Βιοπληροφορική

Ενότητα 3: Ηλεκτρονική διαχείριση βιολογικών δεδομένων

Αν. καθηγητής Αγγελίδης Παντελής
e-mail: paggelidis@uowm.gr
ΕΕΔΙΠ Μπέλλου Σοφία
e-mail: sbellou@uowm.gr

Τμήμα Μηχανικών Πληροφορικής και Τηλεπικοινωνιών

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ
2007-2013
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Άδειες Χρήσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ψηφιακά Μαθήματα στο Πανεπιστήμιο Δυτικής Μακεδονίας**» έχει χρηματοδοτήσει μόνο τη 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

ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σκοπός του μαθήματος

- Δημόσιες βάσεις δεδομένων βιοπληροφορικής.
- Βασικοί ορισμοί (Data repository, Data mart, Data warehouse).
- Ηλεκτρονικά ιατρικά αρχεία.
- Κύκλος ζωής ιατρικών δεδομένων.
- Είδη βάσεων δεδομένων στη Βιοπληροφορική.
- Οργάνωση των βιολογικών βάσεων δεδομένων (πεδία-εγγραφές).
- Σχεσιακές βιολογικές βάσεις δεδομένων.
- Βιβλιογραφικές βάσεις δεδομένων.

Computers and bioinformatics

Computers serve 4 independent functions in bioinformatics:

I. **Communication:** More journal articles available for the researchers in less time and increasingly large proportion of academic research information appears online.

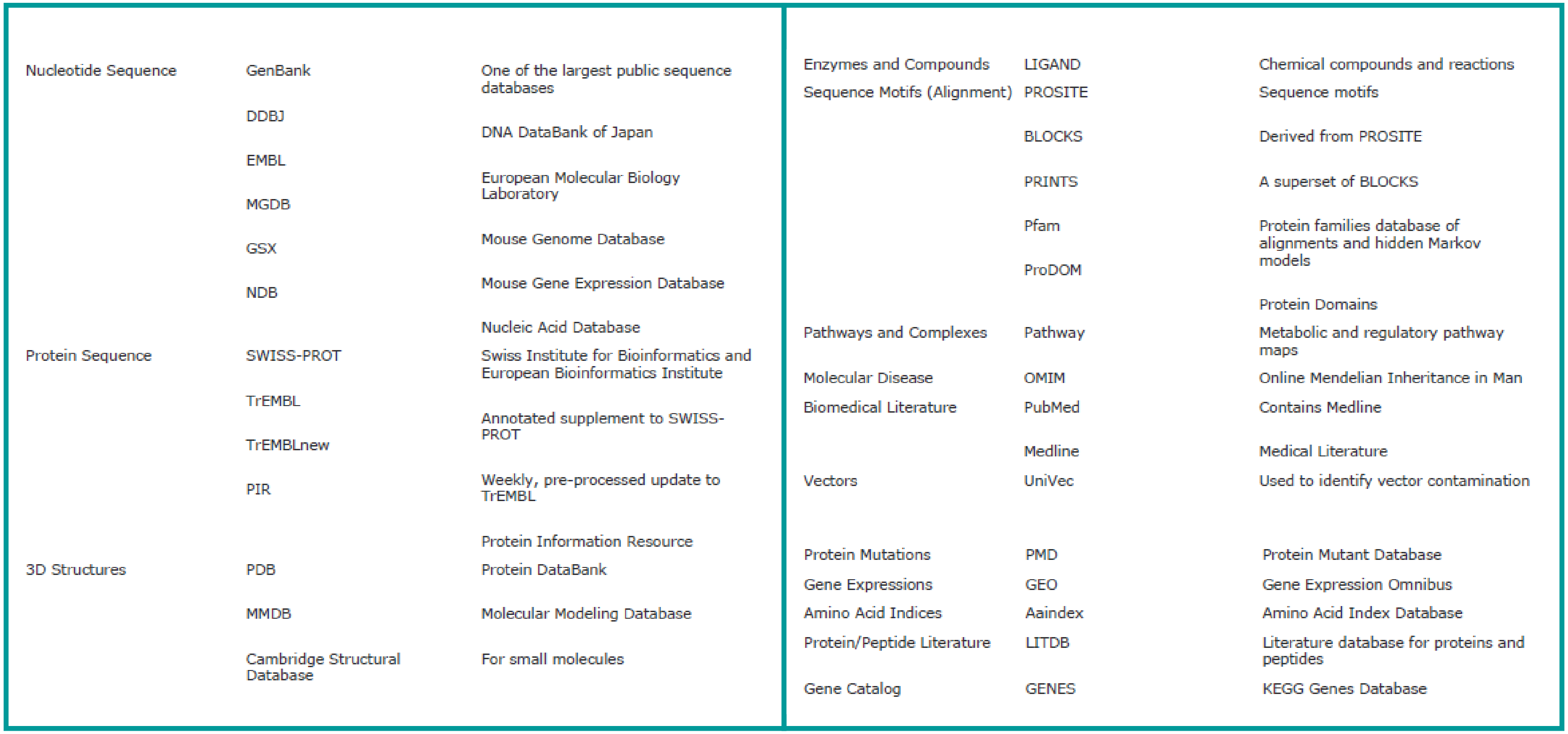
II. **Computations:** Computers are used for tasks that range from searching a reference (or a nucleotide sequence) and visualizing protein folding patterns to simulating 3D protein-protein interaction.

III. **Control:** Computer controlled devices are superior to manual operators, when time matters.

IV. **Storage:** A true database is the data repository, a database used as an information storage facility, with minimal analysis or querying functionality.

https: //www.cartoonstock.com/

Public bioinformatics databases accessible via the Internet

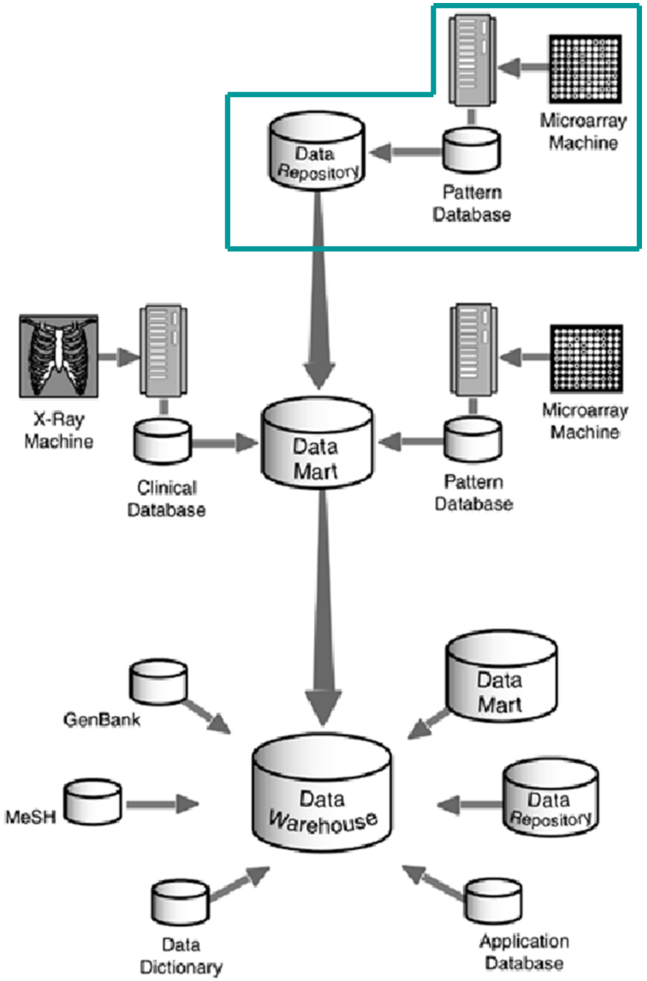


Nucleotide Sequence	GenBank	One of the largest public sequence databases
	DDBJ	DNA DataBank of Japan
	EMBL	European Molecular Biology Laboratory
	MGDB	Mouse Genome Database
	GSX	Mouse Gene Expression Database
	NDB	Nucleic Acid Database
Protein Sequence	SWISS-PROT	Swiss Institute for Bioinformatics and European Bioinformatics Institute
	TrEMBL	Annotated supplement to SWISS-PROT
	TrEMBLnew	Weekly, pre-processed update to TrEMBL
	PIR	Protein Information Resource
3D Structures	PDB	Protein DataBank
	MMDB	Molecular Modeling Database
	Cambridge Structural Database	For small molecules

Enzymes and Compounds	LIGAND	Chemical compounds and reactions
Sequence Motifs (Alignment)	PROSITE	Sequence motifs
	BLOCKS	Derived from PROSITE
	PRINTS	A superset of BLOCKS
	Pfam	Protein families database of alignments and hidden Markov models
	ProDOM	Protein Domains
Pathways and Complexes	Pathway	Metabolic and regulatory pathway maps
Molecular Disease	OMIM	Online Mendelian Inheritance in Man
Biomedical Literature	PubMed	Contains Medline
	Medline	Medical Literature
Vectors	UniVec	Used to identify vector contamination
Protein Mutations	PMD	Protein Mutant Database
Gene Expressions	GEO	Gene Expression Omnibus
Amino Acid Indices	Aaindex	Amino Acid Index Database
Protein/Peptide Literature	LITDB	Literature database for proteins and peptides
Gene Catalog	GENES	KEGG Genes Database

Bergeron, B. *Bioinformatics Computing*

Ορισμοί – Definitions (1/3)

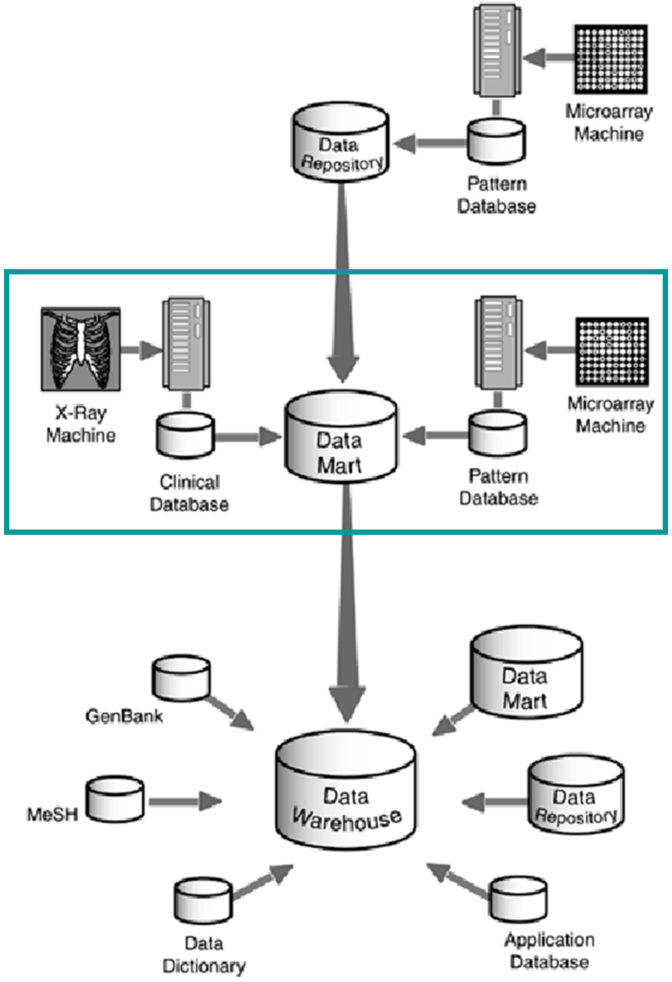


1st level: Data repository

- Simplest level of a true database is the data repository (database used as an information storage facility) with minimal analysis.
- Advantages using data repository instead of original database.
  - Longitudinal studies are possible because all data in the host application are stored in the repository.
  - Offloads the query functions that are available through native applications to the database management system that enables efficient control and management of the data repository.

Bergeron, B. *Bioinformatics Computing*

Ορισμοί – Definitions (2/3)

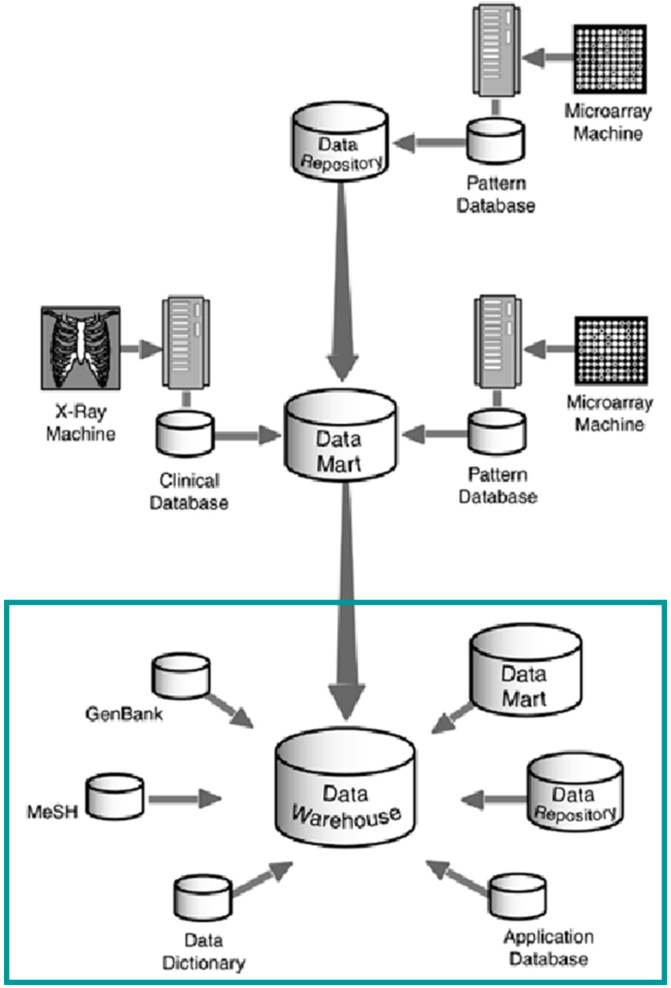


2nd level: Data mart

- Εξειδικευμένα θεματικά υποσύνολα.
- Περιορισμένο εύρος δεδομένων για ένα συγκεκριμένο ερευνητικό θέμα.
- Διαφορά με data repository: Περιέχει πειραματικά/κλινικά δεδομένα που προέρχονται από πολλές βάσεις δεδομένων - data repository.

Bergeron, B. Bioinformatics Computing

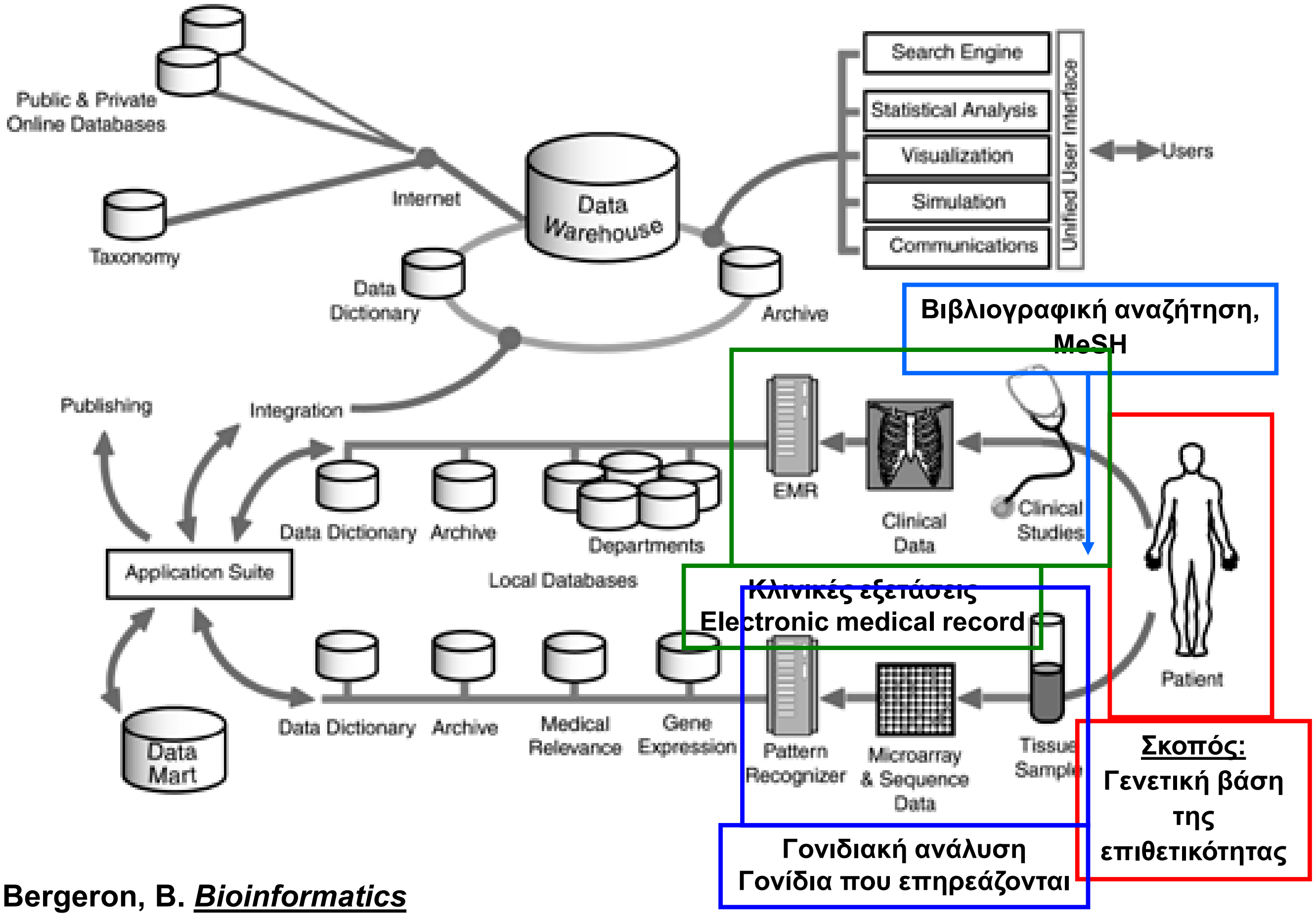
Ορισμοί – Definitions (3/3)



3rd level: Data warehouse

- Κεντρική βάση δεδομένων.
- Σύνολο όψεων από διαφορετικές πηγές.
- Τα δεδομένα:
  - χαρακτηρίζονται από ανομοιομορφία,
  - δεν αντικαθίστανται από καινούργια δεδομένα,
  - χαρακτηρίζονται από την ημερομηνία καταχώρησης (time-stamped) η οποία δεν αλλάζει.

Data management scenario for a pharmacogenomic laboratory (1/3)



Βιβλιογραφική αναζήτηση, MeSH

Κλινικές εξετάσεις
Electronic medical record

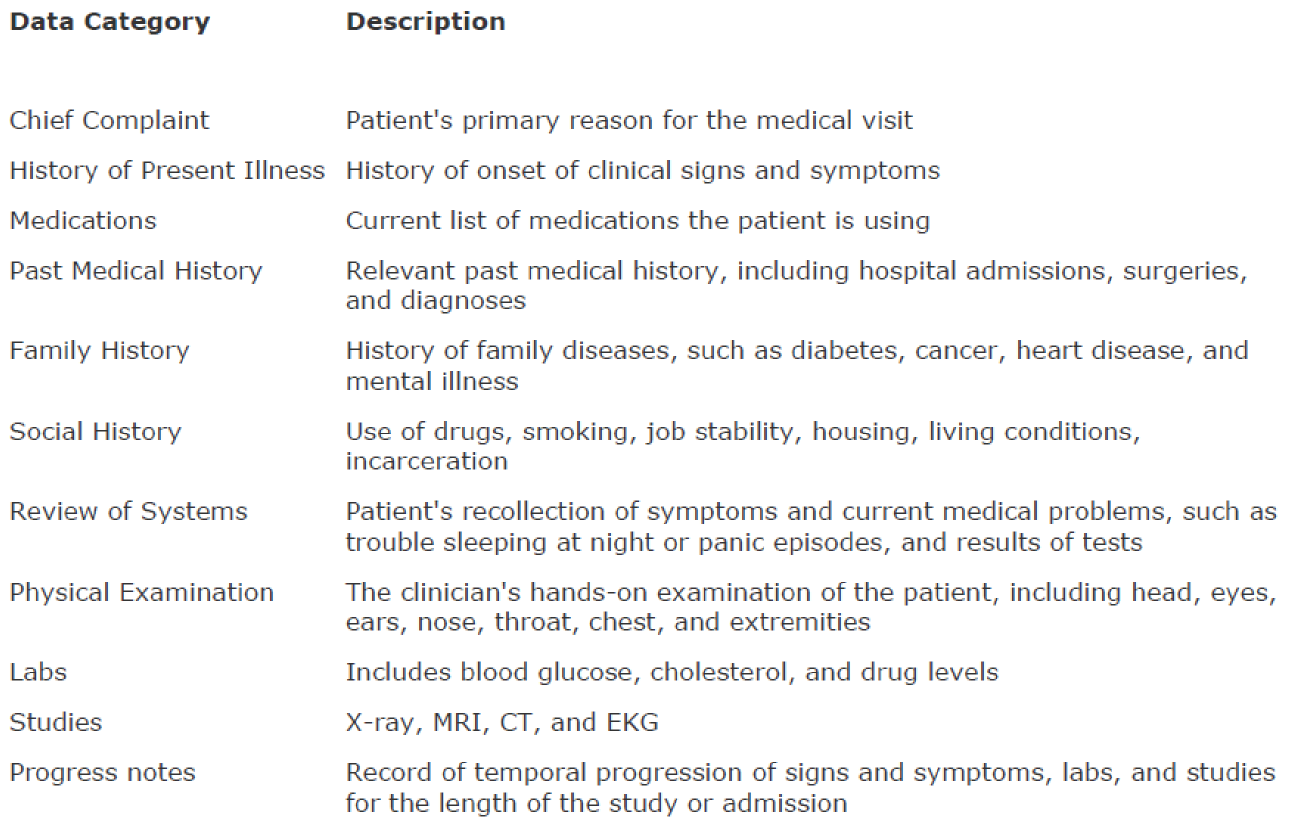
Γονιδιακή ανάλυση
Γονίδια που επηρεάζονται

<u>Σκοπός:</u>
Γενετική βάση της επιθετικότητας

Bergeron, B. *Bioinformatics Computing*

Data management scenario for a pharmacogenomic laboratory (2/3)

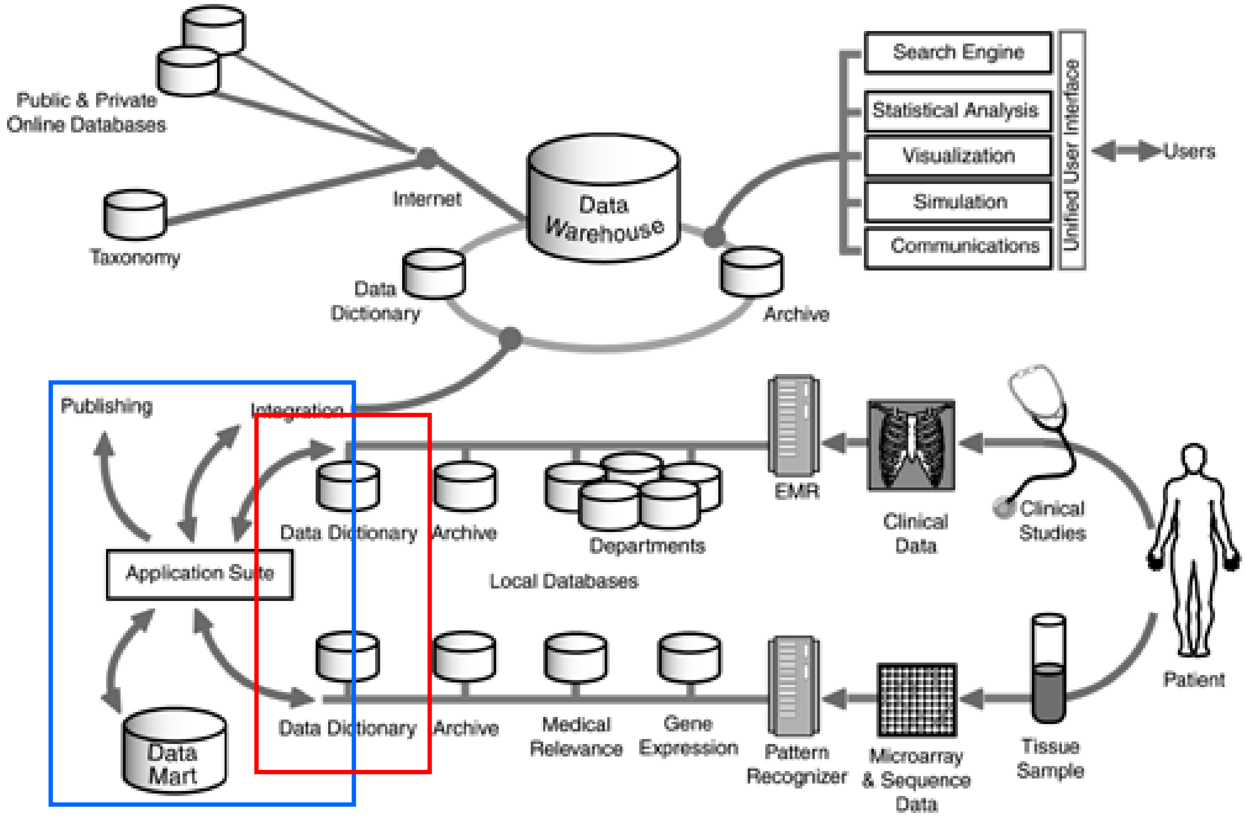
Typical electronic Medical Record (EMR)



Data Category	Description
Chief Complaint	Patient's primary reason for the medical visit
History of Present Illness	History of onset of clinical signs and symptoms
Medications	Current list of medications the patient is using
Past Medical History	Relevant past medical history, including hospital admissions, surgeries, and diagnoses
Family History	History of family diseases, such as diabetes, cancer, heart disease, and mental illness
Social History	Use of drugs, smoking, job stability, housing, living conditions, incarceration
Review of Systems	Patient's recollection of symptoms and current medical problems, such as trouble sleeping at night or panic episodes, and results of tests
Physical Examination	The clinician's hands-on examination of the patient, including head, eyes, ears, nose, throat, chest, and extremities
Labs	Includes blood glucose, cholesterol, and drug levels
Studies	X-ray, MRI, CT, and EKG
Progress notes	Record of temporal progression of signs and symptoms, labs, and studies for the length of the study or admission

