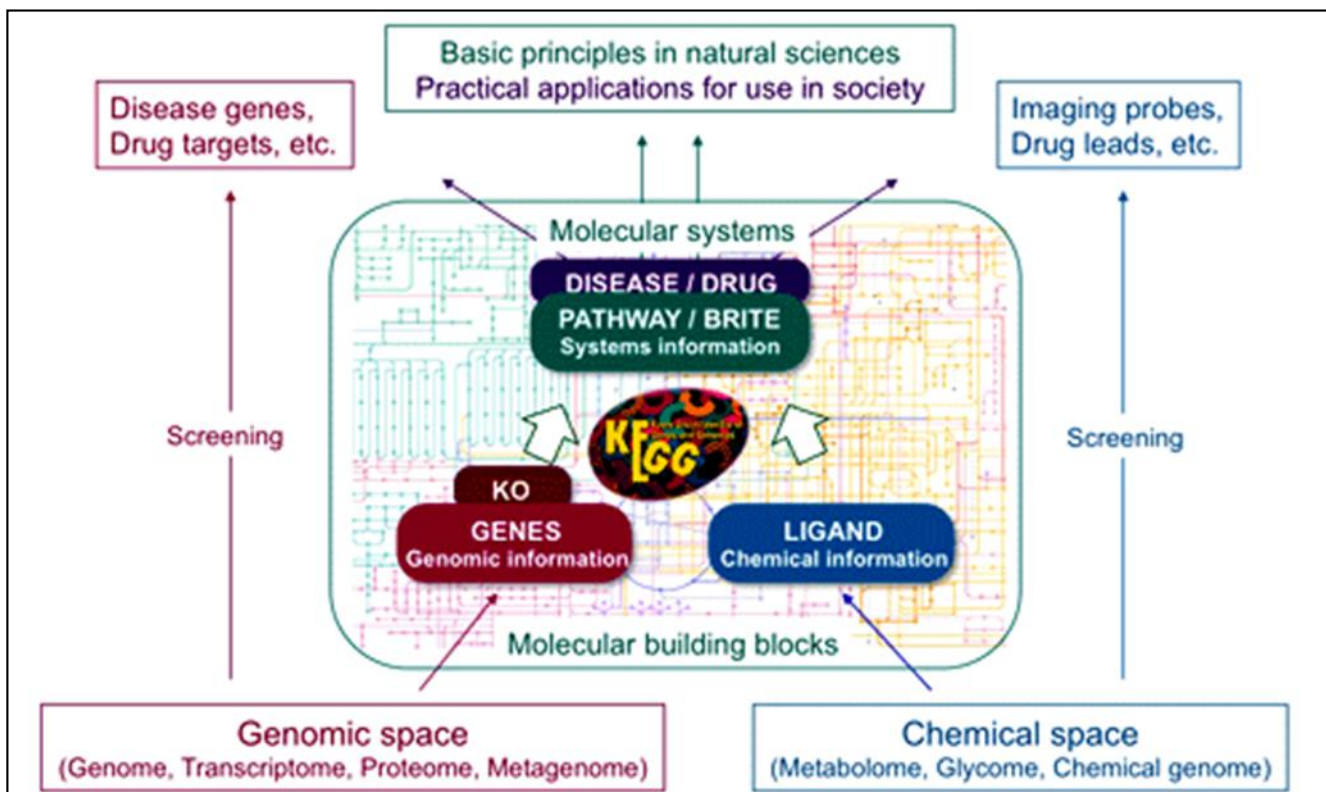
- Class.
- Fold.
- Superfamily.
- Family.
- Protein.





Εξειδικευμένες ΒΔ 1/2

- **KEGG: Kyoto Encyclopedia of Genes and Genomes** <http://www.genome.jp/kegg/>.





Εξειδικευμένες ΒΔ 2/2

- **OMIM: Online Mendelian Inheritance in Man**
<http://www.ncbi.nlm.nih.gov/omim>.
- **Gene Ontology** <http://www.geneontology.org/>.
 - δομημένο ελεγχόμενο λεξιλόγιο (γνωστό ως οντολογία) για την περιγραφή των γονιδιακών προϊόντων σε τρία επίπεδα:
 - cellular component.
 - molecular function.
 - biological process.



Βιβλιογραφικές ΒΔ

- **MEDLINE** (πρόσβαση μέσω της υπηρεσίας **PubMed**) <http://www.ncbi.nlm.nih.gov/pubmed/>.

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed Search

RSS Save search Advanced Help

Show additional filters

Text availability
Abstract available
Free full text available
Full text available

Publication dates
5 years
10 years
Custom range...

Species
Humans
Other Animals

Article types
Clinical Trial
Meta-Analysis
Practice Guideline
Randomized Controlled Trial
Review
Systematic Reviews
more ...

Languages
English
more ...