Bergeron, B. *Bioinformatics Computing*

Data management scenario for a pharmacogenomic laboratory (3/3)

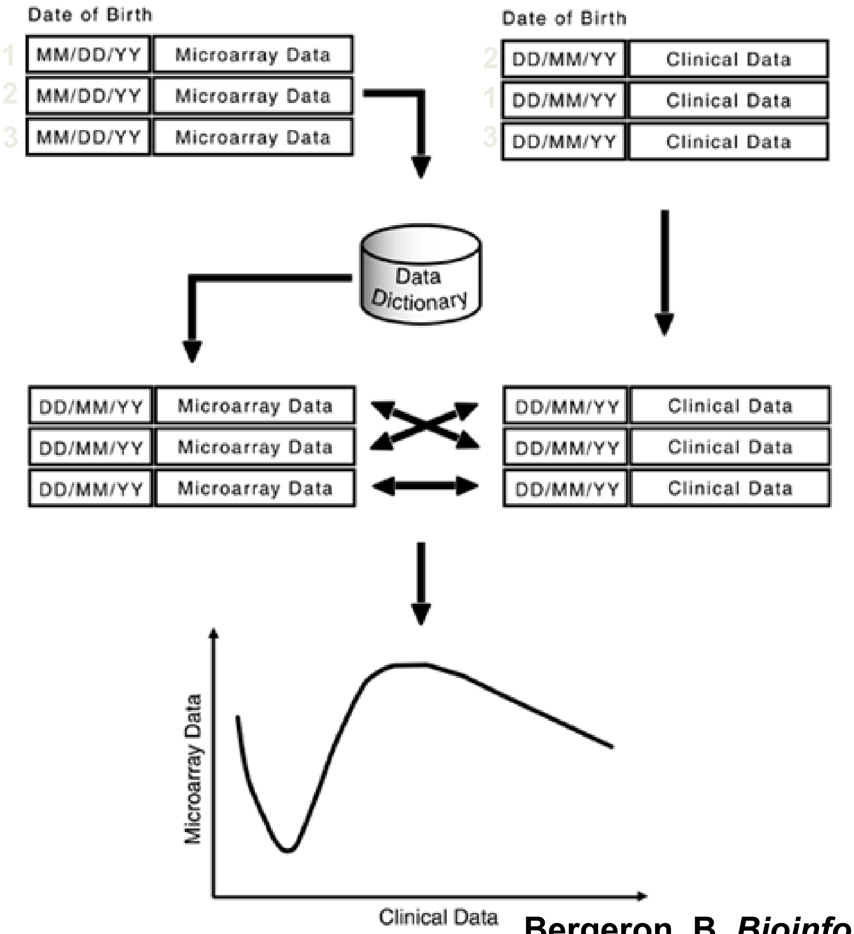


Bergeron, B. *Bioinformatics Computing*

Data modification

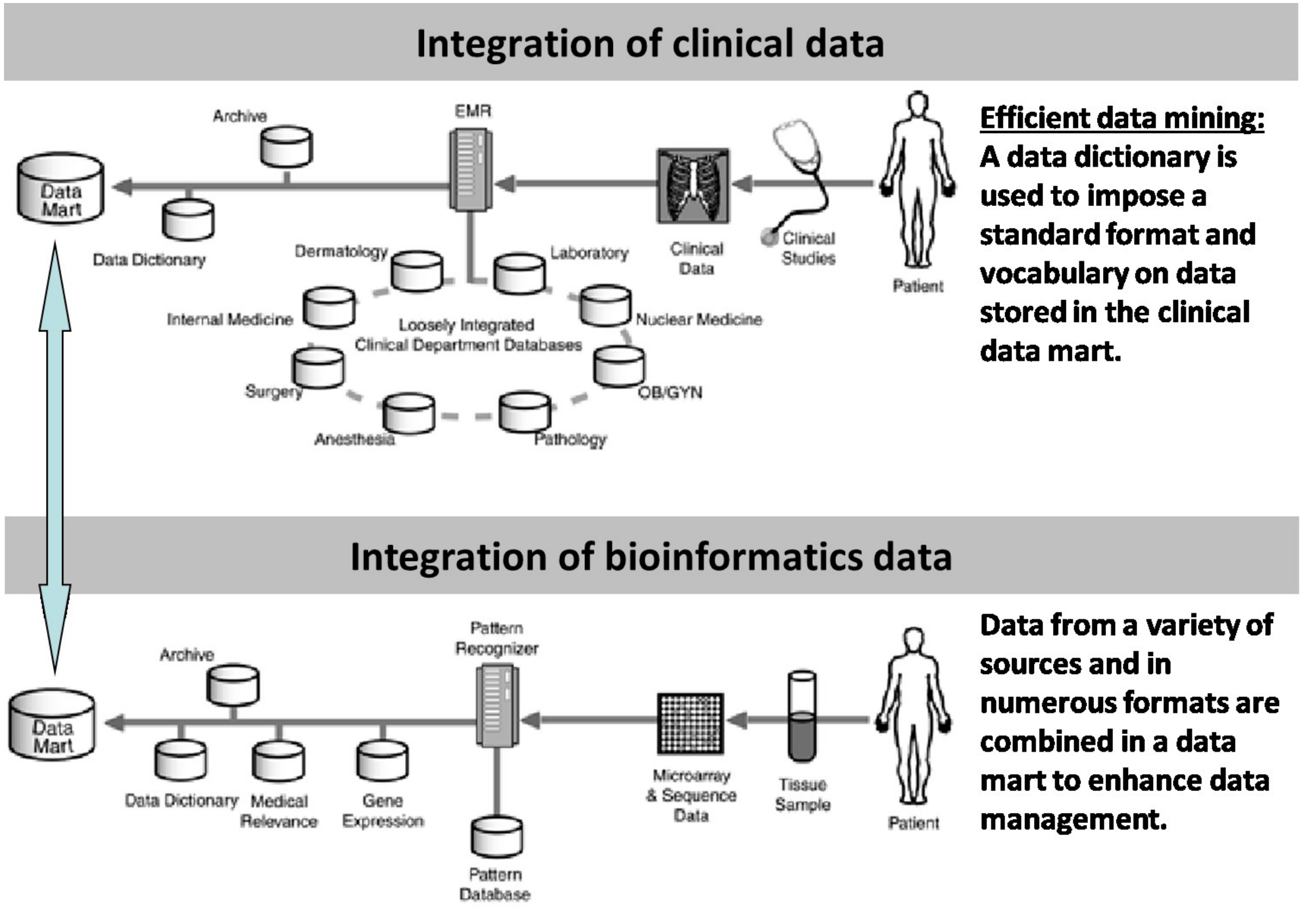
Γονιδιακή ανάλυση δειγμάτων με βιοπληροφορική

Κλινικές δοκιμές σε ασθενείς



Bergeron, B. *Bioinformatics Computing*

Integration of clinical data – Integration of bioinformatics data



Integration of clinical data

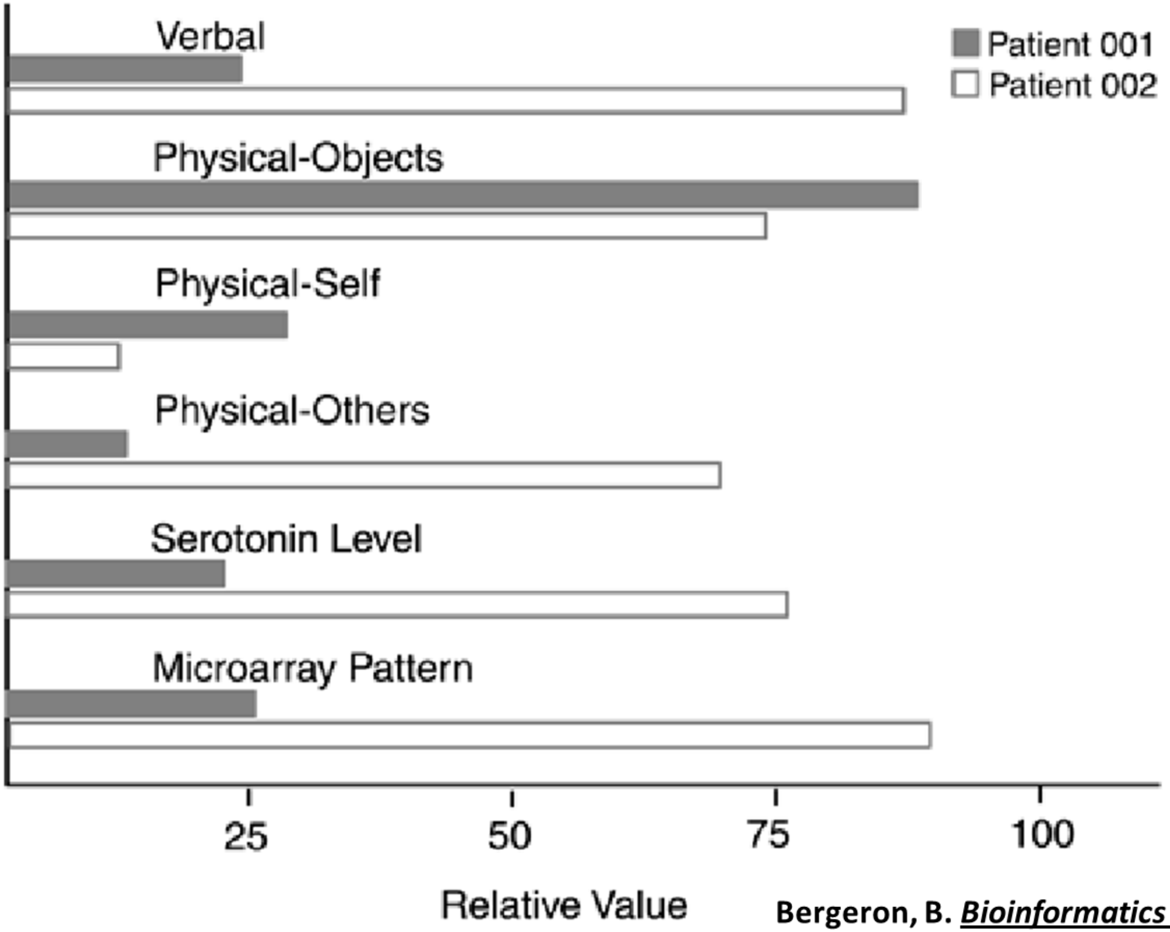
Efficient data mining: **A data dictionary is used to impose a standard format and vocabulary on data stored in the clinical data mart.**

Integration of bioinformatics data

Data from a variety of sources and in numerous formats are combined in a data mart to enhance data management.

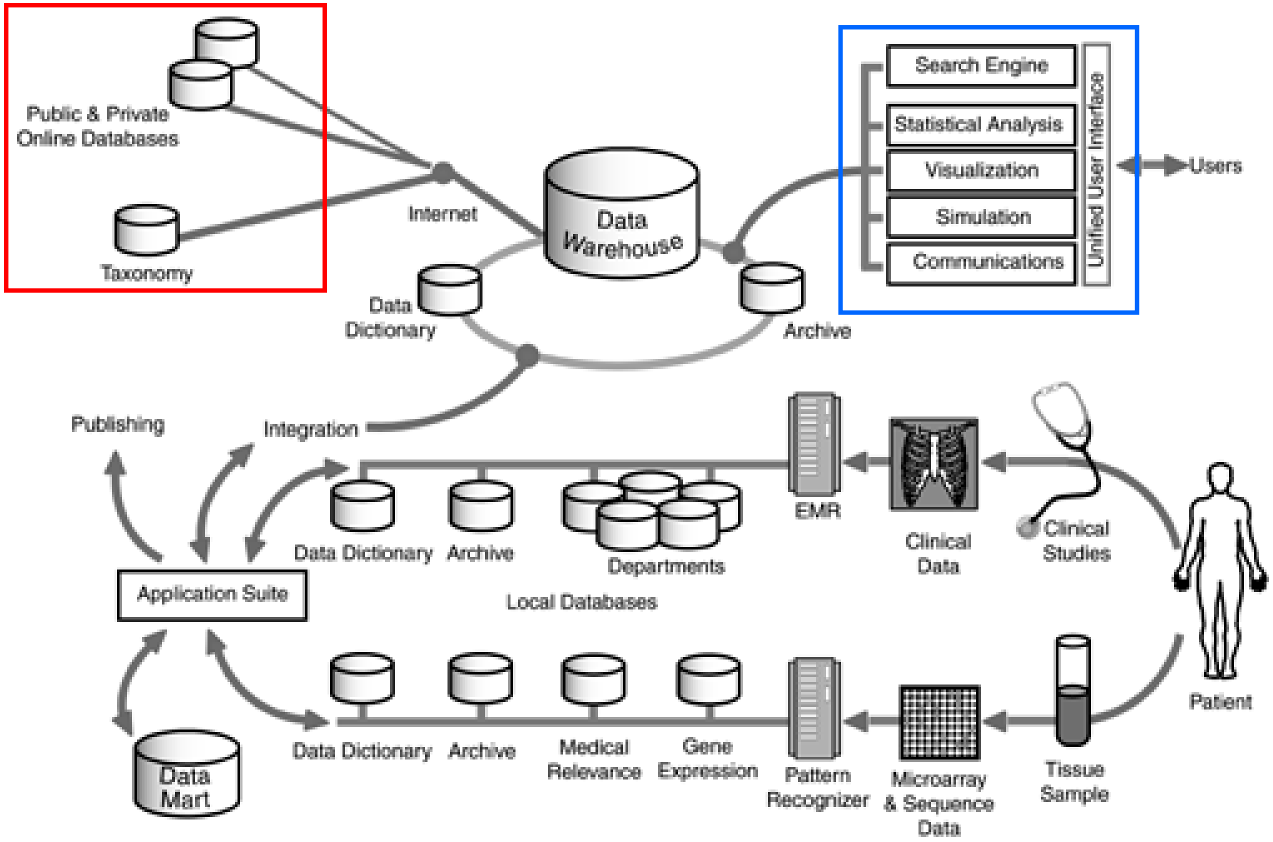
Bergeron, B. *Bioinformatics Computing*

Aggressive behavior - Results



Bergeron, B. ***Bioinformatics Computing***

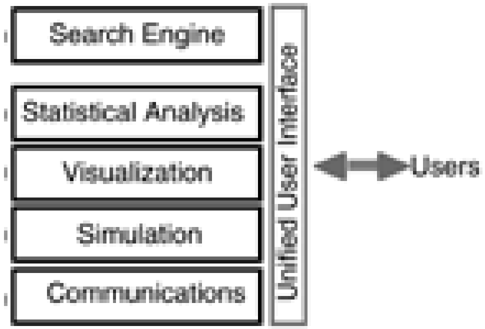
Data management scenario for a pharmacogenomic laboratory



Bergeron, B. ***Bioinformatics Computing***

Genomic & public-funded applications

Application	Examples
Sequence Search	BLAST, BLASTN, CLUSTALW, FASTA, MOTIF, PBLAST, TBLASTIN
Submission	AceDB, Audit, BankIt, Sakura, Sequin, WebIN
Information Retrieval	Entrez, DBGET, IDEAS
Linkage	LocusLink
Portal	KEGG
Structure Match	CD, DALI, SCOP, Searchlite, Structure Explorer, VAST
Visualization	CAD, Cn3D, Mage, RasMol/WebMol, SWISS-PDBViewer, VRML, WebMol
Protein-Protein Interactions	BRITE
Microarray Gene Expression Profiles	Expression
Open-Reading Frame Locator	ORF Finder



Bergeron, B. *Bioinformatics Computing*

FASTA Format
for sequencing analysis

- Διάταξη για αλληλουχίες DNA και πρωτεϊνών.
- Περιέχει γραμμή-επικεφαλίδα, η οποία **αρχίζει πάντα με το σύμβολο <.**
- Δίνει το όνομα ακολουθίας και μία σύντομη περιγραφή γι’ αυτήν.
- Ακολουθείται από τη νουκλεοτιδική (DNA-RNA) ή αμινοξική (πρωτεΐνη) ακολουθία.

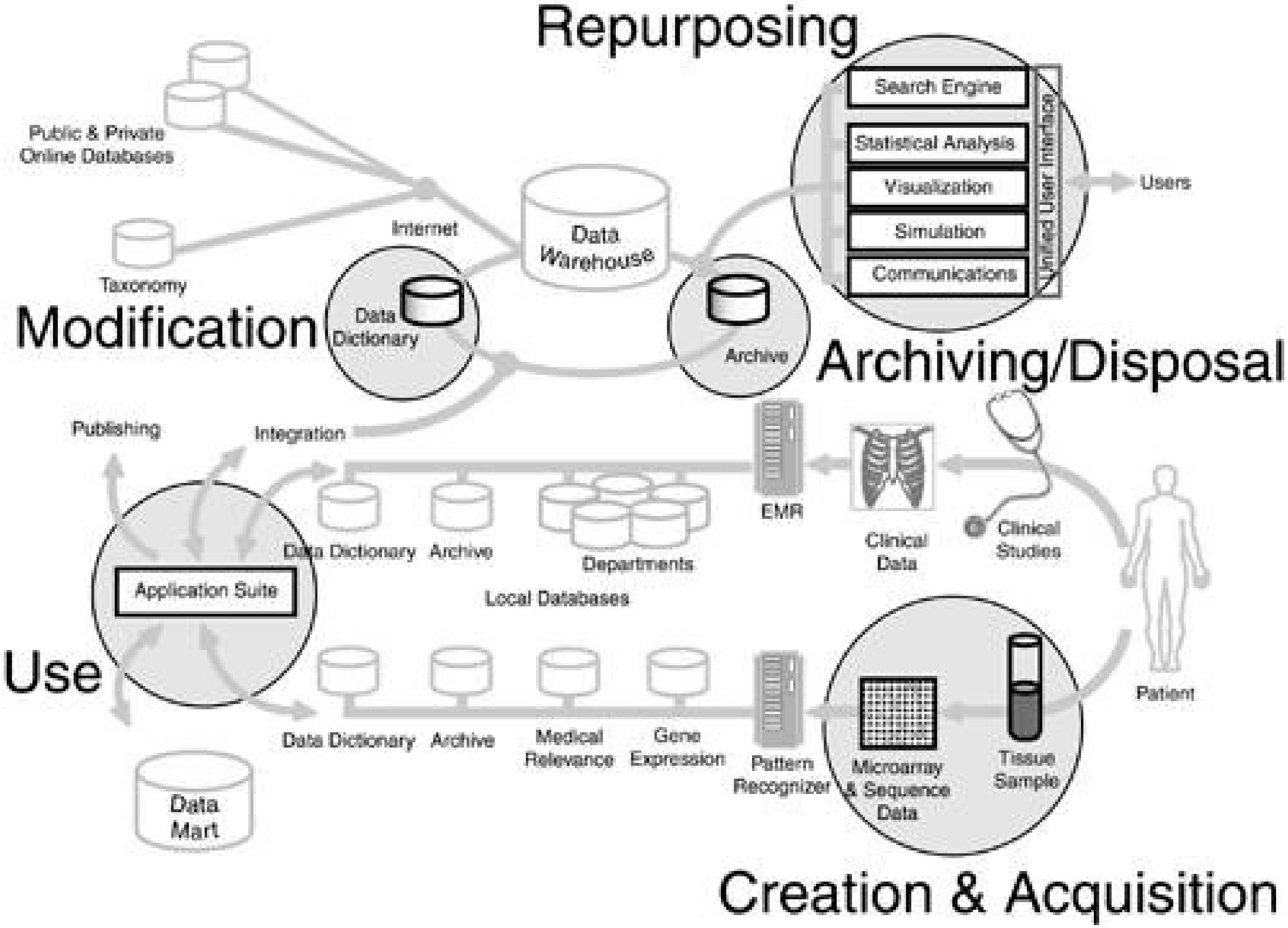
```
1   mvmevgtlda gglrallger aaqcllldcr sffafnaghi agsvnvrfst ivrrrakgam
61  glehivpnae lrgrllagay havvllders aaldgakrdg tlalaagalc rearaaqvff
121 lkggyeafsa scpelcskqs tpmglslpls tsvpdsaesg csscstplyd qggpveilpf
181 lylgsayhas rkdmldalgi talinvsanc pnhfeghyqy ksipvednhk adisswfnea
241 idfidsikna ggrvfvhcqa gisrsaticl aylmrtnrvk ldeafefvkq rrsiispnfs
301 fmgqllqfes qvlaphcsae agspamavld rgtstttvfn fpvsipvhst nsalsylqsp
361 ittspsc
```

```
>gi|184
MVME                                                     EHIVPNAE
LRGRL                                                    PELCSKQS
TPMGL                                                    NVSANC
PNHFE                                                    RTNRVK
LDEAFEFVKQRRSIISPNFSFMGQLLQFESQVLAPHCSAEAGSPAMAVLDRGTSTTTVFNFPVSIPVHST
NSALSYLQSPITTSPSC
```

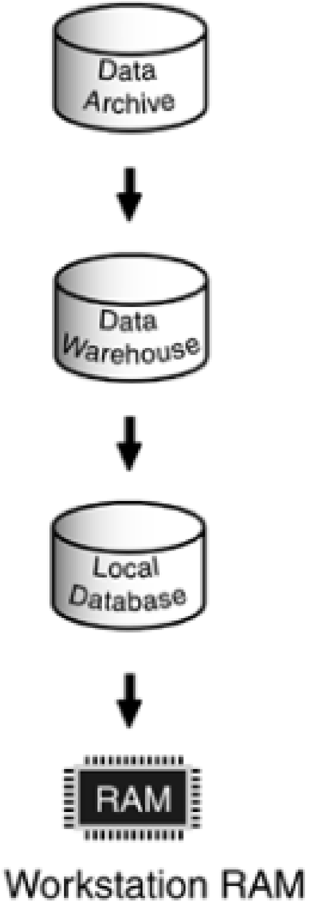
Καταχωρήσεις (entries) στις βιολογικές βάσεις δεδομένων

- Οι καταχωρήσεις έχουν έναν «κύκλο ζωής» στις βάσεις δεδομένων.
- Εξαιτίας της αυξημένης ζήτησης για άμεση πρόσβαση στην νέα πληροφορία, οι καινούργιες καταχωρήσεις είναι διαθέσιμες πριν ολοκληρωθούν και ελεγχθούν από την ερευνητική κοινότητα.
- Οι καταχωρήσεις «ωριμάζουν» μέσω των τάξεων:
  - **Unannotated → Preliminary → Unreviewed → Standard.**
  - **Ασχολίαστη → Προκαταρτική → Μη επαληθεύσιμη → Τελική.**
  - Σπάνια, μία καταχώρηση «πεθαίνει» (an entry '**dies**‘).
  - Περιορισμένος αριθμός καταχωρήσεων έχει αφαιρεθεί από τις βάσεις δεδομένων όταν διαπιστώθηκε ότι ήταν λανθασμένες.

Data Life Cycle



Organic analog of database hierarchy



Αναγκαιότητα…

- Αύξηση βιολογικής πληροφορίας εξαιτίας νέων τεχνολογιών.
  - Για παράδειγμα, αύξηση του αριθμού των γονιδιωμάτων των οργανισμών που έχουν αποκρυπτογραφηθεί.
- Διαθεσιμότητα πληροφορίας στους επιστήμονες.
  - Ενοποίηση πληροφοριών από διαφορετικές πηγές.
  - Προσβασιμότητα σε πληροφορία που δεν μπορεί να δημοσιευτεί αναλυτικά.
- Διαθεσιμότητα πληροφορίας σε επεξεργάσιμη μορφή.
  - Αυτόματη ανάλυση.

Έτσι…

Δημιουργήθηκε η ανάγκη της μεθοδικής οργάνωσής τους σε βάσεις δεδομένων, δηλαδή της ηλεκτρονικής αρχειοθέτησης τους.

Διαφορετικά είδη βάσεων δεδομένων στη Βιοπληροφορική (1/2)

1. **Δεδομένα:**

- **Καταχώρηση στοιχείων και έλεγχος ποιότητας:**
  - ✓ data deposited directly
  - ✓ curators add and update data
  - ✓ treatment of erroneous data: removed, or marked
  - ✓ error checking
  - ✓ consistency, updates
  - ✓ ….

- **Τύπος δεδομένων:**
  - nucleotide sequences (DNA-RNA)
  - protein sequences
  - 3D structures
  - gene expression data
  - metabolic pathways
  - ….

- **Πρωτογενή ή επεξεργασμένα δεδομένα:**
  - Primary databases: direct experimental results
  - Secondary databases: result of analysis on primary databases
  - Consolidation of many databases
  - …

Διαφορετικά είδη βάσεων δεδομένων στη Βιοπληροφορική (2/2)

2. Βάση δεδομένων:

Οργάνωση:

- ✓ flat files.
- ✓ Relational databases.
- ✓ Object-oriented databases.
- ✓ ….

- **Διαθεσιμότητα:**
  - Publicly available, no restriction.
  - Available, but with copyright.
  - Accessible, but not downloadable.
  - Academic, but not freely available.
  - Commercial.

- **Curators:**
  - Large, public institution (EMBL, NCBI).
  - Quasi-academic institute (Swiss institute of Bioinformatics, TIGR,…).
  - Academic group or scientist.
  - Commercial company.

Οργάνωση των βιολογικών βάσεων δεδομένων

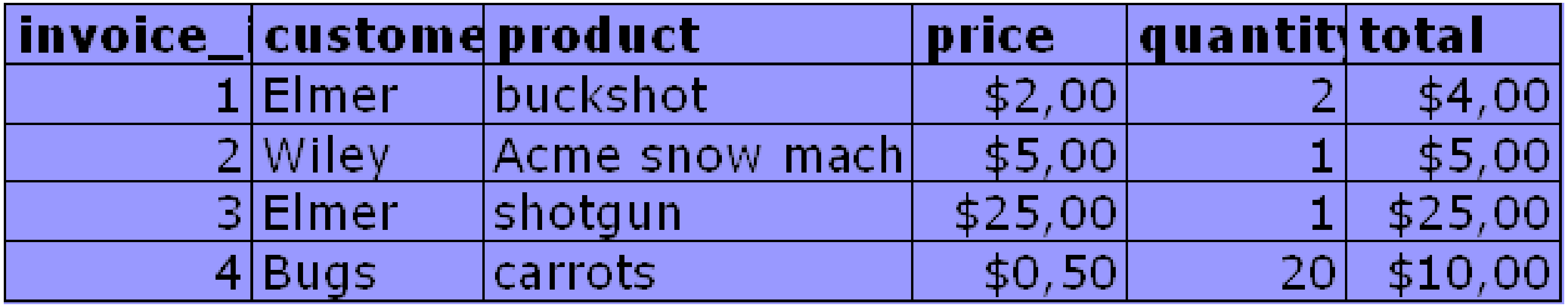
- Ανάλυση των δεδομένων σε:
  - Εγγραφές (records) και πεδία (fields), ή
  - Γραμμές και στήλες, ή
  - Κάρτες και στοιχεία.
- **Εγγραφή (record)**: Συγκεκριμένη ομάδα πληροφοριών που εισάγεται στη βάση δεδομένων. Οι πληροφορίες αυτές αναφέρονται σε **ένα γονίδιο** ή **μία πρωτεΐνη**.
- **Πεδία (fields)**: Τα μέρη στα οποία αναλύεται μία εγγραφή. Ίδια πεδία σε κάθε εγγραφή. Μοναδικό πεδίο σε κάθε εγγραφή.
  - Όνομα πρωτεΐνης / γονιδίου.
  - Αριθμών αμινοξέων / νουκλεοτιδίων.

Βάσεις δεδομένων - Κατηγορίες

- Βάσεις επίπεδης οργάνωσης (flat file):
  - Χειρίζονται μόνο ένα σύνολο από εγγραφές του ίδιου τύπου κάθε φορά.
  - Προορίζονται για τη διαχείριση καταλόγων.
  - Περιορισμένες δυνατότητες / Βοηθητικός ρόλος.

Παράδειγμα

- Filing cabinet for invoices:
  - **Table:** the filing cabinet.
  - **Rows (records):** individual invoices.
  - **Columns (fields):** data on the individual invoices *(customer, product, price, quantity).*



invoice_	custome	product	price	quantit	total
1	Elmer	buckshot	$2,00	2	$4,00
2	Wiley	Acme snow mach	$5,00	1	$5,00
3	Elmer	shotgun	$25,00	1	$25,00
4	Bugs	carrots	$0,50	20	$10,00

Εγγραφές και πεδία των βιολογικών βάσεων δεδομένων

ΠΕΔΙΑ - FIELDS

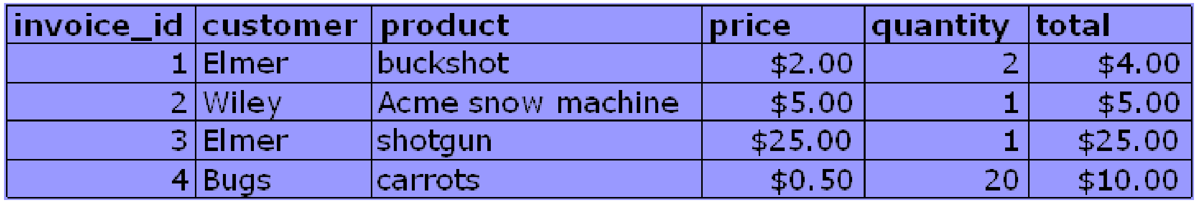
ΕΓΓΡΑΦΕΣ- RECORDS

Accession	**Protein names**	**Organism**	**Length (a.a.)**
P32234	GTP-binding protein	Drosophila melanogaster (Fruit fly)	368
P19084	11S globulin seed storage protein G3	Helianthus annuus (Common sunflower)	493
Q4U9M9	104 kDa microneme/rhoptry antigen (p104)	Theileria annulata	893
P15711	104 kDa microneme/rhoptry antigen (p104)	Theileria parva	924

- **Problem:** Storage of data in one way (e.g. alphabetic order).

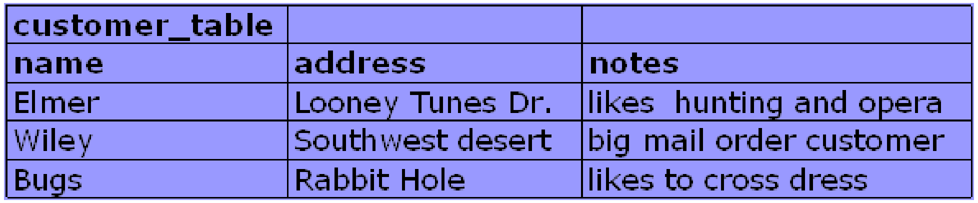
Search on any other criteria?

Παράδειγμα (συνέχεια)

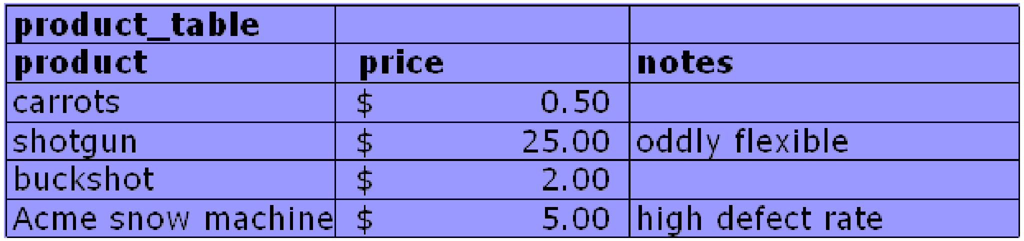


invoice_id	customer	product	price	quantity	total
1	Elmer	buckshot	$2.00	2	$4.00
2	Wiley	Acme snow machine	$5.00	1	$5.00
3	Elmer	shotgun	$25.00	1	$25.00
4	Bugs	carrots	$0.50	20	$10.00

- Σ
- Ε
- Γ
- Λ



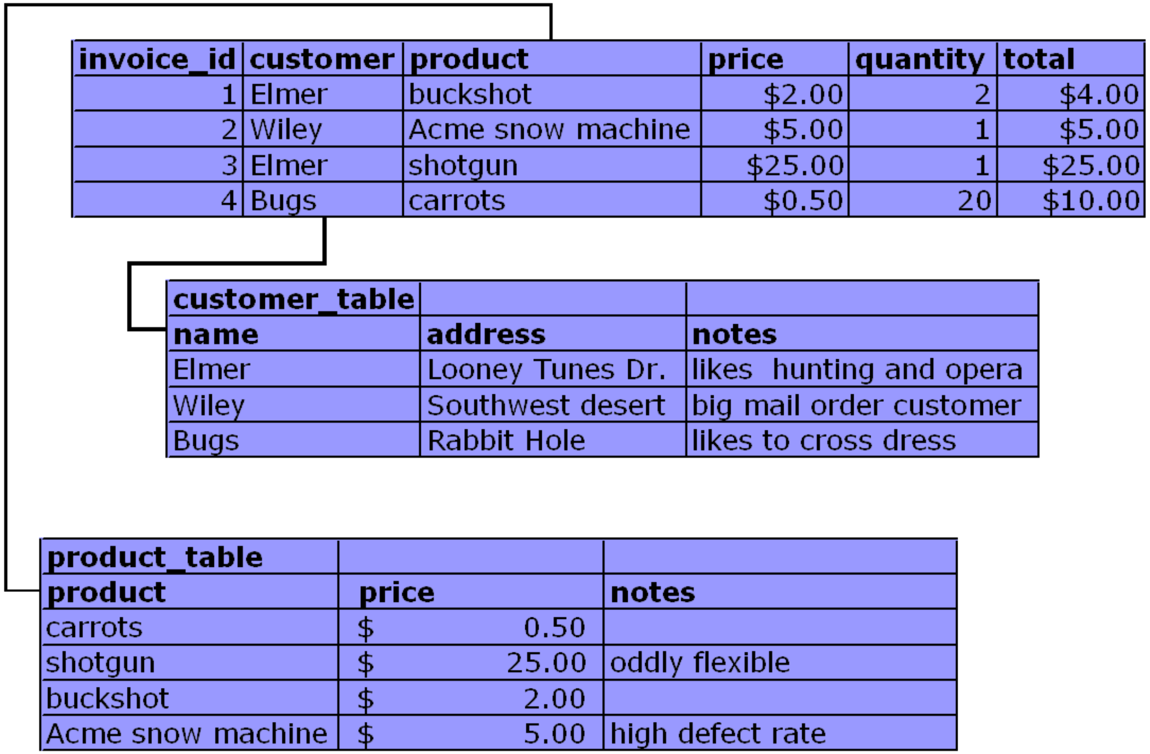
customer_table		
name	**address**	**notes**
Elmer	Looney Tunes Dr.	likes hunting and opera
Wiley	Southwest desert	big mail order customer
Bugs	Rabbit Hole	likes to cross dress



product_table		
product	**price**	**notes**
carrots	$ 0.50	
shotgun	$ 25.00	oddly flexible
buckshot	$ 2.00	
Acme snow machine	$ 5.00	high defect rate

Σχεσιακές βάσεις δεομένων
(Relational biological database)

Relationships can be built between tables and fields:



invoice_id	customer	product	price	quantity	total
1	Elmer	buckshot	$2.00	2	$4.00
2	Wiley	Acme snow machine	$5.00	1	$5.00
3	Elmer	shotgun	$25.00	1	$25.00
4	Bugs	carrots	$0.50	20	$10.00

customer_table		
name	**address**	**notes**
Elmer	Looney Tunes Dr.	likes hunting and opera
Wiley	Southwest desert	big mail order customer
Bugs	Rabbit Hole	likes to cross dress

product_table		
product	**price**	**notes**
carrots	$ 0.50	
shotgun	$ 25.00	oddly flexible
buckshot	$ 2.00	
Acme snow machine	$ 5.00	high defect rate

database "***schema***"

Βάσεις δεδομένων – Κατηγορίες (1/3)

- Σχεσιακές ή συσχετισμένες βιολογικές βάσεις δεδομένων (Relational biological database):
  - Δυνατότητα ταυτόχρονου χειρισμού και σύνδεσης πολλών συλλογών από εγγραφές διαφορετικών τύπων οργανωμένες σε πίνακες.
  - Οι πίνακες της βάσης μπορούν να συσχετιστούν μεταξύ τους, με αποτέλεσμα την ομοιόμορφη κατανομή της πληροφορίας σε όλη τη βάση.

Παράδειγμα σχεσιακής βιολογικής βάσης δεδομένων

ΠΡΩΤΕΪΝΕΣ - PROTEINS

No of protein (primary key)	Protein names	Accession	Length (a.a.)
2	GTP-binding protein	P32234	368
3	11S globulin seed storage protein G3	P19084	493

ΒΙΒΛΙΟΓΡΑΦΙΕΣ - REFERENCES

No of protein (foreign key)	Reference number (primary key)	Title	Date	Authors
2	45	The genome sequence of…	2000	Adams M.D., Celnikew S.E….

ΣΥΓΓΡΑΦΕΙΣ - AUTHORS

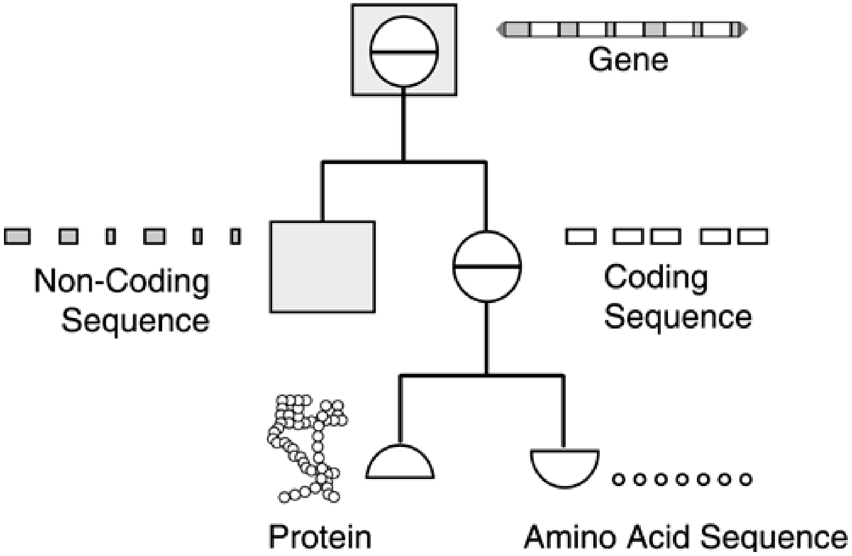
No of author (primary key)	Reference number (foreign key)	Name	Institute
2	45	Adams M.D.	Case Western Reserve University
6	45	Celnikew S.E	University of California

Πλεονεκτήματα σχεσιακών βάσεων δεδομένων

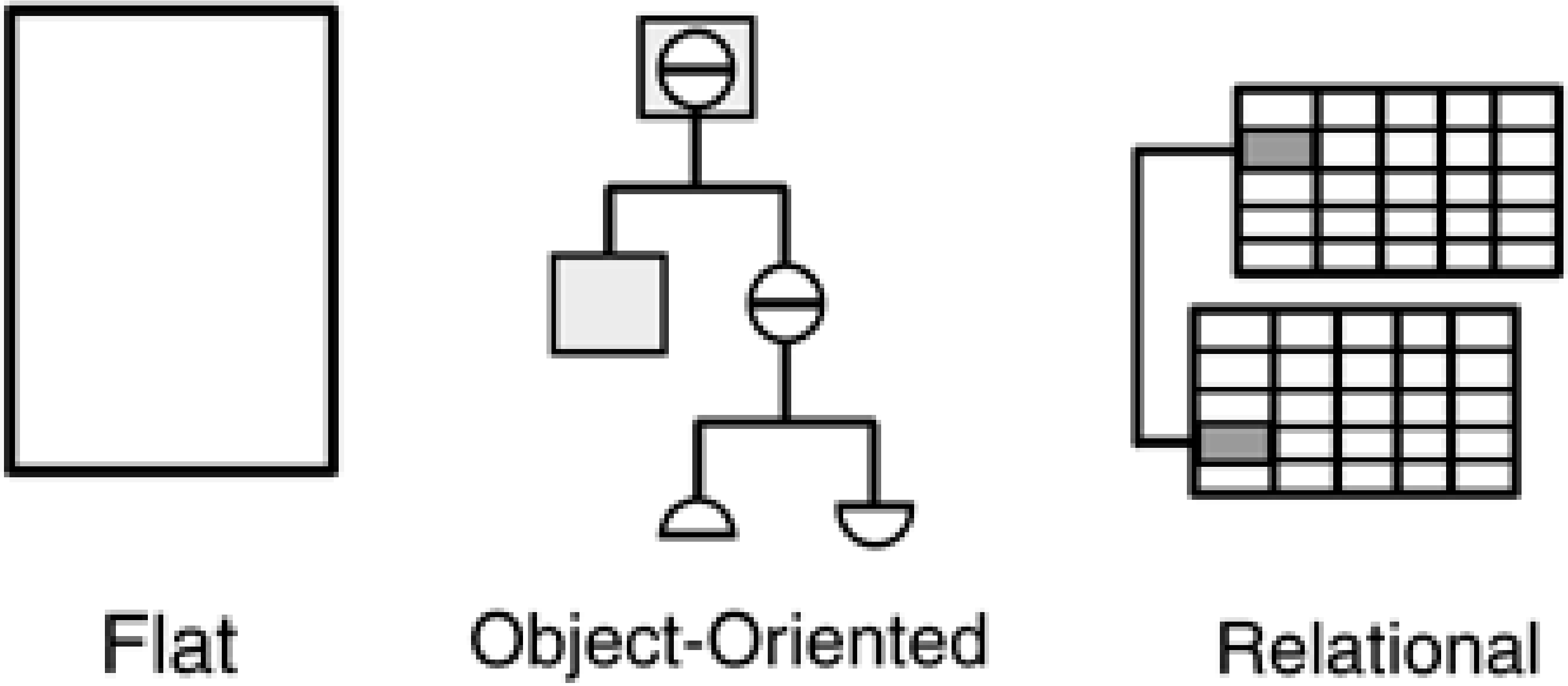
- ✓ Αποφυγή επανάληψης και συσσώρευσης δεδομένων.
- ✓ Εξασφάλιση της πληρότητας και της αξιοπιστίας των δεδομένων.
- ✓ Εύκολη άντληση της πληροφορίας.
- ▪ Πακέτα λογισμικού για τη δημιουργία, αποθήκευση και διαχείριση των βάσεων δεδομένων.
- ▪ Συστήματα διαχείρισης βάσεων δεδομένων (Database Management System - DBMS).

Βάσεις δεδομένων – Κατηγορίες (2/3)

- Αντικειμενοστραφείς Βάσεις Δεδομένων. (Object-oriented biological database):
  - Πολύπλοκες δομές δεδομένων παρουσιάζονται με σύνθετες καταχωρήσεις.
  - Αυτές οι σύνθετες καταχωρήσεις περιέχουν με τη σειρά τους άλλα αντικείμενα.
  - Τελικά οι δομές ομαδοποιούνται σε πολλά επίπεδα.

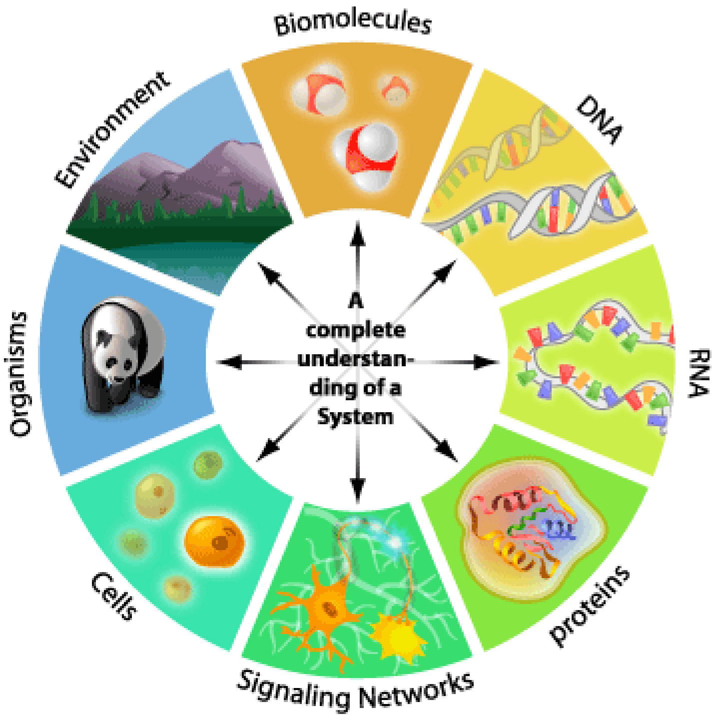


Βάσεις δεδομένων – Κατηγορίες (3/3)



Βιολογικές βάσεις δεδομένων - ΒΒΔ

- **Ορισμός:** Ηλεκτρονικές βιβλιοθήκες βιολογικών πληροφοριών.
- **Οι πληροφορίες συλλέγονται από:**
  - Επιστημονικά πειράματα.
  - Δημοσιευμένη βιβλιογραφία.
  - Πειράματα και αναλύσεις που πραγματοποιούνται στον Η/Υ.



Βιολογικές βάσεις δεδομένων - Κατηγορίες (1/3)

- **Γενικευμένες (Generalised) ή Αρχειακές (Archival): Περιλαμβάνουν πρωτογενή βιολογική πληροφορία.**
  - Πρωτογενείς βάσεις δεδομένων ακολουθιών.
  - Βάσεις δεδομένων με τρισδιάστατες δομές.

```
GATCCCTGCC TGGGCTTTGC CTCTGCAGCC CCCCGCCCCA CAGGTTCACA CCTCGGGTCT
TCTCCACCGC TGCCACACGC CAGAGCCTGT AGCGGGGCCT CAGAGTCTGG GAGGTGGGAC
TCCTGCACCT CAGCCATCAT CAGACCCATG GGGCCACCCA GGGAACCTTG GCAGGGACCA
TTACCAGTGA CCTGCCGAGG CCCCGGACTC TGCCAGCCAG CTGTGCCGGC CACCCTGCCC
CGGACAGTGC CGGTTCATGT GGGAACTAGG GGACGATGTG GTTCTTCGCA TCTGATGATG
AAGGCCCTGG GCCACTTGGC ACGGGCGGGC GCTCCCGAGA TGGATATGAG GAGCCCCCTC
TGCCCAACTC CCAGAAAGGC CGAGGCTCTG CAGCGGGAGG AAGTCCTGCG ATGTCCTGGG
GGGCAGCAGC GCAGGGCACA GGGACAGCCC CCCTCCACAG CTCTTCCTGG CCAGCCCTCC
CCACTATCTG CCAGGAGGTT GCTTCTTCCA GGAGGCTTTT CCCGACCAGC CCAGGGGTCC
AGGGTCTGGG GCTCCCAGCT GCTGTGAGTG CTGCACATTC TCTTGAGGAC AGCCCCCTCC
CTCCCCCACC CACTTCTGGT GCCCACTGTG GCCACAGCAA GCACTGGGGC CTGCACTCAG
GGACCTCGGG GCCTCCTGGG GAGCTGCTGA CCCTAGGCAG AGAGATTGCA CATCCCTAAG
AGTCTACAGA CACCCCAGTG TTTGCCAGTG TTTGCCCGTG TTCACCAGTG TTTGCCAGTG
TTTGCCAGTA TTTGCTCGCC AGTGTTCGCC ACTTGTCCCT CTGGCTGCAA GAGTGACTGG
GTTTGGGCGG GAAGTTGCAG GTCCCTCCAG GACAGTTGGC CGATGACGTG GAGACAGACC
CACCCCCCAA TCCTGGCTCC CTGCAGGACG CGGGGCCCCC CGAGATCCTG GCGGTGCTCA
GCACGACGGG CACCTCCGTG TTCACCAGTC CAATGGGCAC GGAGCGTGGC TTTATTTGCA
TGTCTGGATT CCTAACGACT TCAGCCTCTG CACCTCCTGG GTTTTCCCTG CTGCAAATTG
CCATTTGGCG TCGTCCCCAA TTTCCGGCCA AGGCCGCGTC GTCGTGCTGC TGTGTAATTT
GATGTGTGGA GTTCTAGATA CCAAGTGTCT GTCGGTTTTA GACATCGCAA ACGTCCTTCC
CAGTGTGGCC CGTCCATTCG CTTCTGTGCA GCAAAATCTT TAATTATTTG ATGGCATCAA
AATGTGTGTC CAGTTTTACC TTCTAGTTTA TACTTTCGAA CATTTGTTTG AGAAATCTTT
CTCCCACCTG TGGCTGATAG TGACGTCTTC TAACTTCCCA TTTACTATGT TACATTCAGA
CCCATCATCT TCAGGAAGAC GCTTGTGTGC GAGACGGGTA TGAGGCCCCC ACACCCCGCC
TCAGGACCAC TGTCCATGGT TCCACCCCTG ACCCCGGACT CCGCTCCCCA GACCTCCTAA
```

Βιολογικές βάσεις δεδομένων - Κατηγορίες (2/3)

- **Δευτερεύουσες (Secondary) βιολογικές βάσεις δεδομένων: Τα δεδομένα τους προκύπτουν από ανάλυση των δεδομένων των αρχειακών βάσεων δεδομένων.**
  - ΒΔ στις οποίες έχουν ληφθεί υπόψη οι πρωτεϊνικές αλληλουχίες **στην πρόβλεψη** της τρισδιάστατης δομής των πρωτεϊνών.
  - ΒΔ που, ιεραρχούν/συσχετίζουν οικογένειες πρωτεϊνών, δομικά κοινές πρωτεΐνες, κοινά μοτίβα ακολουθιών DNA και πρωτεϊνών.
  - ΒΔ που καταγράφουν μεταλλαγές ή παραλλαγές στις ακολουθίες DNA ή πρωτεϊνών.

Βιολογικές βάσεις δεδομένων - Κατηγορίες (3/3)

- **Βιολογικές βάσεις δεδομένων ιστοσελίδων που περιλαμβάνουν:**
  - Βάσεις δεδομένων που περιέχουν ως εγγραφές βιολογικές βάσεις δεδομένων.
  - Συνδέσμους μεταξύ των βιολογικών βάσεων δεδομένων.
- **Εξειδικευμένες βάσεις δεδομένων:**
  - ΒΔ μικροσυστοιχιών (microarrays).
  - ΒΔ μεταβολικών μονοπατιών.
- **Βιβλιογραφικές βάσεις δεδομένων.**

Βιβλιογραφικές βάσεις δεδομένων

dblp.uni-trier.de

Computer Science Bibliography

PubMed

Πληροφορική
Ιατρική - Βιολογία
Βιομοριακή Τεχνολογία

Είδη δημοσιεύσεων

- Books – Βιβλία (μονογραφίες).
- Journals – Επιστημονικά περιοδικά.
- Conferences – Επιστημονικά συνέδρια.
- Research articles - Ερευνητικά Άρθρα.
- Review articles – Άρθρα ανασκόπησης.
- Meta – analysis – Ανάλυση αποτελεσμάτων προηγούμενων δημοσιεύσεων.