Clear all

Display Settings: Summary, 20 per page, Sorted by Recently Added

Send to: Filters: [Manage Filters](#)

Results: 1 to 20 of 36819 << First < Prev Page 1 of 1841 Next > Last >>

[Isolation of Basal cells and submucosal gland duct cells from mouse trachea.](#)
1. Hegab AE, Luan Ha V, Attiga YS, Nickerson DW, Gomperts BN.
J Vis Exp. 2012 Sep 14;(67). pii: 3731. doi: 10.3791/3731.
PMID: 23007468 [PubMed - in process]

[Cord-forming mycobacteria induce DNA-meshwork formation by human peripheral blood mononuclear cells.](#)
2. Jónsson BE, Bylund J, Johansson BR, Telemo E, Wold AE.
FEMS Immunol Med Microbiol. 2012 Sep 24. doi: 10.1111/1574-695x.12007. [Epub ahead of print]
PMID: 23005575 [PubMed - as supplied by publisher]

[Reduction of lipid-laden macrophage index after laparoscopic Nissen fundoplication in cystic fibrosis patients after lung transplantation.](#)
3. Hayes D Jr, Kirkby S, McCoy K, Mansour HM, Khosravi M, Strawbridge H, Tobias JD.
Clin Transplant. 2012 Sep 24. doi: 10.1111/ctr.12020. [Epub ahead of print]
PMID: 23004684 [PubMed - as supplied by publisher]

[VD testing for CF patients: testing patients with cystic fibrosis for respiratory infections is crucial during flu season, and the use of agar-based systems is recommended.](#)
4. Porter G.
MLO Med Lab Obs. 2012 Sep;44(9):40-2. No abstract available.
PMID: 23002511 [PubMed - in process]

[Abstracts of the 26th Annual North American cystic Fibrosis Conference, October 11-13, 2012, Orlando, Florida, USA.](#)
5. [No authors listed]

Results by year

Related searches

- cystic fibrosis review
- cystic fibrosis treatment
- aeruginosa cystic fibrosis
- cystic fibrosis lung
- cystic fibrosis pseudomonas

Titles with your search terms

- A CFTR potentiator in patients with **cystic fibrosis** and the G551D rr [N Engl J Med. 2011]
- Genotype and phenotype in **cystic fibrosis**. [Respiration. 2000]
- Effect of VX-770 in persons with **cystic fibrosis** and the G551D-CFTR mu [N Engl J Med. 2010]

See more...

5602 free full-text articles in



Αναζήτηση στις ΒΔ 1/4

● Διαδραστική αναζήτηση:

- κωδικός εγγραφής στη ΒΔ.
- λέξεις κλειδιά.
- νουκλεοτιδικές ή αμινοξικές ακολουθίες (BLAST κ.α.)

Advanced Search Interface

All/Experimental Type/Molecule Type ?

Retrieve all PDB entries or a subset based on experimental method or molecule type

Experimental Method: X-ray Result Count

Molecule Type: Protein

AND

UniProtKB Accession Number(s) ?

Search for structures by entering one or more [UniProtKB](#) Accession Numbers (UniProtKB AC), e.g. P69905

Accession IDs: P01308 Result Count

[Add Search Criteria](#)

Remove Similar Sequences at 90% Identity

Match [all](#) of the above conditions. Clear All Parameters Submit Query



Αναζήτηση στις ΒΔ 2/4

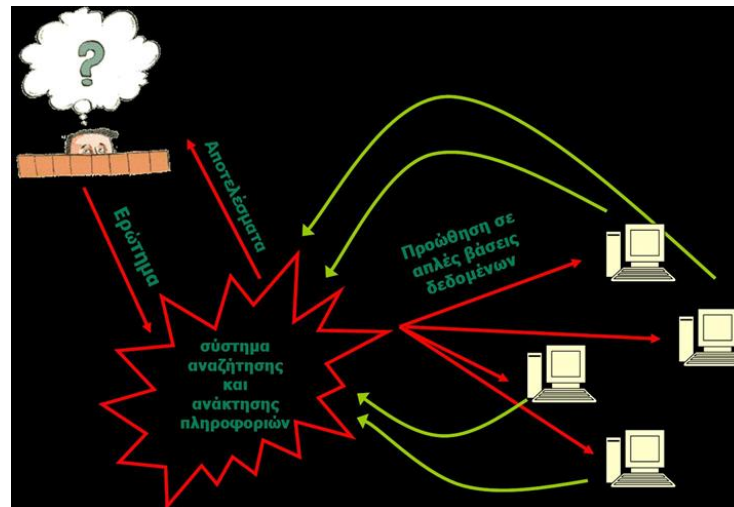
● Πρόσβαση μέσω προγραμμάτων:

- σύνθετες αναζητήσεις / μαζική ανάκτηση δεδομένων.
- δημιουργία συστημάτων επιστημονικών ροών εργασίας (scientific workflows) και διαδικτυακών εφαρμογών.
 - π.χ. in silico protein engineering web server.
- Entrez Programming Utilities (eUtils)
 - <http://www.ncbi.nlm.nih.gov/books/NBK25497/>.