Η δομή μίας δημοσίευσης – Computer science

1. Abstract.
2. Introduction.
3. Method – Algorithm – Tools.
4. Results / Conclusions.
5. References.

Ερευνητικό άρθρο - Research Article (1/2)

BIOINFORMATICS **APPLICATIONS NOTE**

Vol. 27 no. 7 2011, pages 1034–1035
doi:10.1093/bioinformatics/btr047

Data and text mining

Advance Access publication February 9, 2011

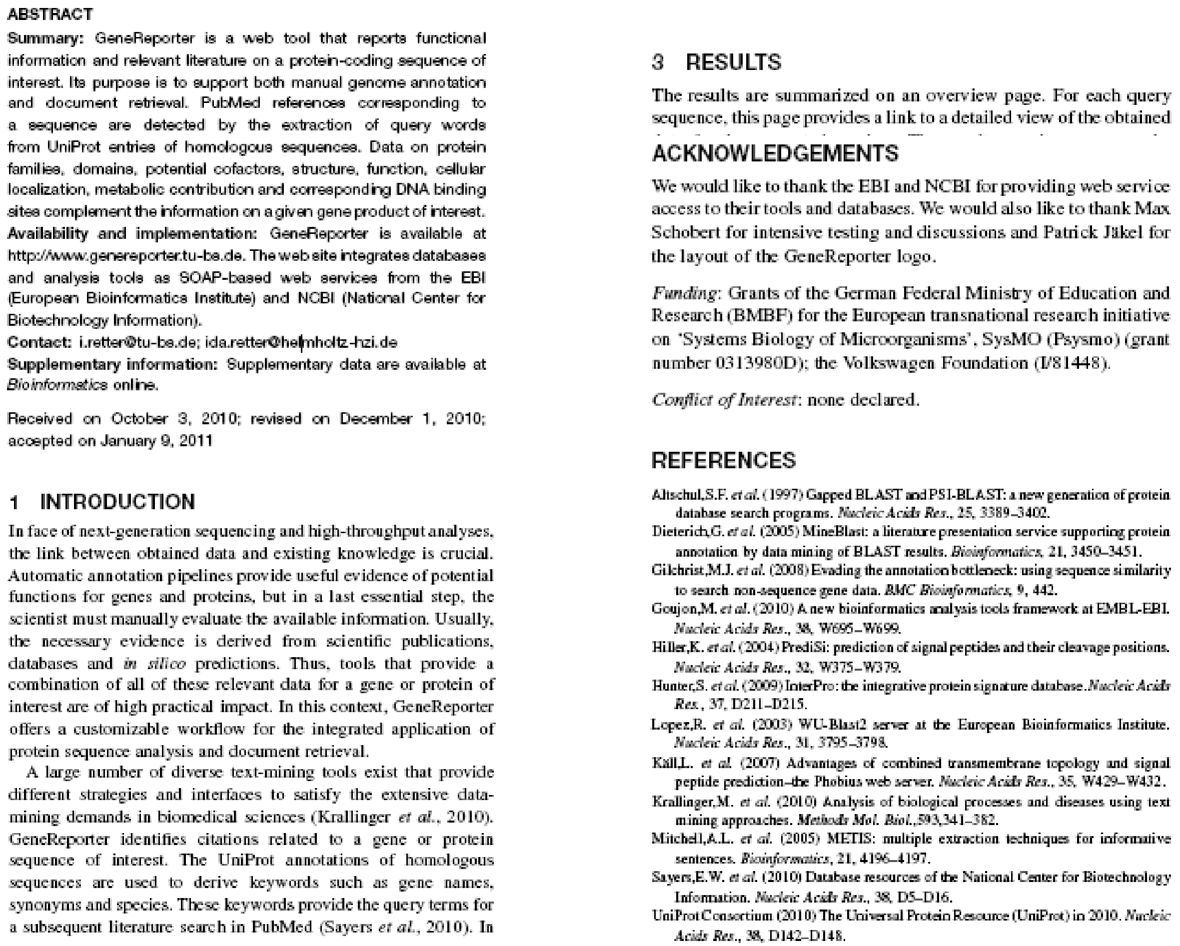
GeneReporter—sequence-based document retrieval and annotation

Annekathrin Bartsch[1], Boyke Bunk[1], Isam Haddad[1], Johannes Klein[1], Richard Münch[1], Thorsten Johl[2], Uwe Kärst[2], Lothar Jänsch[2], Dieter Jahn[1] and Ida Retter[1,*]

[1]Institute for Microbiology, Technische Universität Braunschweig, Spielmannstr. 7, 38106 Braunschweig and [2]Cellular Proteomics Group, Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124 Braunschweig, Germany

Associate Editor: Jonathan Wren

Ερευνητικό άρθρο - Research Article (2/2)



ABSTRACT

Summary: GeneReporter is a web tool that reports functional information and relevant literature on a protein-coding sequence of interest. Its purpose is to support both manual genome annotation and document retrieval. PubMed references corresponding to a sequence are detected by the extraction of query words from UniProt entries of homologous sequences. Data on protein families, domains, potential cofactors, structure, function, cellular localization, metabolic contribution and corresponding DNA binding sites complement the information on a given gene product of interest.

Availability and implementation: GeneReporter is available at http://www.genereporter.tu-bs.de. The web site integrates databases and analysis tools as SOAP-based web services from the EBI (European Bioinformatics Institute) and NCBI (National Center for Biotechnology Information).

Contact: i.retter@tu-bs.de; ida.retter@helmholtz-hzi.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on October 3, 2010; revised on December 1, 2010; accepted on January 9, 2011

1 INTRODUCTION

In face of next-generation sequencing and high-throughput analyses, the link between obtained data and existing knowledge is crucial. Automatic annotation pipelines provide useful evidence of potential functions for genes and proteins, but in a last essential step, the scientist must manually evaluate the available information. Usually, the necessary evidence is derived from scientific publications, databases and *in silico* predictions. Thus, tools that provide a combination of all of these relevant data for a gene or protein of interest are of high practical impact. In this context, GeneReporter offers a customizable workflow for the integrated application of protein sequence analysis and document retrieval.

A large number of diverse text-mining tools exist that provide different strategies and interfaces to satisfy the extensive data-mining demands in biomedical sciences (Krallinger *et al.*, 2010). GeneReporter identifies citations related to a gene or protein sequence of interest. The UniProt annotations of homologous sequences are used to derive keywords such as gene names, synonyms and species. These keywords provide the query terms for a subsequent literature search in PubMed (Sayers *et al.*, 2010). In

3 RESULTS

The results are summarized on an overview page. For each query sequence, this page provides a link to a detailed view of the obtained

ACKNOWLEDGEMENTS

We would like to thank the EBI and NCBI for providing web service access to their tools and databases. We would also like to thank Max Schobert for intensive testing and discussions and Patrick Jäkel for the layout of the GeneReporter logo.

Funding: Grants of the German Federal Ministry of Education and Research (BMBF) for the European transnational research initiative on 'Systems Biology of Microorganisms', SysMO (Psysmo) (grant number 0313980D); the Volkswagen Foundation (I/81448).

Conflict of Interest: none declared.

REFERENCES

Altschul,S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25, 3389–3402.

Dieterich,G. *et al.* (2005) MineBlast: a literature presentation service supporting protein annotation by data mining of BLAST results. *Bioinformatics*, 21, 3450–3451.

Gilchrist,M.J. *et al.* (2008) Evading the annotation bottleneck: using sequence similarity to search non-sequence gene data. *BMC Bioinformatics*, 9, 442.

Goujon,M. *et al.* (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.*, 38, W695–W699.

Hiller,K. *et al.* (2004) PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res.*, 32, W375–W379.

Hunter,S. *et al.* (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.*, 37, D211–D215.

Lopez,R. *et al.* (2003) WU-Blast2 server at the European Bioinformatics Institute. *Nucleic Acids Res.*, 31, 3795–3798.

Käll,L. *et al.* (2007) Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server. *Nucleic Acids Res.*, 35, W429–W432.

Krallinger,M. *et al.* (2010) Analysis of biological processes and diseases using text mining approaches. *Methods Mol. Biol.*,593,341–382.

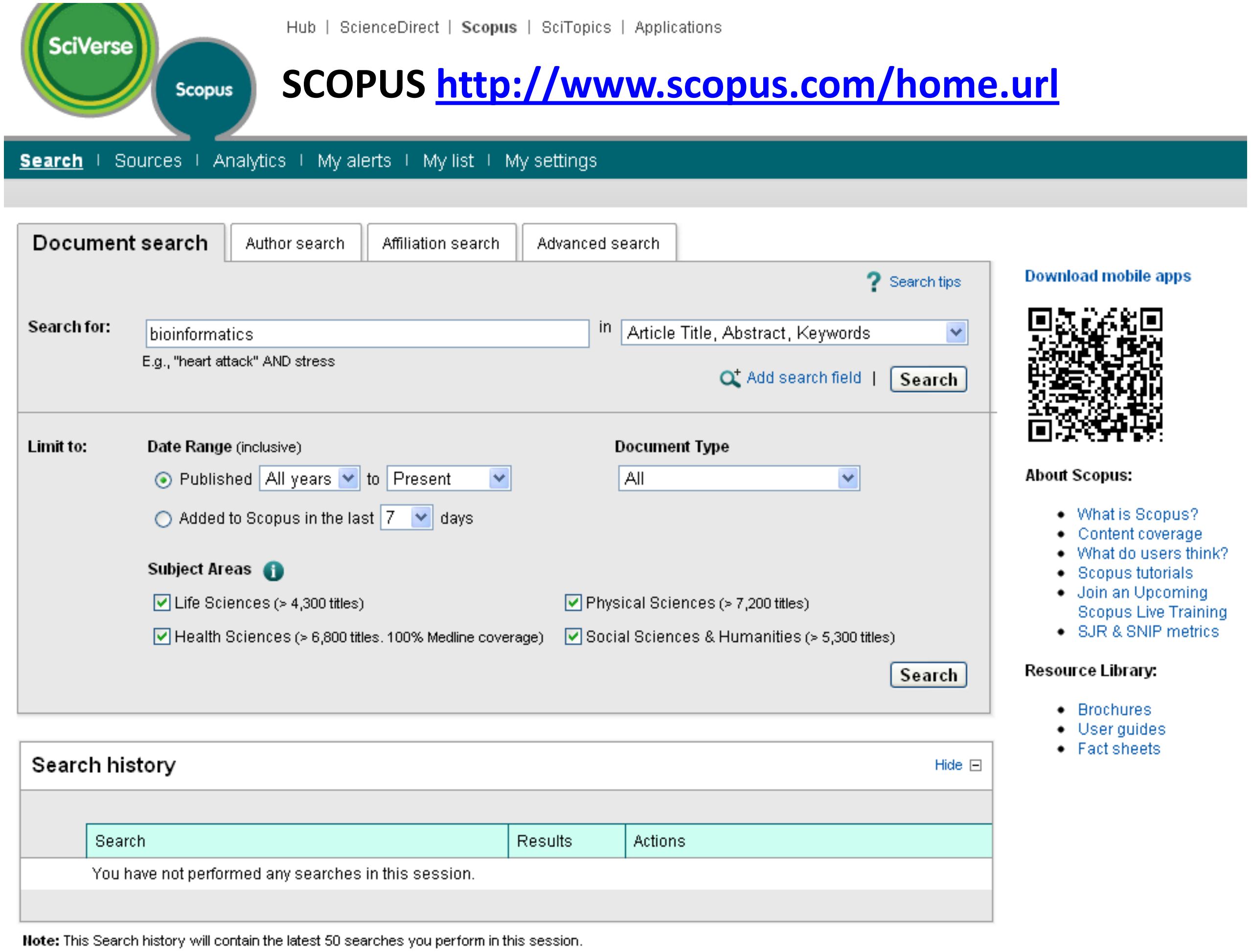
Mitchell,A.L. *et al.* (2005) METIS: multiple extraction techniques for informative sentences. *Bioinformatics*, 21, 4196–4197.

Sayers,E.W. *et al.* (2010) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, 38, D5–D16.

UniProt Consortium (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, 38, D142–D148.

SCOPUS (1/3)

SCOPUS http://www.scopus.com/home.url



Search for: bioinformatics in Article Title, Abstract, Keywords

E.g., "heart attack" AND stress

Limit to:

Date Range (inclusive)

● Published All years to Present

○ Added to Scopus in the last 7 days

Document Type

All

Subject Areas

☑ Life Sciences (> 4,300 titles)

☑ Health Sciences (> 6,800 titles. 100% Medline coverage)

☑ Physical Sciences (> 7,200 titles)

☑ Social Sciences & Humanities (> 5,300 titles)

Download mobile apps

About Scopus:

- What is Scopus?
- Content coverage
- What do users think?
- Scopus tutorials
- Join an Upcoming Scopus Live Training
- SJR & SNIP metrics

Resource Library:

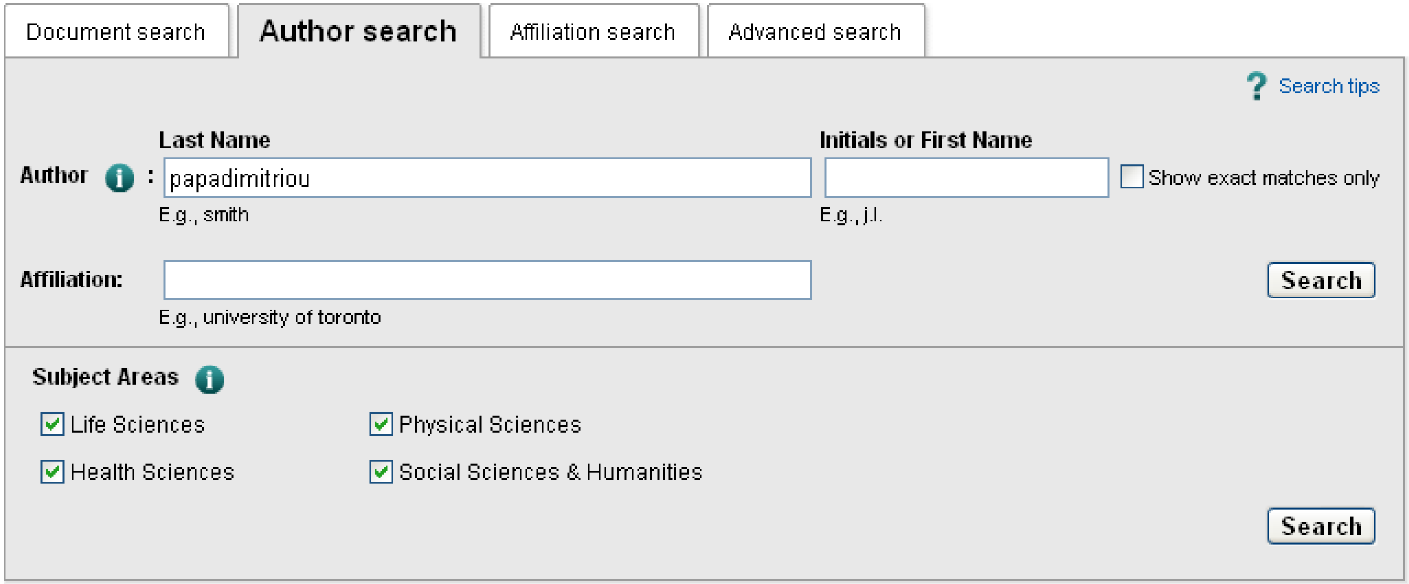
- Brochures
- User guides
- Fact sheets

Search history

Search	Results	Actions
You have not performed any searches in this session.		

Note: This Search history will contain the latest 50 searches you perform in this session.

SCOPUS (2/3)

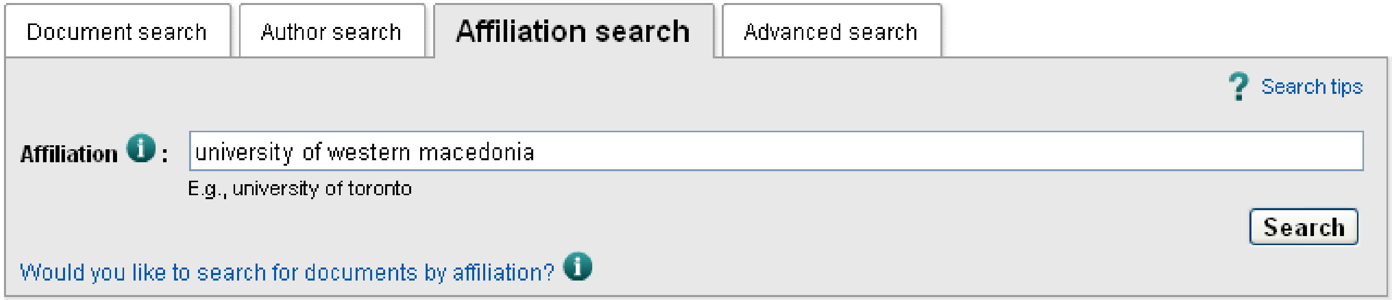


Author : Last Name: papadimitriou — E.g., smith
Initials or First Name: — E.g., j.l.
☐ Show exact matches only

Affiliation: — E.g., university of toronto

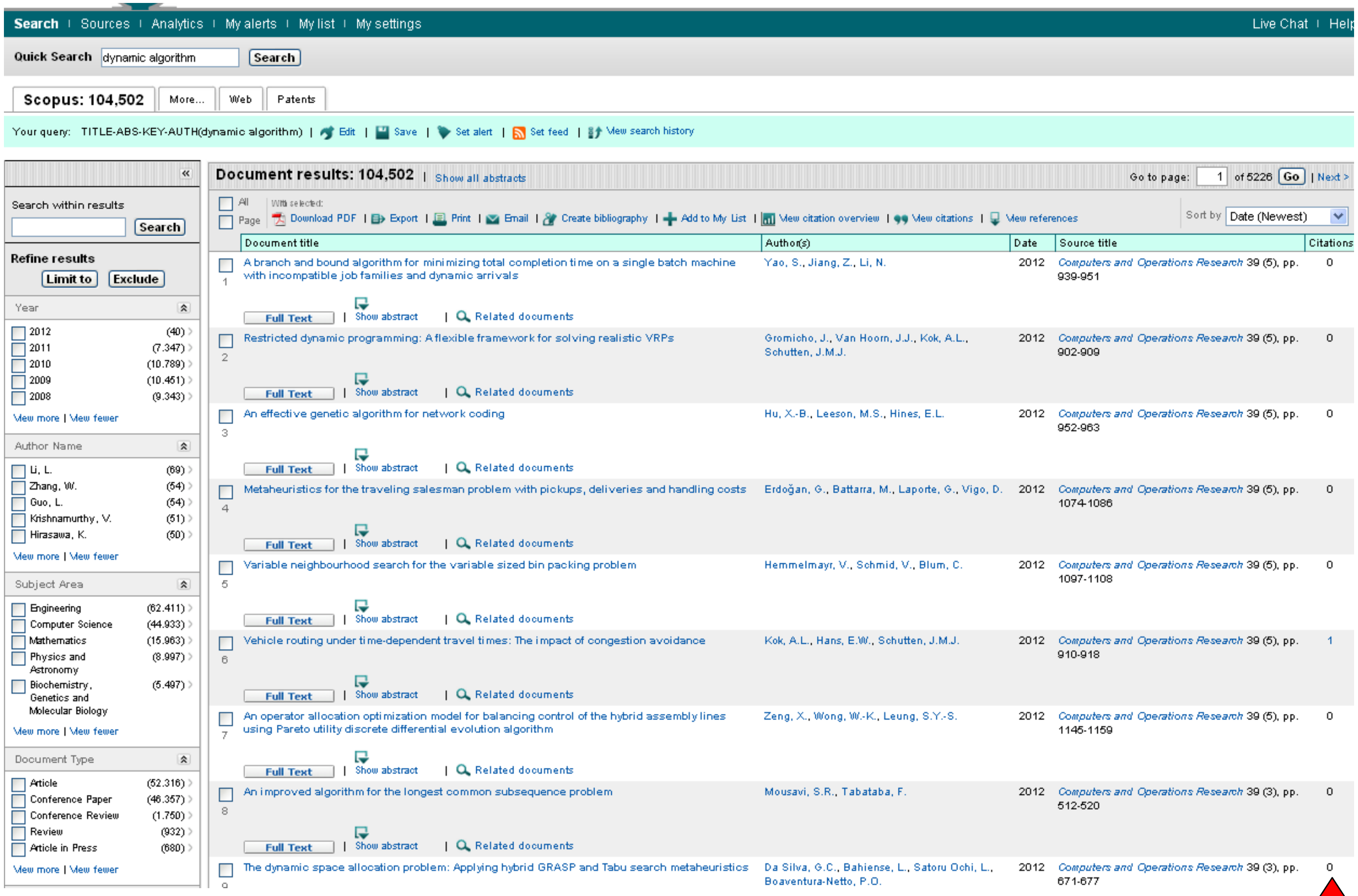
Subject Areas

☑ Life Sciences
☑ Physical Sciences
☑ Health Sciences
☑ Social Sciences & Humanities



Affiliation : university of western macedonia — E.g., university of toronto

SCOPUS (3/3)



Quick Search: dynamic algorithm

Your query: TITLE-ABS-KEY-AUTH(dynamic algorithm)

Search within results

Refine results

Year

☐ 2012	(40)
☐ 2011	(7.347)
☐ 2010	(10.789)
☐ 2009	(10.451)
☐ 2008	(9.343)

View more | View fewer

Author Name

☐ Li, L.	(69)
☐ Zhang, W.	(54)
☐ Guo, L.	(54)
☐ Krishnamurthy, V.	(51)
☐ Hirasawa, K.	(50)

View more | View fewer

Subject Area

☐ Engineering	(62.411)
☐ Computer Science	(44.933)
☐ Mathematics	(15.963)
☐ Physics and Astronomy	(8.997)
☐ Biochemistry, Genetics and Molecular Biology	(5.497)

View more | View fewer

Document Type

☐ Article	(52.316)
☐ Conference Paper	(46.357)
☐ Conference Review	(1.750)
☐ Review	(932)
☐ Article in Press	(680)

View more | View fewer

Document results: 104,502 | Show all abstracts

Go to page: 1 of 5226 Go | Next >

☐ All
☐ Page
With selected:

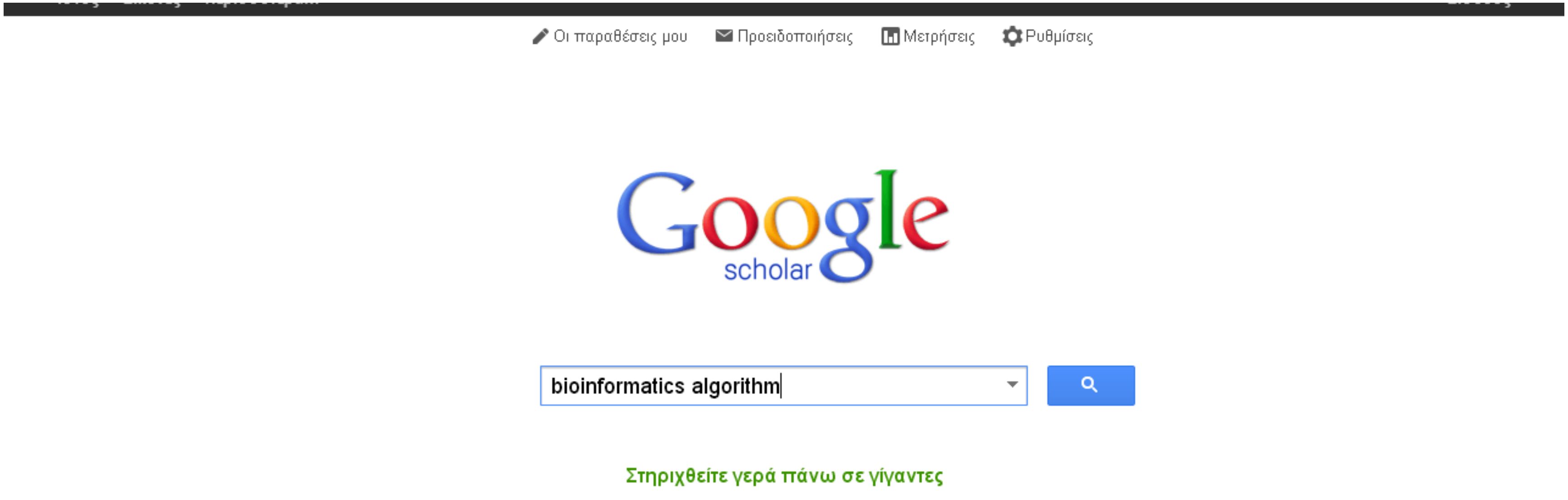
Sort by: Date (Newest)

	Document title	Author(s)	Date	Source title	Citations
☐ 1	A branch and bound algorithm for minimizing total completion time on a single batch machine with incompatible job families and dynamic arrivals	Yao, S., Jiang, Z., Li, N.	2012	*Computers and Operations Research* 39 (5), pp. 939-951	0
☐ 2	Restricted dynamic programming: A flexible framework for solving realistic VRPs	Gromicho, J., Van Hoorn, J.J., Kok, A.L., Schutten, J.M.J.	2012	*Computers and Operations Research* 39 (5), pp. 902-909	0
☐ 3	An effective genetic algorithm for network coding	Hu, X.-B., Leeson, M.S., Hines, E.L.	2012	*Computers and Operations Research* 39 (5), pp. 952-963	0
☐ 4	Metaheuristics for the traveling salesman problem with pickups, deliveries and handling costs	Erdoğan, G., Battarra, M., Laporte, G., Vigo, D.	2012	*Computers and Operations Research* 39 (5), pp. 1074-1086	0
☐ 5	Variable neighbourhood search for the variable sized bin packing problem	Hemmelmayr, V., Schmid, V., Blum, C.	2012	*Computers and Operations Research* 39 (5), pp. 1097-1108	0
☐ 6	Vehicle routing under time-dependent travel times: The impact of congestion avoidance	Kok, A.L., Hans, E.W., Schutten, J.M.J.	2012	*Computers and Operations Research* 39 (5), pp. 910-918	1
☐ 7	An operator allocation optimization model for balancing control of the hybrid assembly lines using Pareto utility discrete differential evolution algorithm	Zeng, X., Wong, W.-K., Leung, S.Y.-S.	2012	*Computers and Operations Research* 39 (5), pp. 1145-1159	0
☐ 8	An improved algorithm for the longest common subsequence problem	Mousavi, S.R., Tabataba, F.	2012	*Computers and Operations Research* 39 (3), pp. 512-520	0
☐ 9	The dynamic space allocation problem: Applying hybrid GRASP and Tabu search metaheuristics	Da Silva, G.C., Bahiense, L., Satoru Ochi, L., Boaventura-Netto, P.O.	2012	*Computers and Operations Research* 39 (3), pp. 671-677	0

Scopus, Πηγές (Sources)

- Agricultural and Biological Science.
- Arts and Humanities.
- Biochemistry, Genetics and Molecular Biology.
- Business, Management and Accounting.
- Chemical Engineering.
- Chemistry.
- **Computer Science.**
- Dentistry.
- Earth and Planetary Sciences.
- Economics, Econometrics and Finance.
- Energy.
- **Engineering.**
- Environmental Science.
- Health Professions.
- Immunology and Microbiology.
- Materials Science.
- Mathematics
- Medicine
- Neuroscience.
- Nursing.
- Pharmacology, Toxicology and Pharmaceutics.
- Physics and Astronomy.
- Psychology.
- Social Sciences.
- Veterinary.

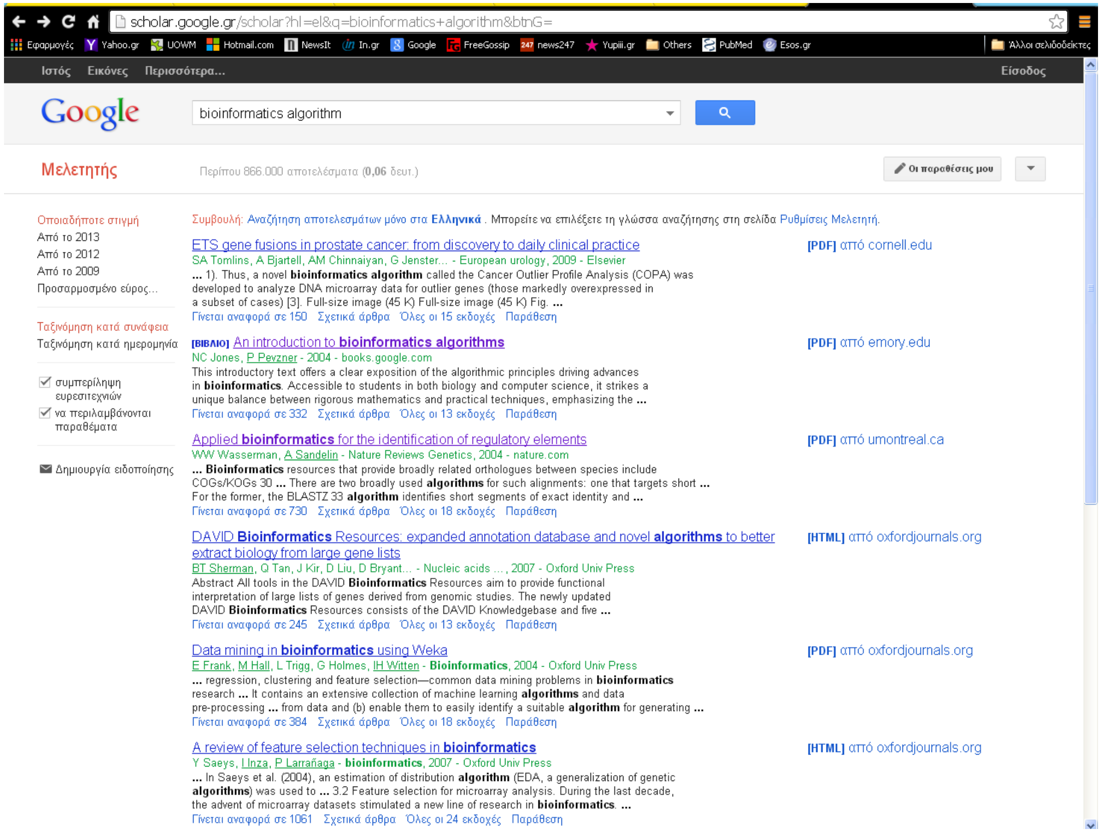
Google (1/6)



bioinformatics algorithm

Στηριχθείτε γερά πάνω σε γίγαντες

Google (2/6)



bioinformatics algorithm

Περίπου 866.000 αποτελέσματα (**0,06** δευτ.)

Συμβουλή: Αναζήτηση αποτελεσμάτων μόνο στα **Ελληνικά** . Μπορείτε να επιλέξετε τη γλώσσα αναζήτησης στη σελίδα Ρυθμίσεις Μελετητή.

ETS gene fusions in prostate cancer: from discovery to daily clinical practice — [PDF] από cornell.edu
SA Tomlins, A Bjartell, AM Chinnaiyan, G Jenster... - European urology, 2009 - Elsevier
... 1). Thus, a novel **bioinformatics algorithm** called the Cancer Outlier Profile Analysis (COPA) was developed to analyze DNA microarray data for outlier genes (those markedly overexpressed in a subset of cases) [3]. Full-size image (45 K) Full-size image (45 K) Fig. ...
Γίνεται αναφορά σε 150 Σχετικά άρθρα Όλες οι 15 εκδοχές Παράθεση

[ΒΙΒΛΙΟ] An introduction to **bioinformatics algorithms** — [PDF] από emory.edu
NC Jones, P Pevzner - 2004 - books.google.com
This introductory text offers a clear exposition of the algorithmic principles driving advances in **bioinformatics**. Accessible to students in both biology and computer science, it strikes a unique balance between rigorous mathematics and practical techniques, emphasizing the ...
Γίνεται αναφορά σε 332 Σχετικά άρθρα Όλες οι 13 εκδοχές Παράθεση

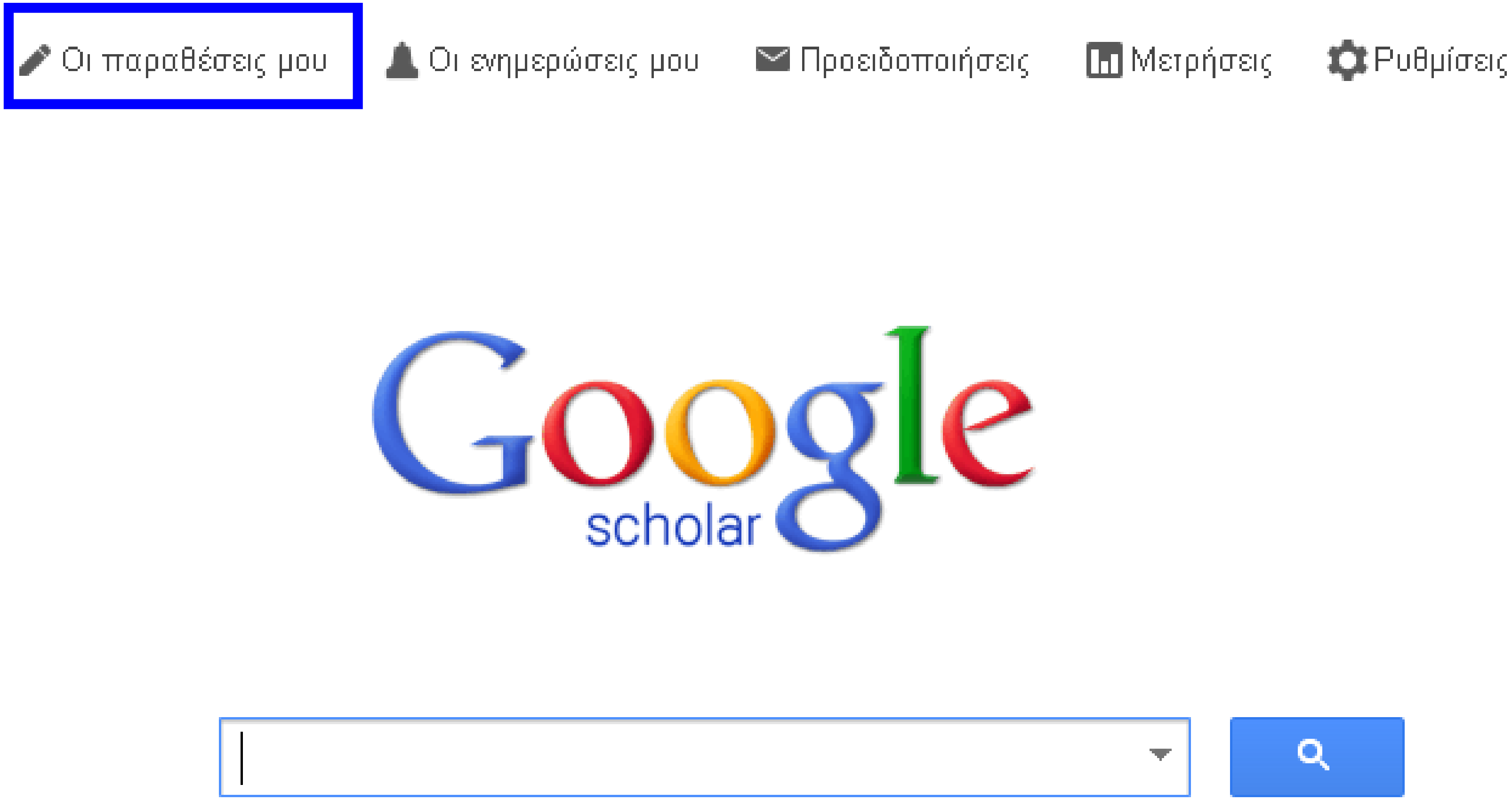
Applied **bioinformatics** for the identification of regulatory elements — [PDF] από umontreal.ca
WW Wasserman, A Sandelin - Nature Reviews Genetics, 2004 - nature.com
... **Bioinformatics** resources that provide broadly related orthologues between species include COGs/KOGs 30 ... There are two broadly used **algorithms** for such alignments: one that targets short ... For the former, the BLASTZ 33 **algorithm** identifies short segments of exact identity and ...
Γίνεται αναφορά σε 730 Σχετικά άρθρα Όλες οι 18 εκδοχές Παράθεση

DAVID **Bioinformatics** Resources: expanded annotation database and novel **algorithms** to better extract biology from large gene lists — [HTML] από oxfordjournals.org
BT Sherman, Q Tan, J Kir, D Liu, D Bryant... - Nucleic acids ..., 2007 - Oxford Univ Press
Abstract All tools in the DAVID **Bioinformatics** Resources aim to provide functional interpretation of large lists of genes derived from genomic studies. The newly updated DAVID **Bioinformatics** Resources consists of the DAVID Knowledgebase and five ...
Γίνεται αναφορά σε 245 Σχετικά άρθρα Όλες οι 13 εκδοχές Παράθεση

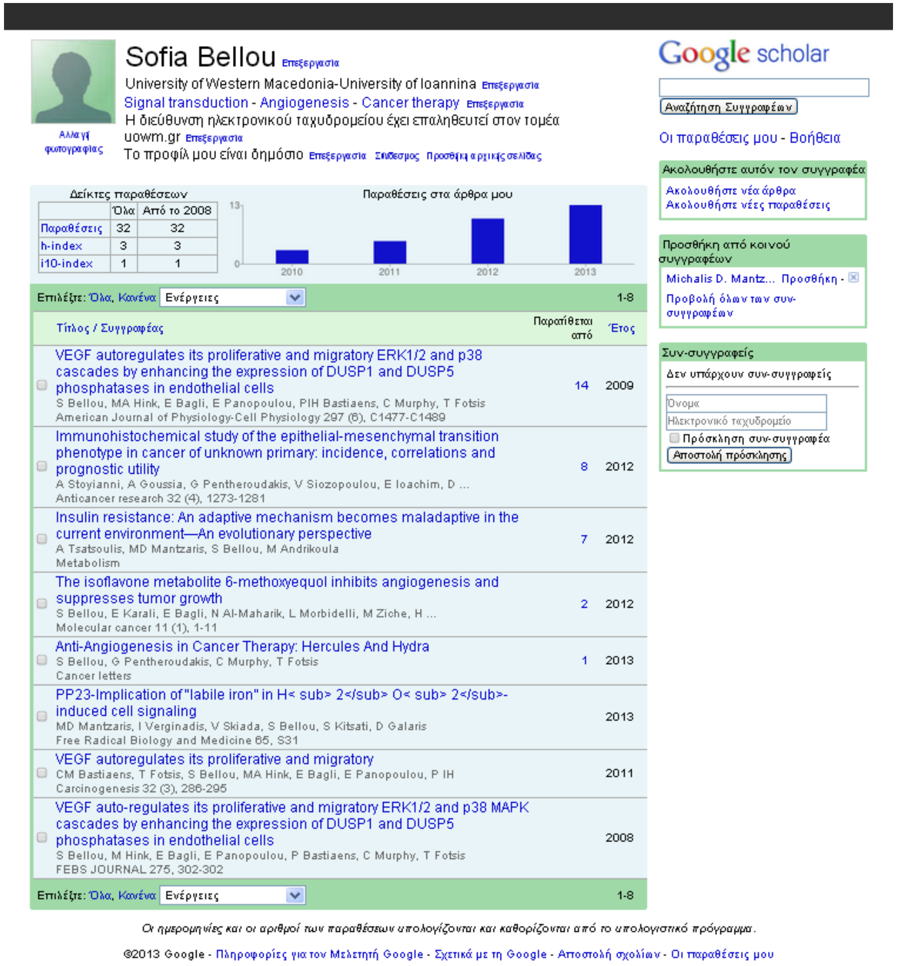
Data mining in **bioinformatics** using Weka — [PDF] από oxfordjournals.org
E Frank, M Hall, L Trigg, G Holmes, IH Witten - **Bioinformatics**, 2004 - Oxford Univ Press
... regression, clustering and feature selection—common data mining problems in **bioinformatics** research ... It contains an extensive collection of machine learning **algorithms** and data pre-processing ... from data and (b) enable them to easily identify a suitable **algorithm** for generating ...
Γίνεται αναφορά σε 384 Σχετικά άρθρα Όλες οι 18 εκδοχές Παράθεση

A review of feature selection techniques in **bioinformatics** — [HTML] από oxfordjournals.org
Y Saeys, I Inza, P Larrañaga - **bioinformatics**, 2007 - Oxford Univ Press
... In Saeys et al. (2004), an estimation of distribution **algorithm** (EDA, a generalization of genetic **algorithms**) was used to ... 3.2 Feature selection for microarray analysis. During the last decade, the advent of microarray datasets stimulated a new line of research in **bioinformatics**. ...
Γίνεται αναφορά σε 1061 Σχετικά άρθρα Όλες οι 24 εκδοχές Παράθεση

Google (3/6)


Οι παραθέσεις μου
Οι ενημερώσεις μου
Προειδοποιήσεις
Μετρήσεις
Ρυθμίσεις
Google
scholar

Google (4/6)



Sofia Bellou

University of Western Macedonia-University of Ioannina

Signal transduction - Angiogenesis - Cancer therapy

Η διεύθυνση ηλεκτρονικού ταχυδρομείου έχει επαληθευτεί στον τομέα uowm.gr

Το προφίλ μου είναι δημόσιο

Οι παραθέσεις μου - Βοήθεια

Ακολουθήστε αυτόν τον συγγραφέα
- Ακολουθήστε νέα άρθρα
- Ακολουθήστε νέες παραθέσεις

Προσθήκη από κοινού συγγραφέων
- Michalis D. Mantz... Προσθήκη - ☒
- Προβολή όλων των συν-συγγραφέων

Συν-συγγραφείς

Δεν υπάρχουν συν-συγγραφείς

Όνομα

Ηλεκτρονικό ταχυδρομείο

☐ Πρόσκληση συν-συγγραφέα

Αποστολή πρόσκλησης

Δείκτες παραθέσεων

	Όλα	Από το 2008
Παραθέσεις	32	32
h-index	3	3
i10-index	1	1

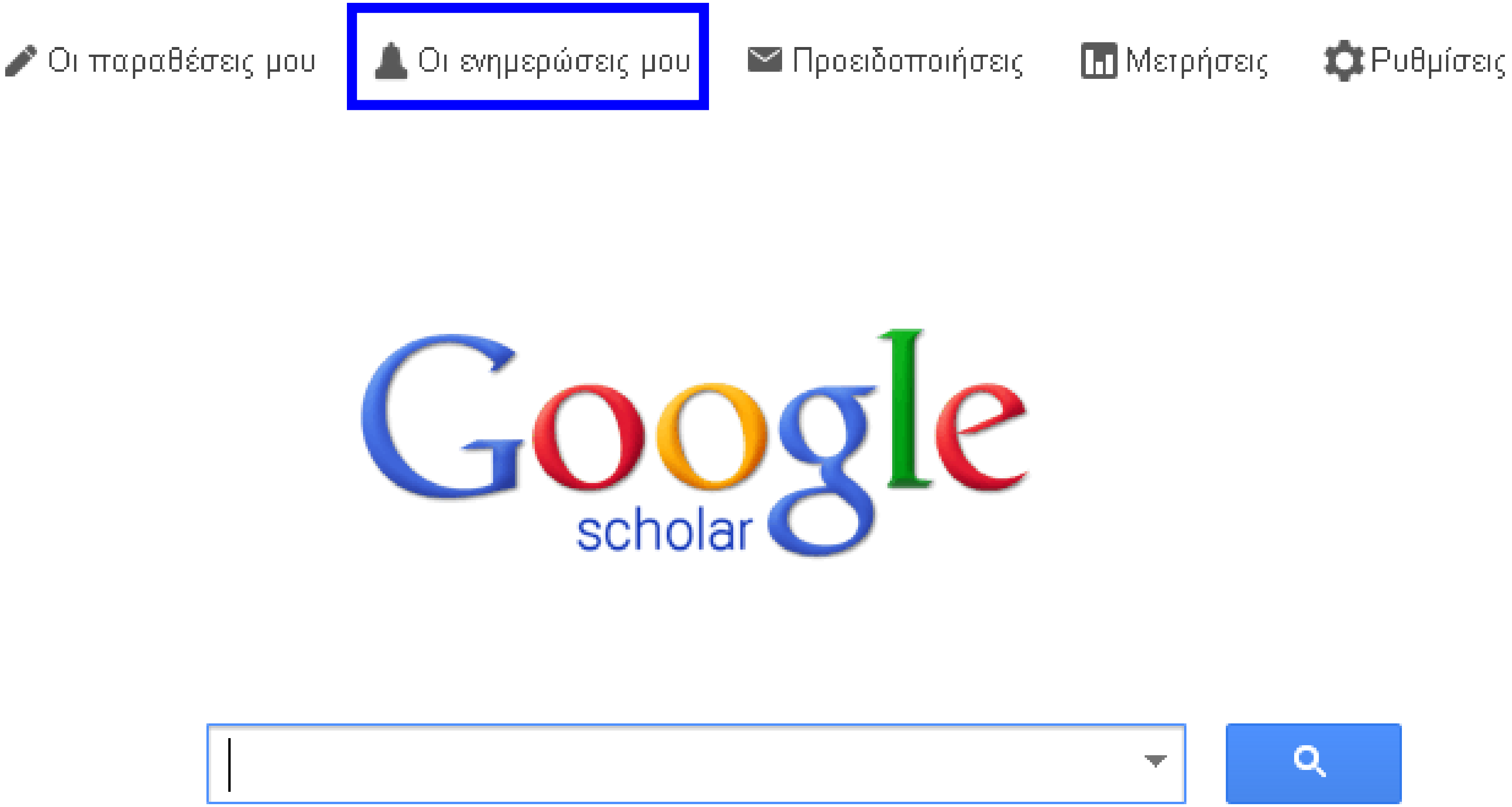
Παραθέσεις στα άρθρα μου

	Τίτλος / Συγγραφέας	Παρατίθεται από	Έτος
☐	VEGF autoregulates its proliferative and migratory ERK1/2 and p38 cascades by enhancing the expression of DUSP1 and DUSP5 phosphatases in endothelial cells S Bellou, MA Hink, E Bagli, E Panopoulou, PIH Bastiaens, C Murphy, T Fotsis American Journal of Physiology-Cell Physiology 297 (6), C1477-C1489	14	2009
☐	Immunohistochemical study of the epithelial-mesenchymal transition phenotype in cancer of unknown primary: incidence, correlations and prognostic utility A Stoyianni, A Goussia, G Pentheroudakis, V Siozopoulou, E Ioachim, D ... Anticancer research 32 (4), 1273-1281	8	2012
☐	Insulin resistance: An adaptive mechanism becomes maladaptive in the current environment—An evolutionary perspective A Tsatsoulis, MD Mantzaris, S Bellou, M Andrikoula Metabolism	7	2012
☐	The isoflavone metabolite 6-methoxyequol inhibits angiogenesis and suppresses tumor growth S Bellou, E Karali, E Bagli, N Al-Maharik, L Morbidelli, M Ziche, H ... Molecular cancer 11 (1), 1-11	2	2012
☐	Anti-Angiogenesis in Cancer Therapy: Hercules And Hydra S Bellou, G Pentheroudakis, C Murphy, T Fotsis Cancer letters	1	2013
☐	PP23-Implication of "labile iron" in H< sub> 2</sub> O< sub> 2</sub>-induced cell signaling MD Mantzaris, I Verginadis, V Skiada, S Bellou, S Kitsati, D Galaris Free Radical Biology and Medicine 65, S31		2013
☐	VEGF autoregulates its proliferative and migratory CM Bastiaens, T Fotsis, S Bellou, MA Hink, E Bagli, E Panopoulou, P IH Carcinogenesis 32 (3), 286-295		2011
☐	VEGF auto-regulates its proliferative and migratory ERK1/2 and p38 MAPK cascades by enhancing the expression of DUSP1 and DUSP5 phosphatases in endothelial cells S Bellou, M Hink, E Bagli, E Panopoulou, P Bastiaens, C Murphy, T Fotsis FEBS JOURNAL 275, 302-302		2008

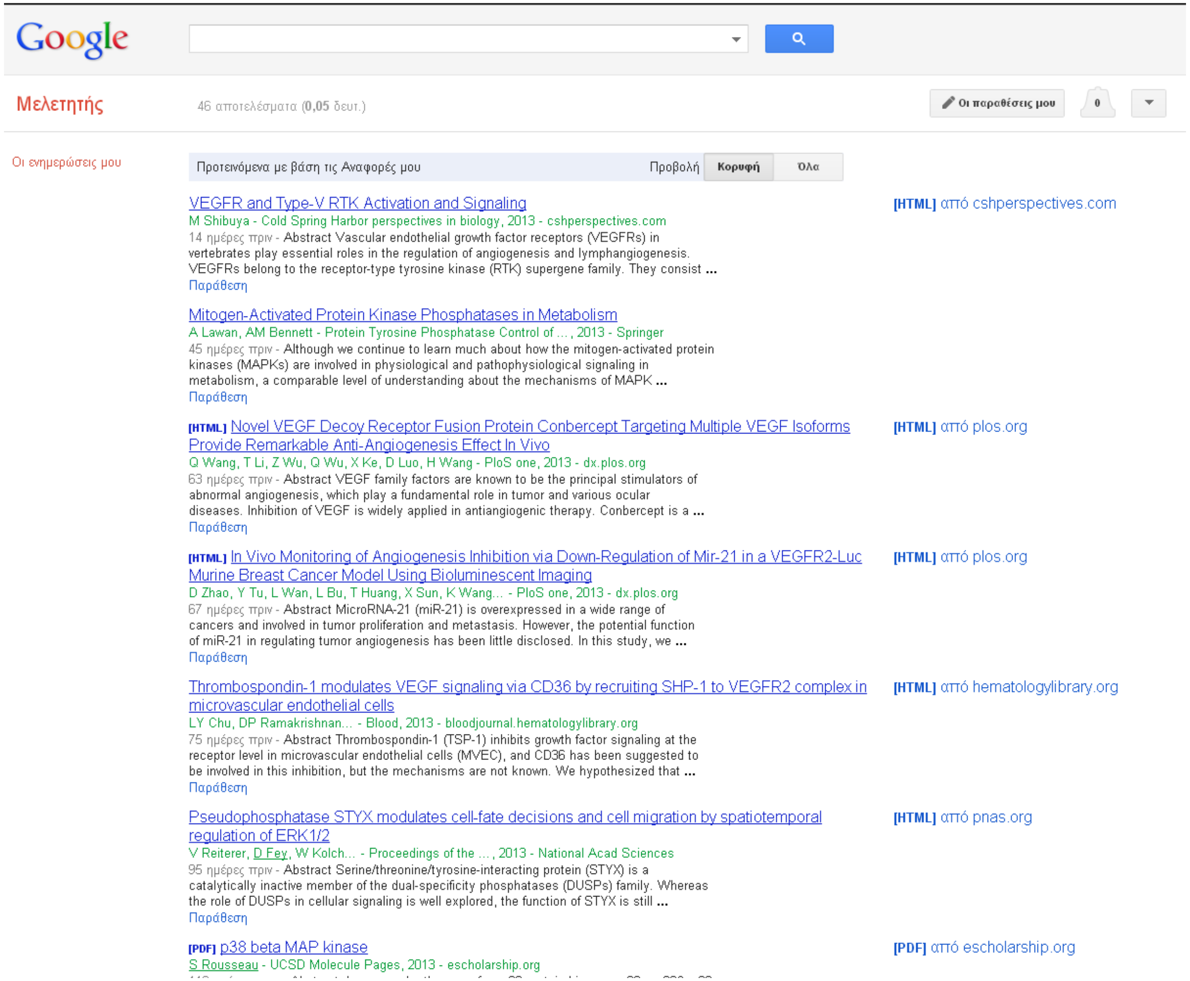
Οι ημερομηνίες και οι αριθμοί των παραθέσεων υπολογίζονται και καθορίζονται από το υπολογιστικό πρόγραμμα.

©2013 Google - Πληροφορίες για τον Μελετητή Google - Σχετικά με τη Google - Αποστολή σχολίων - Οι παραθέσεις μου

Google (5/6)


Οι παραθέσεις μου
Οι ενημερώσεις μου
Προειδοποιήσεις
Μετρήσεις
Ρυθμίσεις
Google
scholar

Google (6/6)



46 αποτελέσματα (**0,05** δευτ.)