Αναζήτηση στις ΒΔ 3/4

- Ολοκληρωμένα συστήματα ανάκτησης πληροφοριών:
 - ταυτόχρονη αναζήτηση σε πολλές ΒΔ.
 - απλές ή σύνθετες αναζητήσεις.
 - εκκίνηση εφαρμογών επί των αποτελεσμάτων.
 - ενσωμάτωση ΒΔ που έχουν δημιουργηθεί από το χρήστη.





Αναζήτηση στις ΒΔ 4/4

● Ολοκληρωμένα συστήματα ανάκτησης πληροφοριών:

- Entrez <http://www.ncbi.nlm.nih.gov/sites/gquery>.
- SRS <http://srs.ebi.ac.uk/>.
- EB-eye <http://www.ebi.ac.uk/ebisearch>.

The screenshot displays the EBI search interface. At the top, there is a search bar with 'tpi1' entered and a 'Go' button. Below the search bar, there are navigation tabs for 'Databases', 'Tools', 'EBI Groups', 'Training', 'Industry', 'About Us', and 'Help'. The main content area is divided into several sections:

- All results:** A list of categories including Gene & Protein Summaries, Genomes, Nucleotide Sequences, Protein Sequences, Gene Expression, Molecular Interactions, Protein Families, Literature, and EBI Web Site.
- Gene & Protein Summaries (includes expression, structures, literature...):** This section shows two entries: 'Triosephosphate isomerase 1' (TPH1) from *Homo sapiens* and 'Triosephosphate isomerase 1 Gene' (Tpi1) from *Mus musculus*. There is a 'View all organisms in this group...' link.
- Genomes / HGNC:** This section shows the HGNC entry for TPH1, including its approved symbol, name, status, locus type, chromosome, and references to UniProtKB, Intenz, and Ensembl Gene.
- Genomes / Ensembl Gene:** This section shows the Ensembl entry for triosephosphate isomerase 1 Gene in *Mus musculus*, including its source, accession number, and references to Taxonomy, Ensembl Genomes Gene, UniProtKB, Ensembl, GO, and EMBL-Bank.



Προοπτικές

- **Χειρισμός σφαλμάτων**
 - σφάλματα στα δεδομένα ή/και στο σχολιασμό.
- **Αποδοτική αναζήτηση**
 - στοχευμένα αποτελέσματα.
 - ταχύτητα αναζήτησης.
- **Ολοκλήρωση δεδομένων (data integration)**
 - ΒΔ πολυάριθμες, αυτόνομες, ετερογενείς και μερικώς επικαλυπτόμενες.
 - προτυποποίηση.



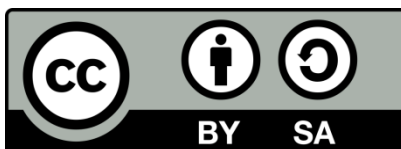
Βιβλιογραφία

- David Mount, "Bioinformatics: Sequence and Genome Analysis", Cold Spring Harbor Laboratory Press; 2nd edition (March 12, 2013).
- Jonathan Pevsner, "Bioinformatics and Functional Genomics", Wiley-Blackwell; 2nd edition (May 4, 2009).
- Andreas D. Baxevanis, B. F. Francis Ouellette, "Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins", Wiley-Interscience; 3rd edition (October 29, 2004).



Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδεια χρήσης, η άδεια χρήσης αναφέρεται ρητώς.





Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα Γεωπονικού Πανεπιστημίου Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



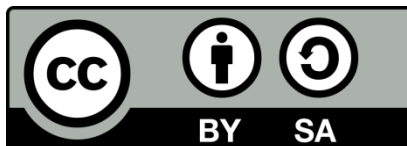
Σημείωμα Αναφοράς

Copyright Γεωπονικό Πανεπιστήμιο Αθηνών 2015. Τμήμα Βιοτεχνολογίας, Θηραίου Τριάς. «Βιοπληροφορική». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση:
<https://mediasrv.aua.gr/eclass/courses/OCDB100/>



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων, π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



Η άδεια αυτή ανήκει στις άδειες που ακολουθούν τις προδιαγραφές του Ορισμού Ανοικτής Γνώσης [2], είναι ανοικτό πολιτιστικό έργο [3] και για το λόγο αυτό αποτελεί ανοικτό περιεχόμενο [4].

[1] <http://creativecommons.org/licenses/by-sa/4.0/>

[2] <http://opendefinition.org/okd/ellinika/>

[3] <http://freedomdefined.org/Definition/EI>

[4] <http://opendefinition.org/buttons/>



Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
 - το Σημείωμα Αδειοδότησης
 - τη δήλωση Διατήρησης Σημειωμάτων
 - το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)
- μαζί με τους συνοδευόμενους υπερσυνδέσμους.