Προτεινόμενα με βάση τις Αναφορές μου

VEGFR and Type-V RTK Activation and Signaling — **[HTML]** από cshperspectives.com
M Shibuya - Cold Spring Harbor perspectives in biology, 2013 - cshperspectives.com
14 ημέρες πριν - Abstract Vascular endothelial growth factor receptors (VEGFRs) in vertebrates play essential roles in the regulation of angiogenesis and lymphangiogenesis. VEGFRs belong to the receptor-type tyrosine kinase (RTK) supergene family. They consist **...**
Παράθεση

Mitogen-Activated Protein Kinase Phosphatases in Metabolism
A Lawan, AM Bennett - Protein Tyrosine Phosphatase Control of ..., 2013 - Springer
45 ημέρες πριν - Although we continue to learn much about how the mitogen-activated protein kinases (MAPKs) are involved in physiological and pathophysiological signaling in metabolism, a comparable level of understanding about the mechanisms of MAPK **...**
Παράθεση

[HTML] Novel VEGF Decoy Receptor Fusion Protein Conbercept Targeting Multiple VEGF Isoforms Provide Remarkable Anti-Angiogenesis Effect In Vivo — **[HTML]** από plos.org
Q Wang, T Li, Z Wu, Q Wu, X Ke, D Luo, H Wang - PloS one, 2013 - dx.plos.org
63 ημέρες πριν - Abstract VEGF family factors are known to be the principal stimulators of abnormal angiogenesis, which play a fundamental role in tumor and various ocular diseases. Inhibition of VEGF is widely applied in antiangiogenic therapy. Conbercept is a **...**
Παράθεση

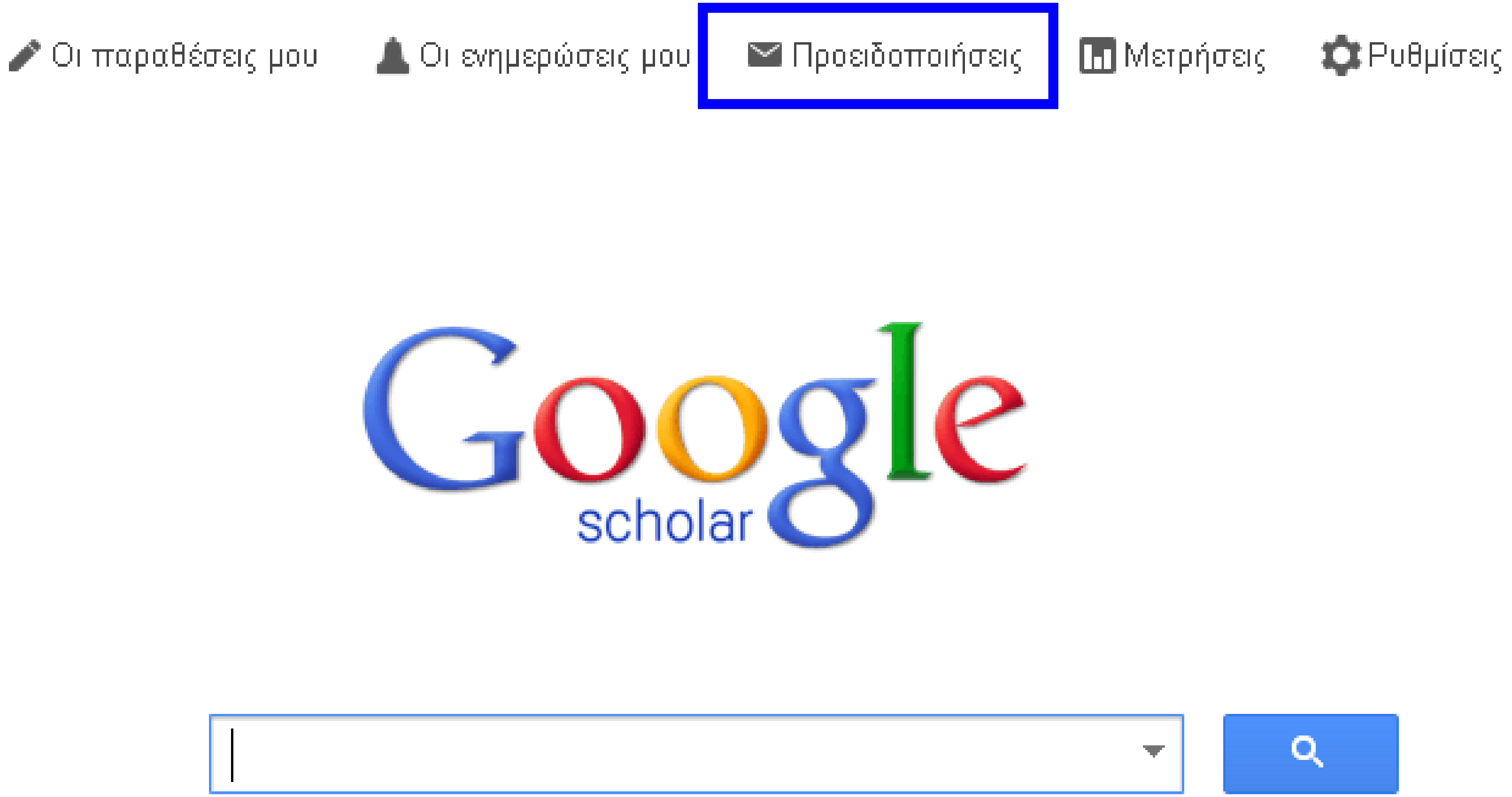
[HTML] In Vivo Monitoring of Angiogenesis Inhibition via Down-Regulation of Mir-21 in a VEGFR2-Luc Murine Breast Cancer Model Using Bioluminescent Imaging — **[HTML]** από plos.org
D Zhao, Y Tu, L Wan, L Bu, T Huang, X Sun, K Wang... - PloS one, 2013 - dx.plos.org
67 ημέρες πριν - Abstract MicroRNA-21 (miR-21) is overexpressed in a wide range of cancers and involved in tumor proliferation and metastasis. However, the potential function of miR-21 in regulating tumor angiogenesis has been little disclosed. In this study, we **...**
Παράθεση

Thrombospondin-1 modulates VEGF signaling via CD36 by recruiting SHP-1 to VEGFR2 complex in microvascular endothelial cells — **[HTML]** από hematologylibrary.org
LY Chu, DP Ramakrishnan... - Blood, 2013 - bloodjournal.hematologylibrary.org
75 ημέρες πριν - Abstract Thrombospondin-1 (TSP-1) inhibits growth factor signaling at the receptor level in microvascular endothelial cells (MVEC), and CD36 has been suggested to be involved in this inhibition, but the mechanisms are not known. We hypothesized that **...**
Παράθεση

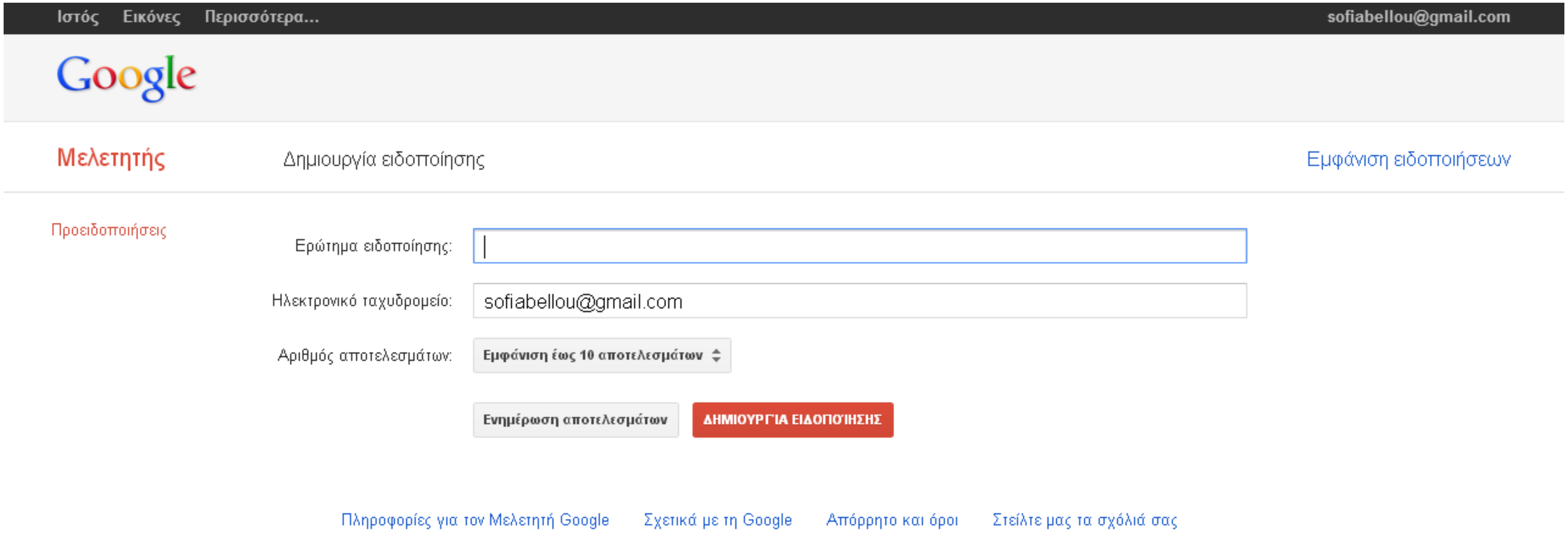
Pseudophosphatase STYX modulates cell-fate decisions and cell migration by spatiotemporal regulation of ERK1/2 — **[HTML]** από pnas.org
V Reiterer, D Fey, W Kolch... - Proceedings of the ..., 2013 - National Acad Sciences
95 ημέρες πριν - Abstract Serine/threonine/tyrosine-interacting protein (STYX) is a catalytically inactive member of the dual-specificity phosphatases (DUSPs) family. Whereas the role of DUSPs in cellular signaling is well explored, the function of STYX is still **...**
Παράθεση

[PDF] p38 beta MAP kinase — **[PDF]** από escholarship.org
S Rousseau - UCSD Molecule Pages, 2013 - escholarship.org

Google (7/8)


Οι παραθέσεις μου
Οι ενημερώσεις μου
Προειδοποιήσεις
Μετρήσεις
Ρυθμίσεις
Google
scholar

Google (8/8)



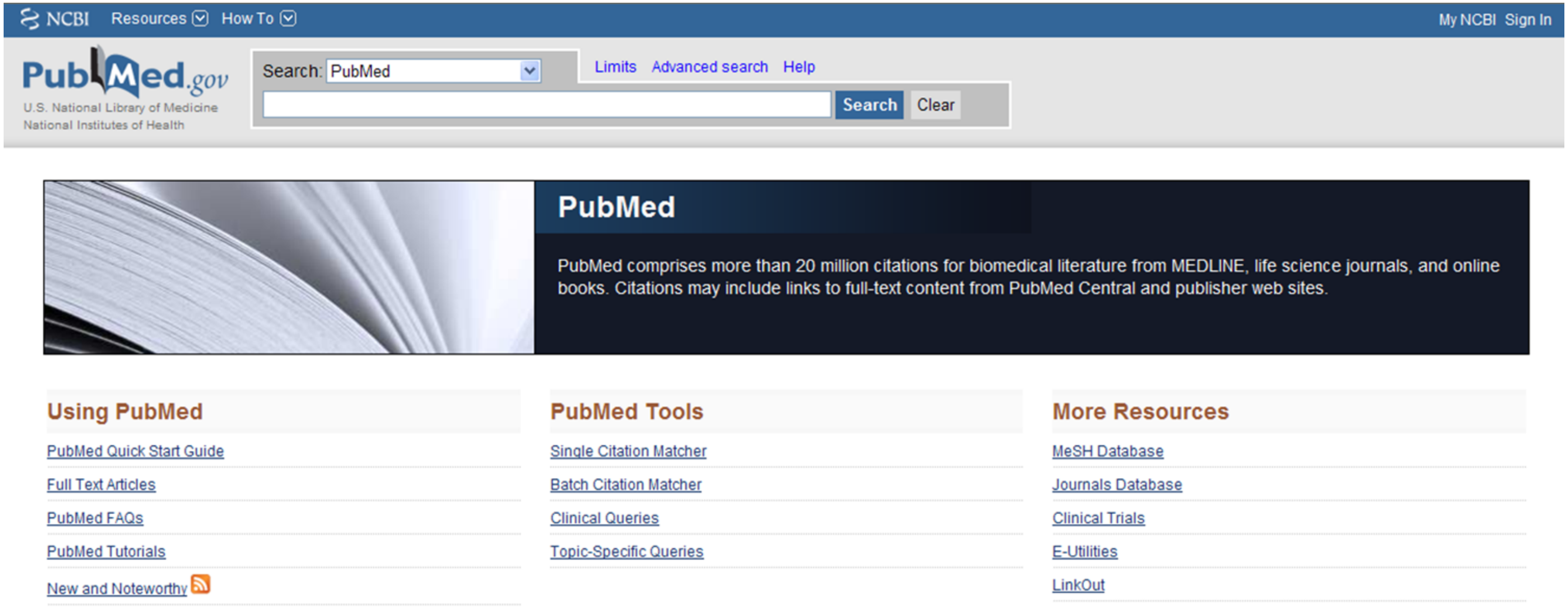
Ερώτημα ειδοποίησης:

Ηλεκτρονικό ταχυδρομείο: sofiabellou@gmail.com

Αριθμός αποτελεσμάτων: Εμφάνιση έως 10 αποτελεσμάτων

PubMed – NCBI

(National Center for Biotechnology Information)



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

PubMed

PubMed comprises more than 20 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

Using PubMed

- PubMed Quick Start Guide
- Full Text Articles
- PubMed FAQs
- PubMed Tutorials
- New and Noteworthy

PubMed Tools

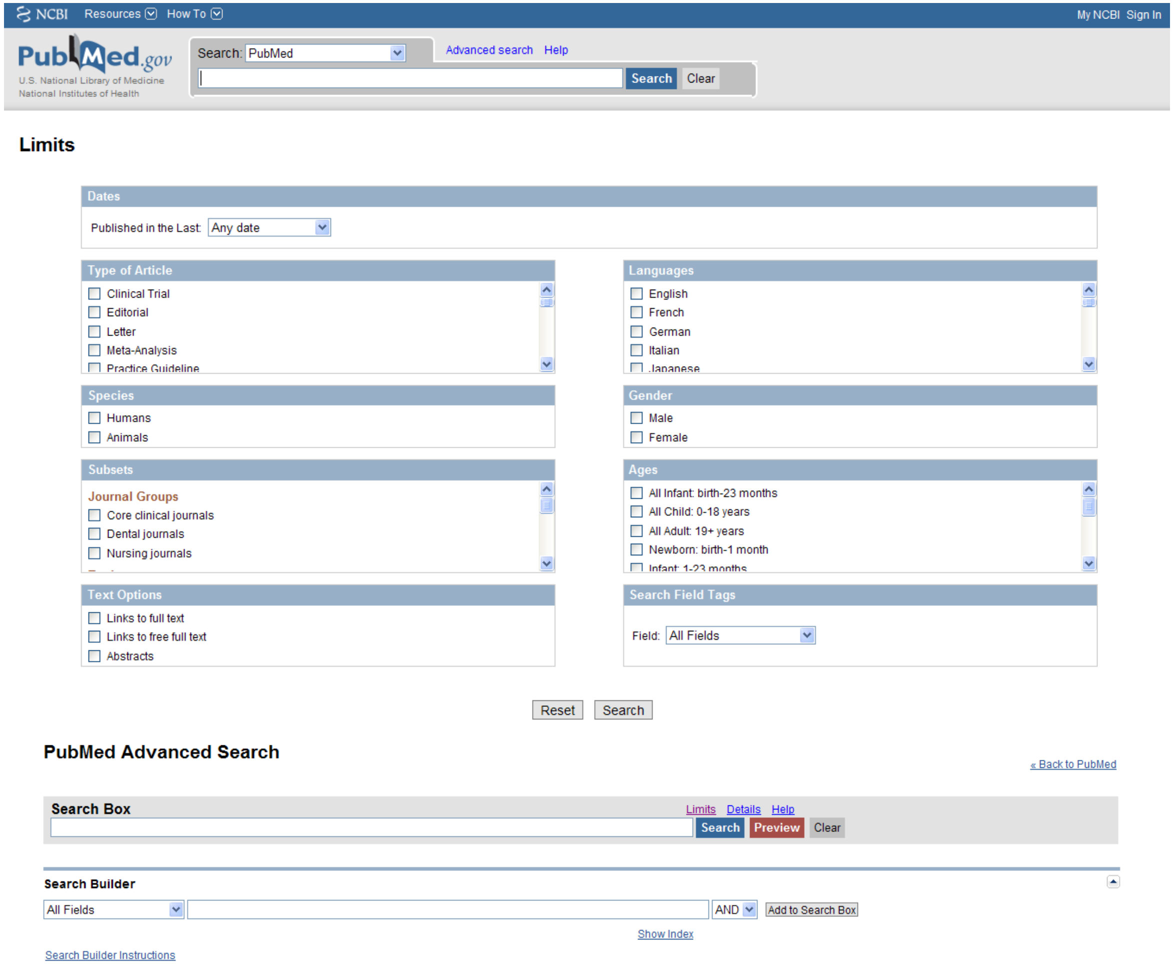
- Single Citation Matcher
- Batch Citation Matcher
- Clinical Queries
- Topic-Specific Queries

More Resources

- MeSH Database
- Journals Database
- Clinical Trials
- E-Utilities
- LinkOut

http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed

PubMed



Search: PubMed

Limits

Dates

Published in the Last: Any date

Type of Article

☐ Clinical Trial
☐ Editorial
☐ Letter
☐ Meta-Analysis
☐ Practice Guideline

Languages

☐ English
☐ French
☐ German
☐ Italian
☐ Japanese

Species

☐ Humans
☐ Animals

Gender

☐ Male
☐ Female

Subsets

Journal Groups
☐ Core clinical journals
☐ Dental journals
☐ Nursing journals

Ages

☐ All Infant: birth-23 months
☐ All Child: 0-18 years
☐ All Adult: 19+ years
☐ Newborn: birth-1 month
☐ Infant: 1-23 months

Text Options

☐ Links to full text
☐ Links to free full text
☐ Abstracts

Search Field Tags

Field: All Fields

PubMed Advanced Search

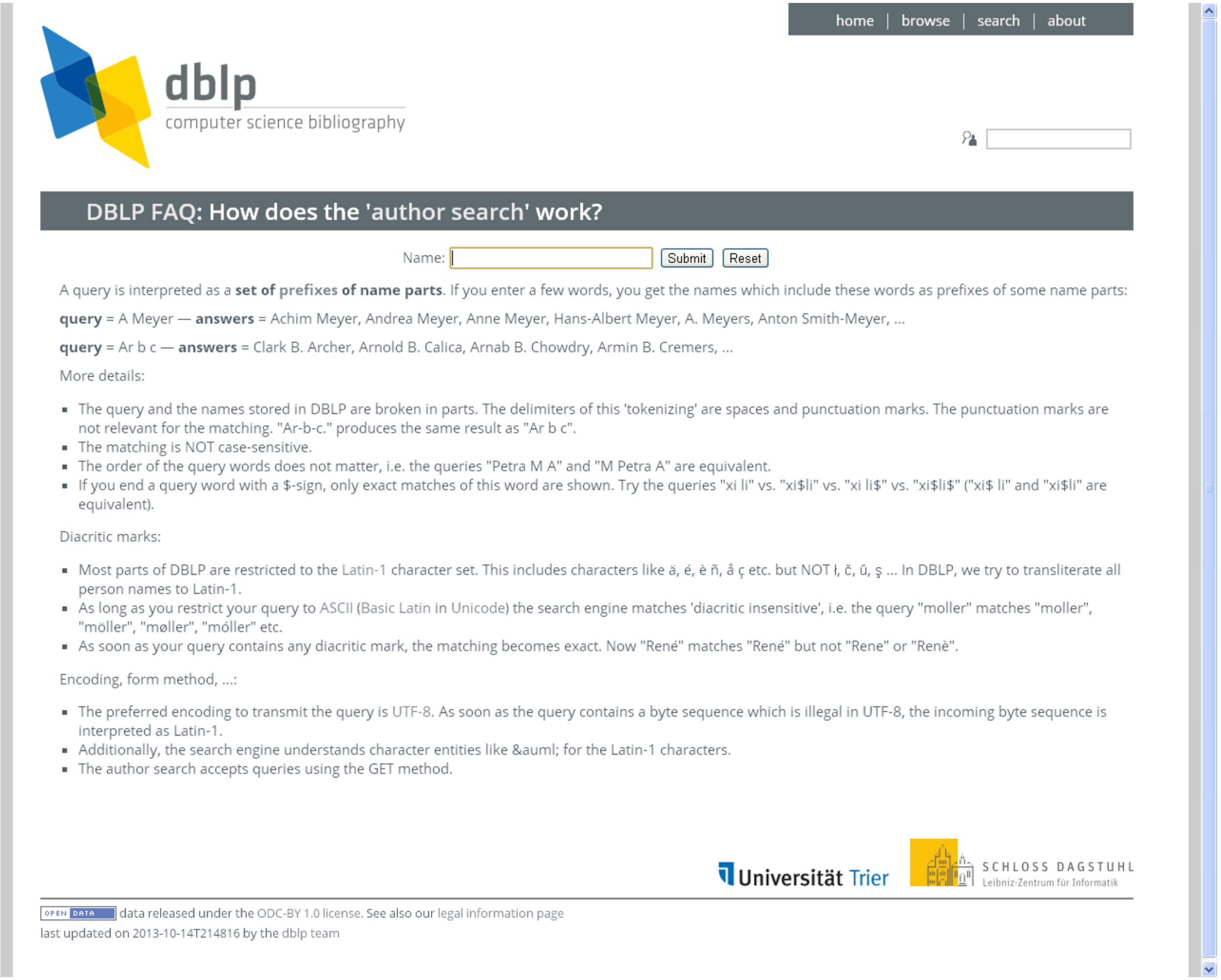
Search Box

Search Builder

All Fields
AND

DBLP Database

DBLP Database
DataBase systems and Logic Programming
or
Digital Bibliography & Library Project.



DBLP FAQ: How does the 'author search' work?

Name: [ ] Submit Reset

A query is interpreted as a **set of prefixes of name parts**. If you enter a few words, you get the names which include these words as prefixes of some name parts:

query = A Meyer — **answers** = Achim Meyer, Andrea Meyer, Anne Meyer, Hans-Albert Meyer, A. Meyers, Anton Smith-Meyer, ...

query = Ar b c — **answers** = Clark B. Archer, Arnold B. Calica, Arnab B. Chowdry, Armin B. Cremers, ...

More details:

- The query and the names stored in DBLP are broken in parts. The delimiters of this 'tokenizing' are spaces and punctuation marks. The punctuation marks are not relevant for the matching. "Ar-b-c." produces the same result as "Ar b c".
- The matching is NOT case-sensitive.
- The order of the query words does not matter, i.e. the queries "Petra M A" and "M Petra A" are equivalent.
- If you end a query word with a $-sign, only exact matches of this word are shown. Try the queries "xi li" vs. "xi$li" vs. "xi li$" vs. "xili" ("xi$ li" and "xi$li" are equivalent).

Diacritic marks:

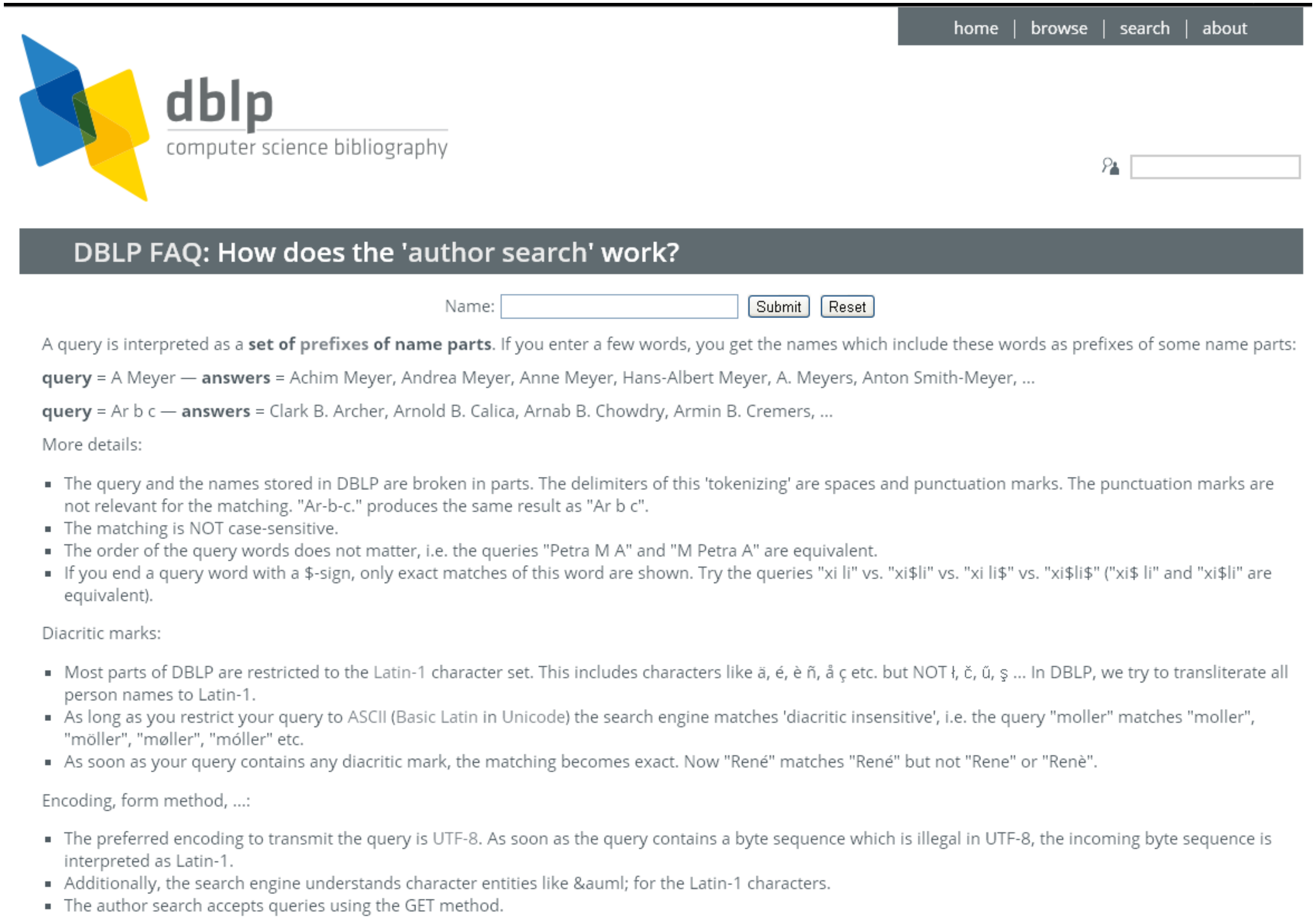
- Most parts of DBLP are restricted to the Latin-1 character set. This includes characters like ä, é, è ñ, å ç etc. but NOT ł, č, ű, ş ... In DBLP, we try to transliterate all person names to Latin-1.
- As long as you restrict your query to ASCII (Basic Latin in Unicode) the search engine matches 'diacritic insensitive', i.e. the query "moller" matches "moller", "möller", "møller", "móller" etc.
- As soon as your query contains any diacritic mark, the matching becomes exact. Now "René" matches "René" but not "Rene" or "Renè".

Encoding, form method, ...:

- The preferred encoding to transmit the query is UTF-8. As soon as the query contains a byte sequence which is illegal in UTF-8, the incoming byte sequence is interpreted as Latin-1.
- Additionally, the search engine understands character entities like ä for the Latin-1 characters.
- The author search accepts queries using the GET method.

last updated on 2013-10-14T214816 by the dblp team

DBLP



DBLP FAQ: How does the 'author search' work?

A query is interpreted as a **set of prefixes of name parts**. If you enter a few words, you get the names which include these words as prefixes of some name parts:

query = A Meyer — **answers** = Achim Meyer, Andrea Meyer, Anne Meyer, Hans-Albert Meyer, A. Meyers, Anton Smith-Meyer, ...

query = Ar b c — **answers** = Clark B. Archer, Arnold B. Calica, Arnab B. Chowdry, Armin B. Cremers, ...

More details:

- The query and the names stored in DBLP are broken in parts. The delimiters of this 'tokenizing' are spaces and punctuation marks. The punctuation marks are not relevant for the matching. "Ar-b-c." produces the same result as "Ar b c".
- The matching is NOT case-sensitive.
- The order of the query words does not matter, i.e. the queries "Petra M A" and "M Petra A" are equivalent.
- If you end a query word with a $-sign, only exact matches of this word are shown. Try the queries "xi li" vs. "xi$li" vs. "xi li$" vs. "xili" ("xi$ li" and "xi$li" are equivalent).

Diacritic marks:

- Most parts of DBLP are restricted to the Latin-1 character set. This includes characters like ä, é, è ñ, å ç etc. but NOT ł, č, ű, ş ... In DBLP, we try to transliterate all person names to Latin-1.
- As long as you restrict your query to ASCII (Basic Latin in Unicode) the search engine matches 'diacritic insensitive', i.e. the query "moller" matches "moller", "möller", "møller", "móller" etc.
- As soon as your query contains any diacritic mark, the matching becomes exact. Now "René" matches "René" but not "Rene" or "Renè".

Encoding, form method, ...:

- The preferred encoding to transmit the query is UTF-8. As soon as the query contains a byte sequence which is illegal in UTF-8, the incoming byte sequence is interpreted as Latin-1.
- Additionally, the search engine understands character entities like ä for the Latin-1 characters.
- The author search accepts queries using the GET method.

Πρωτογενείς βάσεις δεδομένων DNA & πρωτεϊνών

NCBI

EMBL-EBI

DDBJ

DNA Data Bank of Japan

Πρωτογενείς βάσεις δεδομένων DNA (Primary sequence database)

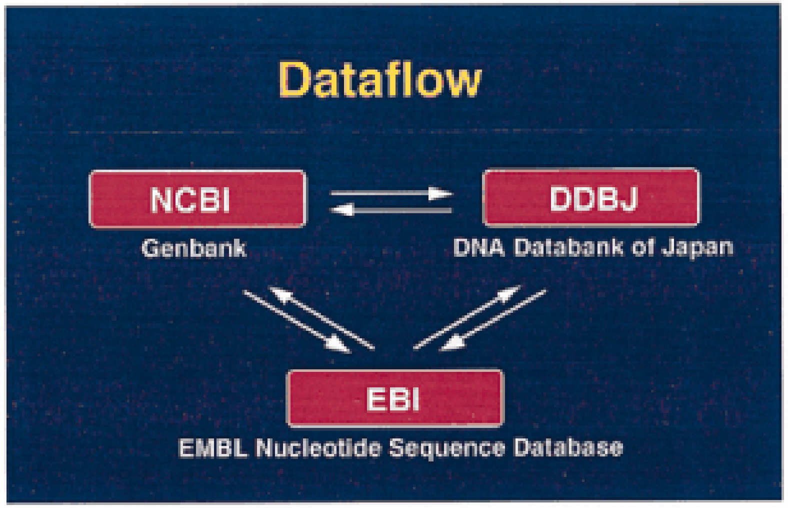
1. NCBI - National Centre of Biotechnology information (U.S.A.).
2. EMBL - European Molecular Biology Laboratory (Europe).
3. DDBJ - DNA Data Banf of Japan (Japan).

Σκοπός: Συλλογή και διάθεση στην επιστημονική κοινότητα ακολουθιών DNA και RNA από πλήρως ή μερικώς αποκρυπτογραφημένα γονιδιώματα οργανισμών, θέτοντας κοινούς κανόνες ταξινόμησης και σχολιασμού των δεδομένων.

Τα δεδομένα υποβάλλονται από διάφορες ανά τον κόσμο επιστημονικές ομάδες σε μία από τις τρεις βάσεις δεδομένων και **μέσω της καθημερινής ανταλλαγής στοιχείων** ενημερώνονται και οι υπόλοιπες.

INSDC

- International Nucleotide Sequence Database Collaboration.
- Made up of:
  - GenBank,
  - European Nucleotide Archive, and
  - DNA Data Bank of Japan.
- Represents an archival repository of all sequences.



Identifiers and Accession numbers

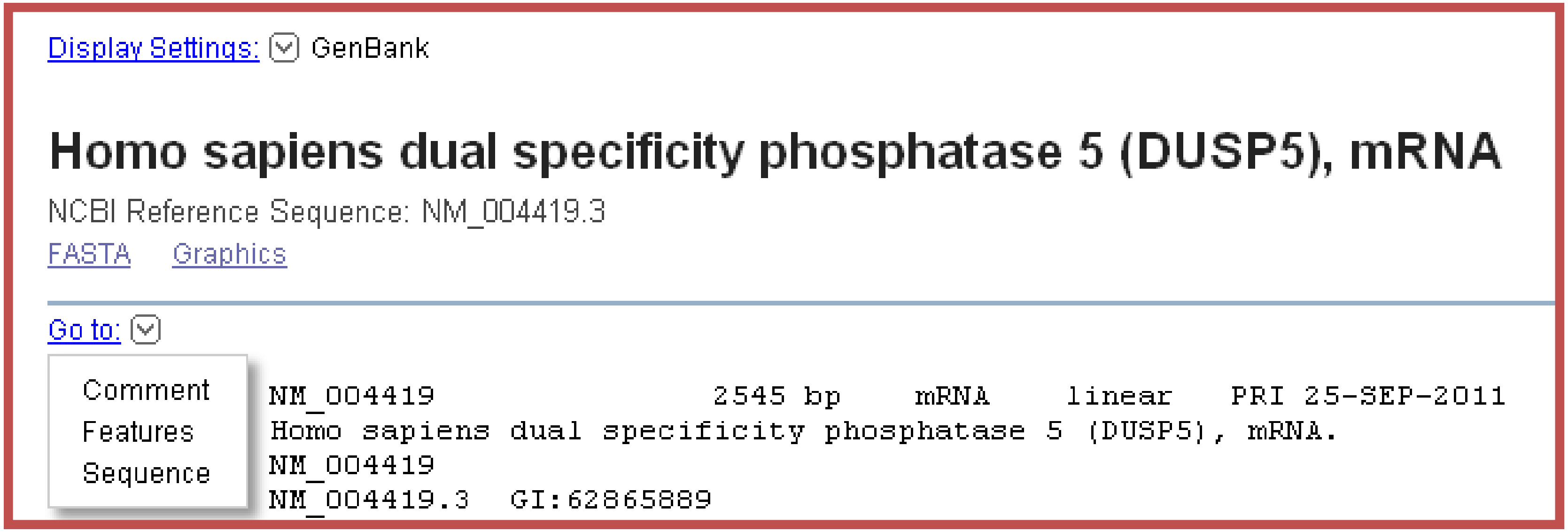
- **Identifier:** Συμβολοσειρά από γράμματα και ψηφία που συνήθως είναι «δυσνόητα».
  - Example: TPIS_CHICK (Triose Phosphate Isomerase from chicken (gallus gallus)) in protein database.
  - The identifier can change (based on the curator).
- **Accession code:** Συμβολοσειρά από γράμματα και ψηφία που αναγνωρίζει μοναδικά την καταχώρηση στη βάση δεδομένων.
  - The accession number for TPIS_CHICK is P00940.
  - Accession number should not changed!!

NCBI - GENBANK

- Υπό την αιγίδα του Εθνικού Ινστιτούτου Υγείας (National Institutes of Health - NIH).
- Διαθέτει νουκλεοτιδικές αλληλουχίες από περισσότερους από 300,000 οργανισμούς.
- Δεδομένα:
  - Απευθείας υποβολές των αποτελεσμάτων ερευνών διαφόρων μικρών ερευνητικών ομάδων, ή
  - Υποβολές εύρεσης της αλληλουχίας του γονιδιώματος διαφόρων οργανισμών από μεγάλα ερευνητικά κέντρα.
- Η τελευταία έκδοση της GENBANK περιλαμβάνει:
  - 126,551,501,141 βάσεις DNA.
  - 135,440,924 εγγραφές.

Format αρχείου GenBank

- Αποτελείται από 3 μέρη:
    1. Σχόλια (Comment).
    2. Χαρακτηριστικά (Features).
    3. Ακολουθία (Sequence).



Homo sapiens dual specificity phosphatase 5 (DUSP5), mRNA

NCBI Reference Sequence: NM_004419.3

```
NM_004419                    2545 bp    mRNA    linear   PRI 25-SEP-2011
Homo sapiens dual specificity phosphatase 5 (DUSP5), mRNA.
NM_004419
NM_004419.3   GI:62865889
```

GENBANK Format – DNA (1/2)

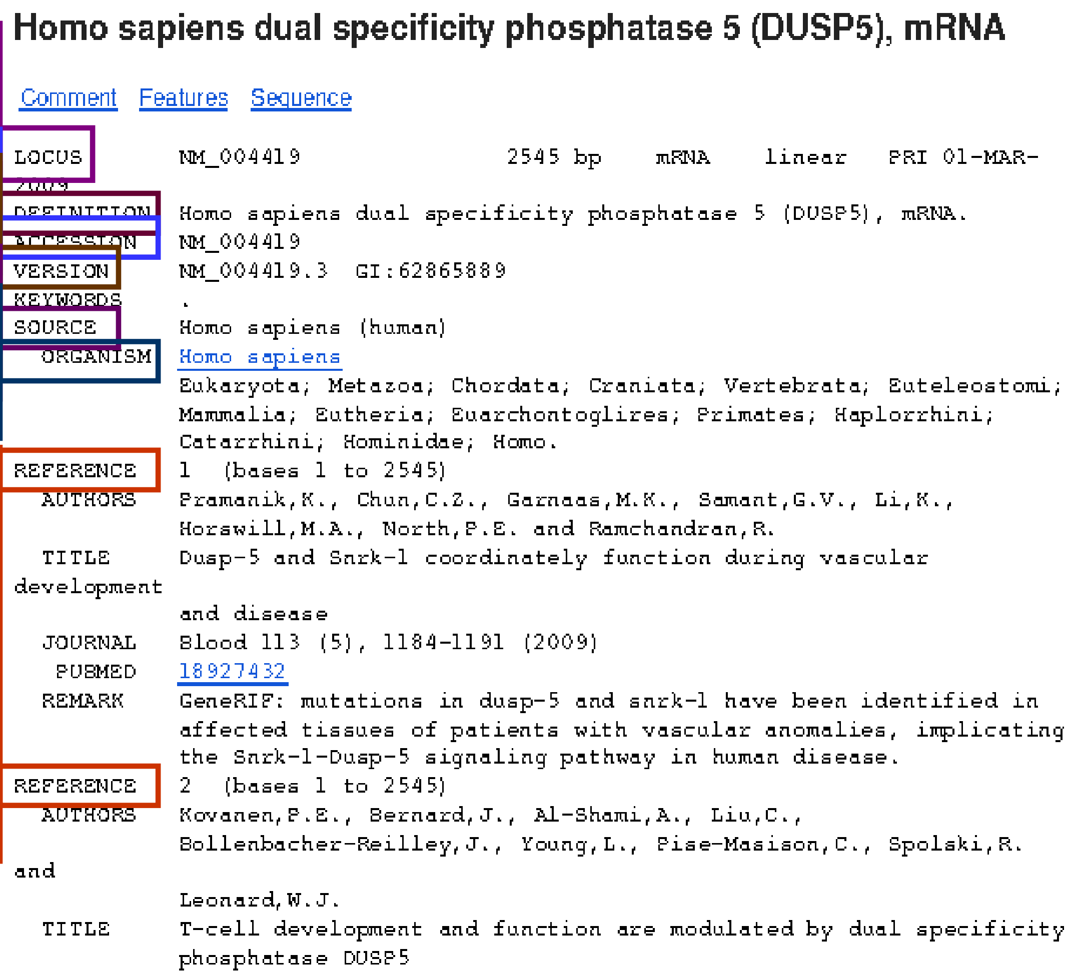
<u>Locus:</u> Η γραμμή περιλαμβάνει **αριθμό υποβολής εγγραφής (Accession**

<u>Version:</u> Ο **αριθμός υποβολής** της εγγραφής της ακολουθίας και η

<u>Source:</u> Το κοινό **όνομα** του

<u>Organism:</u> Το **επίσημο επιστημονικό όνομα** του οργανισμού και η συστηματική αναγνώρισή του.

<u>Reference:</u> Παράθεση των **άρθρων** που περιέχουν δεδομένα για την εγγραφή, με αναφορά στον **συγγραφέα**, τον **τίτλο του άρθρου**, τα στοιχεία του **περιοδικού** στο οποίο δημοσιεύτηκε, το **μοναδικό αριθμό καταχώρησης** του άρθρου στις βιβλιογραφικές βάσεις δεδομένων MEDLINE & PUBMED.



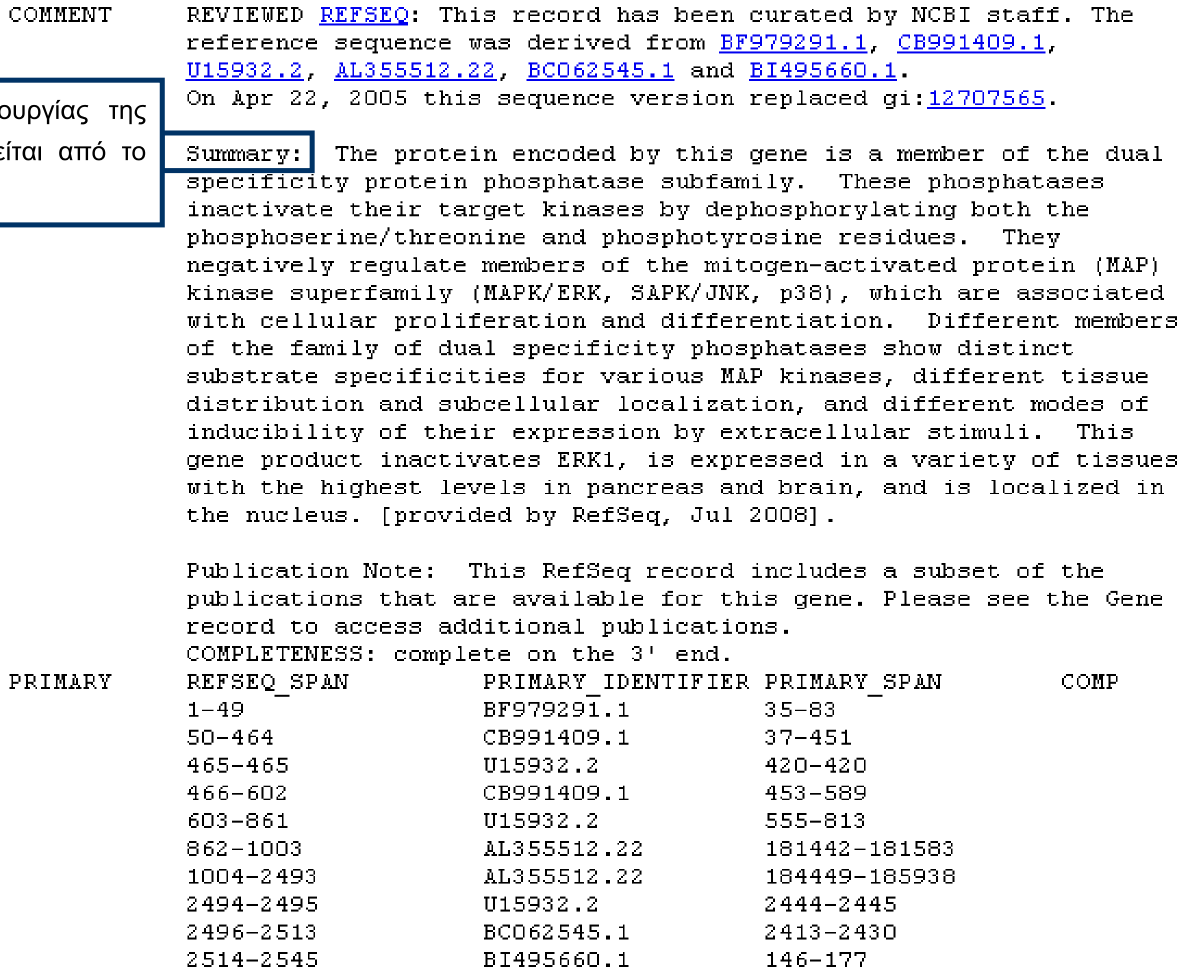
Homo sapiens dual specificity phosphatase 5 (DUSP5), mRNA

```
LOCUS       NM_004419              2545 bp    mRNA    linear   PRI 01-MAR-
2009
DEFINITION  Homo sapiens dual specificity phosphatase 5 (DUSP5), mRNA.
ACCESSION   NM_004419
VERSION     NM_004419.3  GI:62865889
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2545)
  AUTHORS   Pramanik,K., Chun,C.Z., Garnaas,M.K., Samant,G.V., Li,K.,
            Horswill,M.A., North,P.E. and Ramchandran,R.
  TITLE     Dusp-5 and Snrk-1 coordinately function during vascular
development

            and disease
  JOURNAL   Blood 113 (5), 1184-1191 (2009)
   PUBMED   18927432
  REMARK    GeneRIF: mutations in dusp-5 and snrk-1 have been identified in
            affected tissues of patients with vascular anomalies, implicating
            the Snrk-1-Dusp-5 signaling pathway in human disease.
REFERENCE   2  (bases 1 to 2545)
  AUTHORS   Kovanen,P.E., Bernard,J., Al-Shami,A., Liu,C.,
            Bollenbacher-Reilley,J., Young,L., Pise-Masison,C., Spolski,R.
and

            Leonard,W.J.
  TITLE     T-cell development and function are modulated by dual specificity
            phosphatase DUSP5
```

Πανεπιστήμιο Δυτικής Μακεδονίας

GeneBank - Comment

Summary: Περίληψη λειτουργίας της πρωτεΐνης που κωδικοποιείται από το γονίδιο



```
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from BF979291.1, CB991409.1,
            U15932.2, AL355512.22, BC062545.1 and BI495660.1.
            On Apr 22, 2005 this sequence version replaced gi:12707565.

            Summary: The protein encoded by this gene is a member of the dual
            specificity protein phosphatase subfamily.  These phosphatases
            inactivate their target kinases by dephosphorylating both the
            phosphoserine/threonine and phosphotyrosine residues.  They
            negatively regulate members of the mitogen-activated protein (MAP)
            kinase superfamily (MAPK/ERK, SAPK/JNK, p38), which are associated
            with cellular proliferation and differentiation.  Different members
            of the family of dual specificity phosphatases show distinct
            substrate specificities for various MAP kinases, different tissue
            distribution and subcellular localization, and different modes of
            inducibility of their expression by extracellular stimuli.  This
            gene product inactivates ERK1, is expressed in a variety of tissues
            with the highest levels in pancreas and brain, and is localized in
            the nucleus. [provided by RefSeq, Jul 2008].

            Publication Note:  This RefSeq record includes a subset of the
            publications that are available for this gene. Please see the Gene
            record to access additional publications.
            COMPLETENESS: complete on the 3' end.
PRIMARY     REFSEQ_SPAN         PRIMARY_IDENTIFIER PRIMARY_SPAN        COMP
            1-49                BF979291.1         35-83
            50-464              CB991409.1         37-451
            465-465             U15932.2           420-420
            466-602             CB991409.1         453-589
            603-861             U15932.2           555-813
            862-1003            AL355512.22        181442-181583
            1004-2493           AL355512.22        184449-185938
            2494-2495           U15932.2           2444-2445
            2496-2513           BC062545.1         2413-2430
            2514-2545           BI495660.1         146-177
```

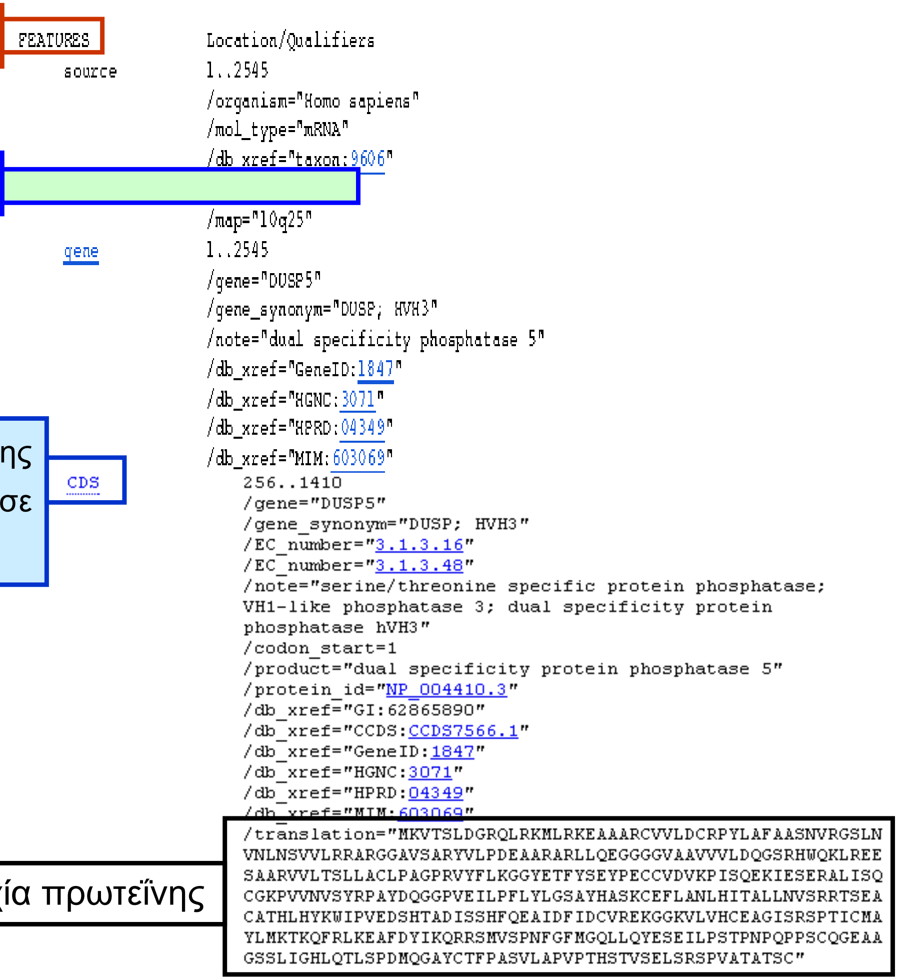
GENBANK - Features

Χαρακτηριστικά

Χρωμόσωμα

Coding sequence: Αρχή της μεταγραφής του γονιδίου σε πρωτεΐνη – ATG: start codon

Αλληλουχία πρωτεΐνης



```
FEATURES             Location/Qualifiers
     source          1..2545
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /map="10q25"
     gene            1..2545
                     /gene="DUSP5"
                     /gene_synonym="DUSP; HVH3"
                     /note="dual specificity phosphatase 5"
                     /db_xref="GeneID:1847"
                     /db_xref="HGNC:3071"
                     /db_xref="HPRD:04349"
                     /db_xref="MIM:603069"
     CDS                256..1410
                        /gene="DUSP5"
                        /gene_synonym="DUSP; HVH3"
                        /EC_number="3.1.3.16"
                        /EC_number="3.1.3.48"
                        /note="serine/threonine specific protein phosphatase;
                        VH1-like phosphatase 3; dual specificity protein
                        phosphatase hVH3"
                        /codon_start=1
                        /product="dual specificity protein phosphatase 5"
                        /protein_id="NP_004410.3"
                        /db_xref="GI:62865890"
                        /db_xref="CCDS:CCDS7566.1"
                        /db_xref="GeneID:1847"
                        /db_xref="HGNC:3071"
                        /db_xref="HPRD:04349"
                        /db_xref="MIM:603069"
                        /translation="MKVTSLDGRQLRKMLRKEAAARCVVLDCRPYLAFAASNVRGSLN
                        VNLNSVVLRRARGGAVSARYVLPDEAARARLLQEGGGGVAAVVVLDQGSRHWQKLREE
                        SAARVVLTSLLACLPAGPRVYFLKGGYETFYSEYPECCVDVKPISQEKIESERALISQ
                        CGKPVVNVSYRPAYDQGGPVEILPFLYLGSAYHASKCEFLANLHITALLNVSRRTSEA
                        CATHLHYKWIPVEDSHTADISSHFQEAIDFIDCVREKGGKVLVHCEAGISRSPTICMA
                        YLMKTKQFRLKEAFDYIKQRRSMVSPNFGFMGQLLQYESEILPSTPNPQPPSCQGEAA
                        GSSLIGHLQTLSPDMQGAYCTFPASVLAPVPTHSTVSELSRSPVATATSC"
```

GENBANK Format – DNA (2/2)

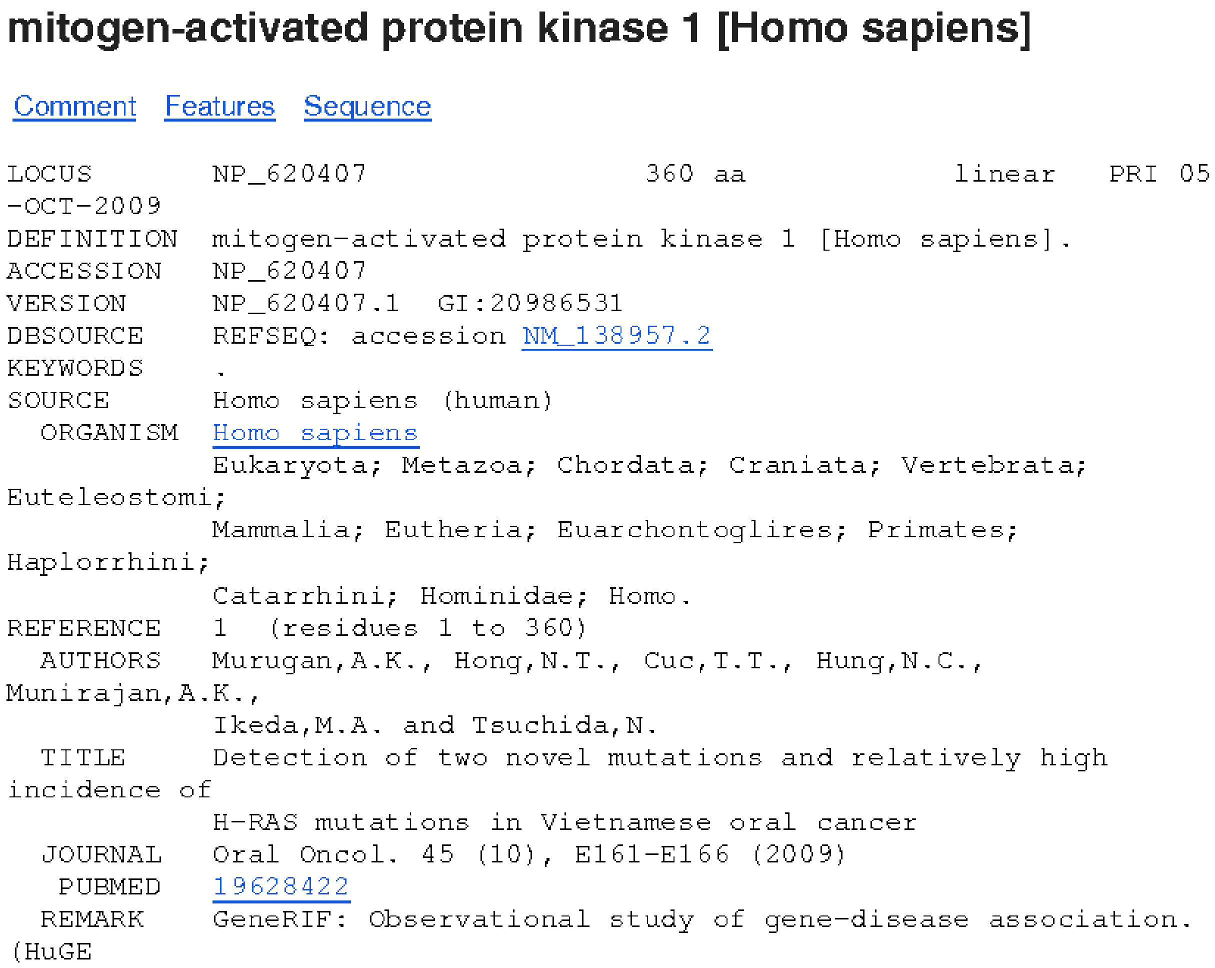
Αλληλουχία DNA

Start codon (256): ATG

Stop codon (1410): GAT

```
ORIGIN
        1 actcattcac ataaaacgct gcgcggccgg cggaatcccc ggcttctagg gcggcgagcg
       61 gccgggctgg ctatcgagcg agcggggcgg gaacgcggag ttgcgccgcc gctcgggcgc
      121 cgggctccgt cgcggccgca gccccgcggg tcgccctccc gtgcctcgcc cgcggacacc
      181 ctggccgtgg acaccctggc cgtgggcacc cgcggggcgc gcggcgcggg gccgctggcc
      241 ggcggcggcg gcggcatgaa ggtcacgtcg ctcgacgggc gccagctgcg caagatgctc
      301 cgcaaggagg cggcggcgcg ctgcgtggtg ctcgactgcc ggccctatct ggccttcgct
      361 gcctcgaacg tgcgcggctc gctcaacgtc aacctcaact cggtggtgct gcggcgggcc
      421 cggggcggcg cggtgtcggc gcgctacgtg ctgcccgacg aggcggcgcg cgcgcggctc
      481 ctgcaggagg gcggcggcgg cgtcgcggcc gtggtggtgc tggaccaggg cagccgccac
      541 tggcagaagc tgcgagagga gagcgccgcg cgtgtcgtcc tcacctcgct actcgcttgc
      601 ctacccgccg gcccgcgggt ctacttcctc aaagggggat atgagacttt ctactcggaa
      661 tatcctgagt gttgcgtgga tgtaaaaccc atttcacaag agaagattga gagtgagaga
      721 gccctcatca gccagtgtgg aaaaccagtg gtaaatgtca gctacaggcc agcttatgac
      781 cagggtggcc cagttgaaat ccttcccttc ctctaccttg gaagtgccta ccatgcatcc
      841 aagtgcgagt tcctcgccaa cctgcacatc acagccctgc tgaatgtctc ccgacggacc
      901 tccgaggcct gcgcgaccca cctacactac aaatggatcc ctgtggaaga cagccacacg
      961 gctgacatta gctcccactt tcaagaagca atagacttca ttgactgtgt cagggaaaag
     1021 ggaggcaagg tcctggtcca ctgtgaggct gggatctccc gttcacccac catctgcatg
     1081 gcttacctta tgaagaccaa gcagttccgc ctgaaggagg ccttcgatta catcaagcag
     1141 aggaggagca tggtctcgcc caactttggc ttcatgggcc agctcctgca gtacgaatct
     1201 gagatcctgc cctccacgcc caacccccag cctccctcct gccaagggga ggcagcaggc
     1261 tcttcactga taggccattt gcagacactg agccctgaca tgcagggtgc ctactgcaca
     1321 ttccctgcct cggtgctggc accggtgcct acccactcaa cagtctcaga gctcagcaga
     1381 agccctgtgg caacggccac atcctgctaa aactgggatg gaggaatcgg cccagcccca
     1441 agagcaactg tgatttttgt ttttaagact catggacatt tcatacctgt gcaatactga
     1501 agacctcatt ctgtcatgct gccccagtga gatagtgagt ggtcaccagg cttgcaaatg
     1561 aacttcagac ggacctcagg gtaggttctc gggactgaag gaaggccaag ccattacggg
     1621 agcacagcat gtgctgacta ctgtacttcc agacccctgc cctcttggga ctgcccagtc
     1681 cttgcacctc agagttcgcc ttttcatttc aagcataagg caataaatac ctgcagcaac
     1741 gtgggagaaa gaagttgctg gaccaggaga aaaggcagtt atgaagccaa ttcattttga
     1801 aggaagcaca atttccacct tatttttga actttggcag tttcaatgtc tgtctctgtt
     1861 gcttcggggc ataagctgat caccgtctag ttgggaaagt aaccctacag ggtttgtagg
     1921 gacatgatca gcatcctgat ttgaaccctg aaatgttgtg tagacaccct cttgggtcca
     1981 atgaggtagt tggttgaagt agcaagatgt tggcttttct ggattttttt tgccatgggt
     2041 tcttcactga ccttggactt tggcatgatt cttagtcata cttgaacttg tctcattcca
     2101 cctcttctca gagcaactct tcctttggga aaagagttct tcagatcata gaccaaaaaa
     2161 gtcatacctt cgaggtggta gcagtagatt ccaggaggag aagggtactt gctaggtatc
     2221 ctgggtcagt ggcggtgcaa actggtttcc tcagctgcct gtccttctgt gtgcttatgt
     2281 ctcttgtgac aattgttttc ctccctgccc ctggaggttg tcttcaagct gtggacttct
     2341 gggatttgca gattttgcaa cgtggtacta cttttttttt ctttttgtct gttagttatt
     2401 tctccagggg aaaaggcaat aattttctaa gacccgtgtg aatgtgaaga aaagcagtat
     2461 gttactggtt gttgttgttg ttcttgtttt ttatagtgta aaataaaaat agtaaaagga
     2521 gaaaagcaaa aaaaaaaaaa aaaaa
//
```

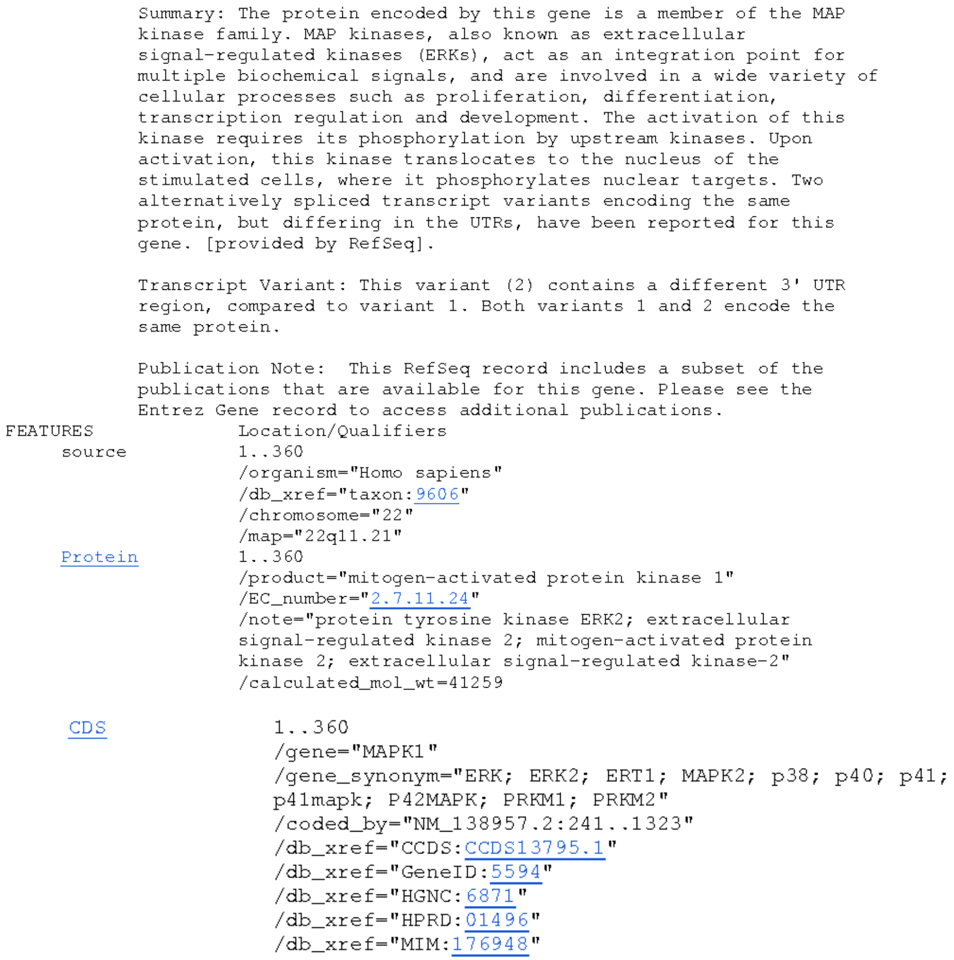
Παράδειγμα GENBANK – Πρωτεΐνη (1/3)



mitogen-activated protein kinase 1 [Homo sapiens]

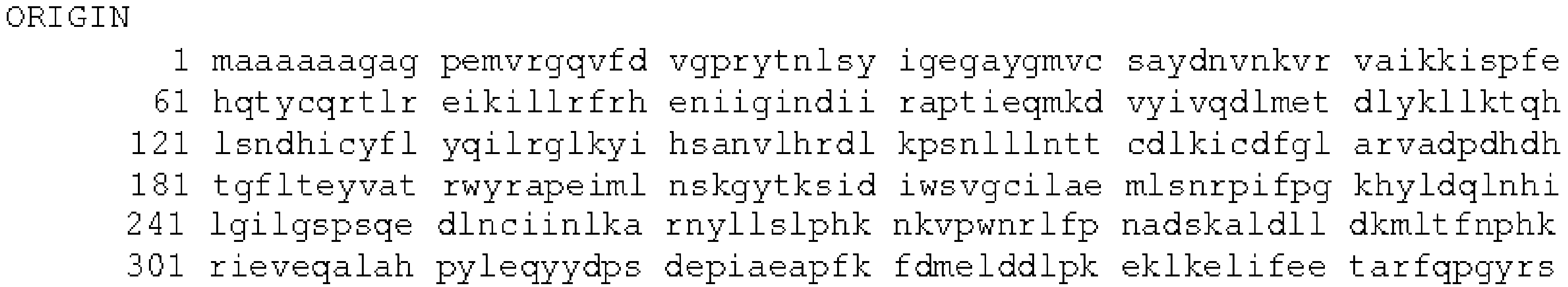
```
LOCUS       NP_620407                360 aa            linear   PRI 05
-OCT-2009
DEFINITION  mitogen-activated protein kinase 1 [Homo sapiens].
ACCESSION   NP_620407
VERSION     NP_620407.1  GI:20986531
DBSOURCE    REFSEQ: accession NM_138957.2
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates;
Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 360)
  AUTHORS   Murugan,A.K., Hong,N.T., Cuc,T.T., Hung,N.C.,
Munirajan,A.K.,
            Ikeda,M.A. and Tsuchida,N.
  TITLE     Detection of two novel mutations and relatively high
incidence of
            H-RAS mutations in Vietnamese oral cancer
  JOURNAL   Oral Oncol. 45 (10), E161-E166 (2009)
   PUBMED   19628422
  REMARK    GeneRIF: Observational study of gene-disease association.
(HuGE
```

Παράδειγμα GENBANK – Πρωτεΐνη (2/3)



```
                    Summary: The protein encoded by this gene is a member of the MAP
                    kinase family. MAP kinases, also known as extracellular
                    signal-regulated kinases (ERKs), act as an integration point for
                    multiple biochemical signals, and are involved in a wide variety of
                    cellular processes such as proliferation, differentiation,
                    transcription regulation and development. The activation of this
                    kinase requires its phosphorylation by upstream kinases. Upon
                    activation, this kinase translocates to the nucleus of the
                    stimulated cells, where it phosphorylates nuclear targets. Two
                    alternatively spliced transcript variants encoding the same
                    protein, but differing in the UTRs, have been reported for this
                    gene. [provided by RefSeq].

                    Transcript Variant: This variant (2) contains a different 3' UTR
                    region, compared to variant 1. Both variants 1 and 2 encode the
                    same protein.

                    Publication Note:  This RefSeq record includes a subset of the
                    publications that are available for this gene. Please see the
                    Entrez Gene record to access additional publications.
FEATURES             Location/Qualifiers
     source          1..360
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.21"
     Protein         1..360
                     /product="mitogen-activated protein kinase 1"
                     /EC_number="2.7.11.24"
                     /note="protein tyrosine kinase ERK2; extracellular
                     signal-regulated kinase 2; mitogen-activated protein
                     kinase 2; extracellular signal-regulated kinase-2"
                     /calculated_mol_wt=41259

     CDS             1..360
                     /gene="MAPK1"
                     /gene_synonym="ERK; ERK2; ERT1; MAPK2; p38; p40; p41;
                     p41mapk; P42MAPK; PRKM1; PRKM2"
                     /coded_by="NM_138957.2:241..1323"
                     /db_xref="CCDS:CCDS13795.1"
                     /db_xref="GeneID:5594"
                     /db_xref="HGNC:6871"
                     /db_xref="HPRD:01496"
                     /db_xref="MIM:176948"
```

Παράδειγμα GENBANK – Πρωτεΐνη (3/3)

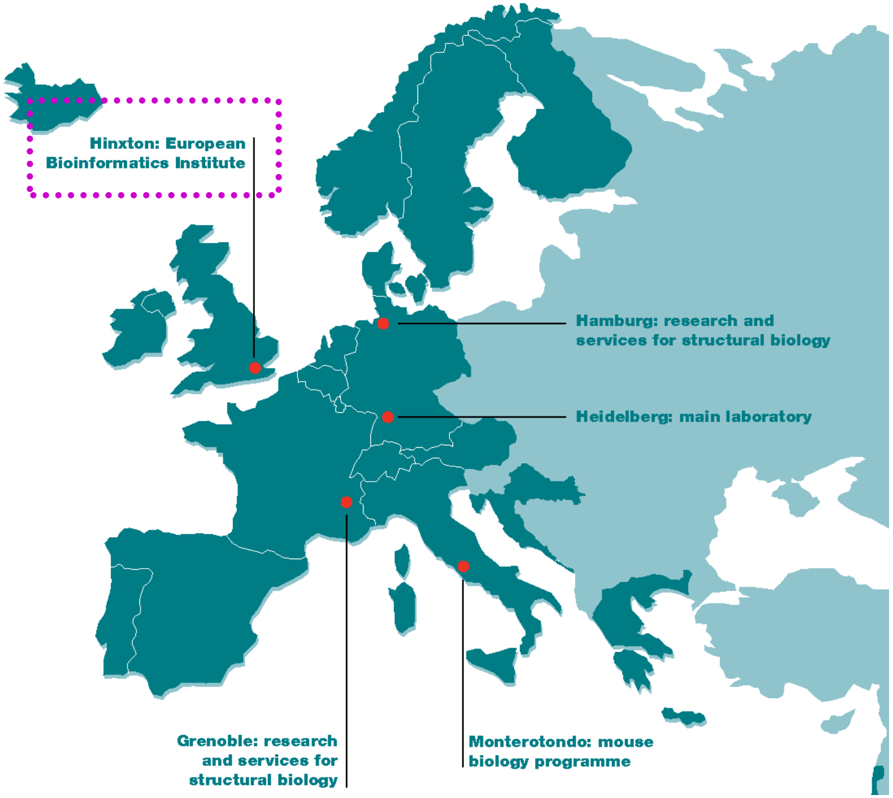


```
ORIGIN
        1 maaaaaagag pemvrgqvfd vgprytnlsy igegaygmvc saydnvnkvr vaikkispfe
       61 hqtycqrtlr eikillrfrh eniigindii raptieqmkd vyivqdlmet dlykllktqh
      121 lsndhicyfl yqilrglkyi hsanvlhrdl kpsnlllntt cdlkicdfgl arvadpdhdh
      181 tgflteyvat rwyrapeiml nskgytksid iwsvgcilae mlsnrpifpg khyldqlnhi
      241 lgilgspsqe dlnciinlka rnyllslphk nkvpwnrlfp nadskaldll dkmltfnphk
      301 rieveqalah pyleqyydps depiaeapfk fdmelddlpk eklkelifee tarfqpgyrs
```

EMBL (European Molecular Biology Laboratory) Nucleotide Sequence Database (also known as EMBL-Bank)

- **EMBL:** Ευρωπαϊκός οργανισμός έρευνας με το κύριο ερευνητικό του κέντρο στην Χαϊδελβέργη – Γερμανία
- 1 κεντρικό ερευνητικό κέντρο και 4 περιφερειακά
- EMBL-Bank: Εδρεύει στο EBI

- Λειτουργεί κατά αντιστοιχία με την GENBANK.
- Η τελευταία έκδοση της EMBL-Bank περιλαμβάνει:
  - 283,748,816,763 βάσεις DNA
  - 163,656,234 εγγραφές



EMBL – EBI / Ευρωπαϊκό Ινστιτούτο Βιοπληροφορικής

- **EMBL-Bank:** Νουκλεοτιδική (DNA) βάση δεδομένων.
- **UniProt:** Πρωτεϊνική βάση δεδομένων.
- Η κάθε γραμμή αρχίζει με συγκεκριμένο αναγνωριστικό 2 χαρακτήρων.

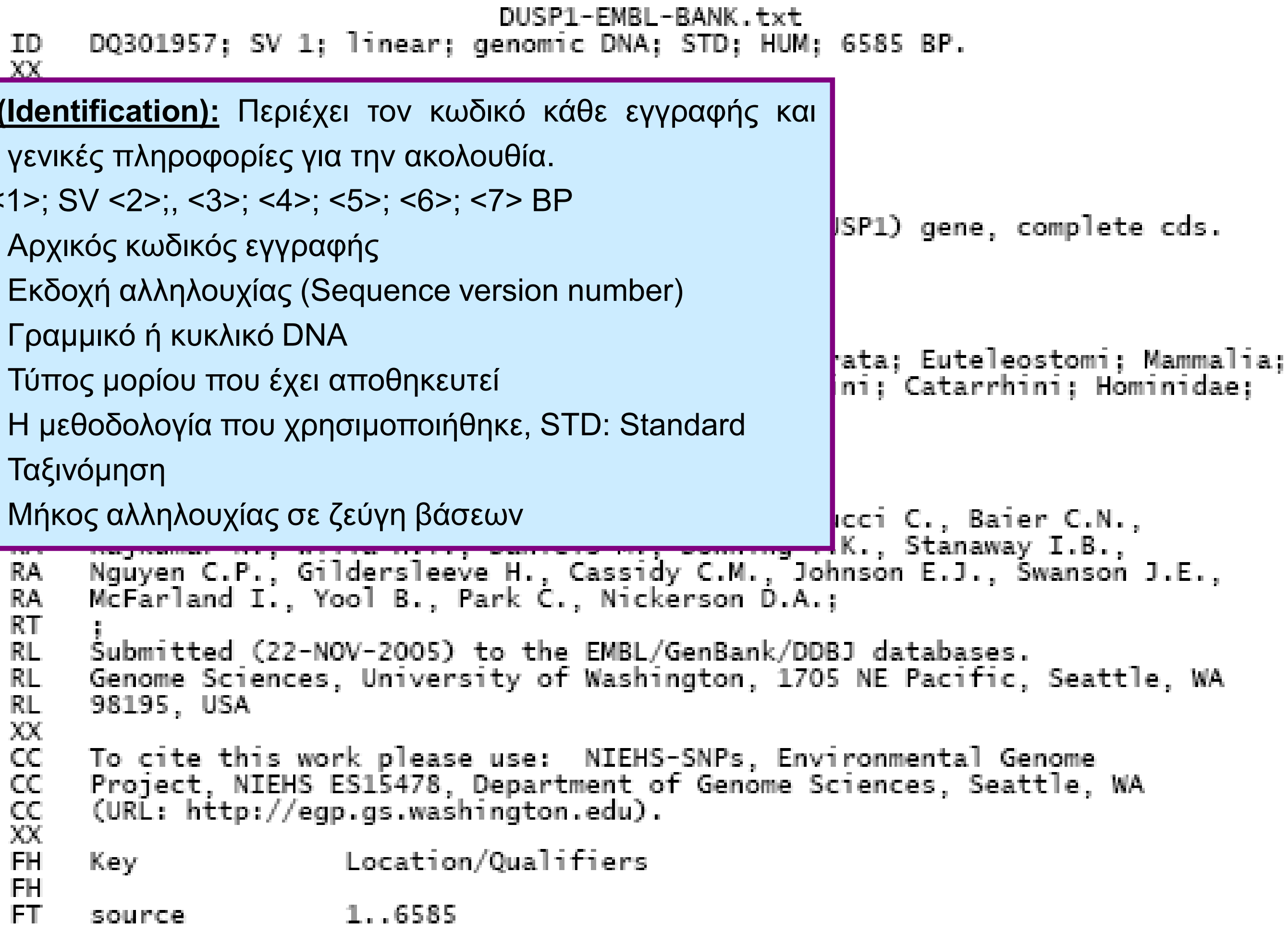
EMBL Format - DNA

- Δομή γραμμών που η καθεμία αρχίζει με ένα συγκεκριμένο αναγνωριστικό που αποτελείται από 2 χαρακτήρες.

Abbreviations
Συντομογραφίες

ID - identification
AC - accession number
PR - project identifier
DT - date
DE - description
KW - keyword
OS - organism species
OC - organism classification
OG - organelle
RN - reference number
RC - reference comment
RP - reference positions
RX - reference cross-reference
RG - reference group
RA - reference author(s)
RT - reference title
RL - reference location
DR - database cross-reference
CC - comments or notes
AH - assembly header
AS - assembly information
FH - feature table header
FT - feature table data
XX - spacer line
SQ - sequence header
CO - contig/construct line
bb - (blanks) sequence data
// - termination line

EMBL - Bank format (DNA database) (1/4)



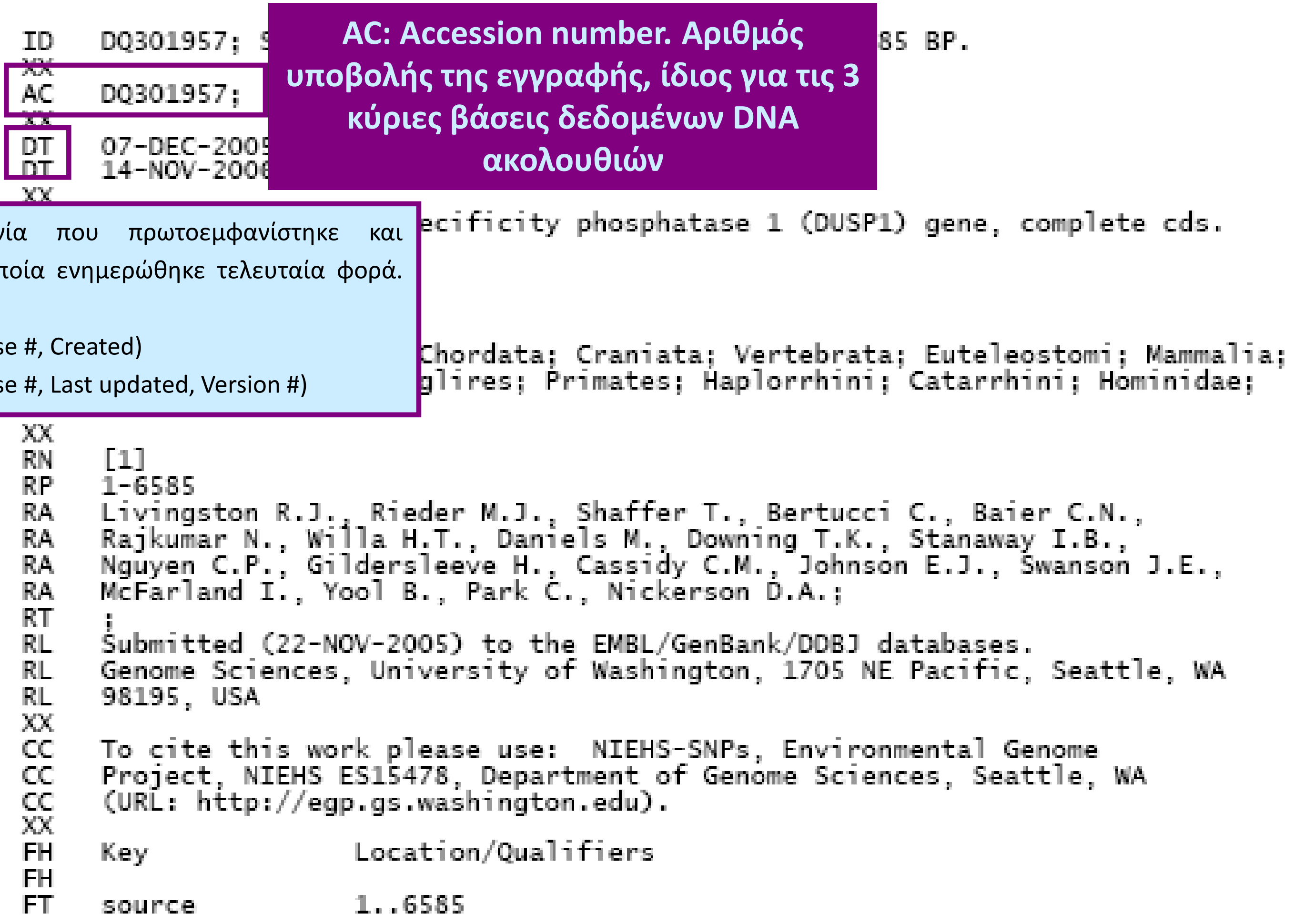
```
                         DUSP1-EMBL-BANK.txt
ID   DQ301957; SV 1; linear; genomic DNA; STD; HUM; 6585 BP.
XX

                                      USP1) gene, complete cds.

                                      rata; Euteleostomi; Mammalia;
                                      ini; Catarrhini; Hominidae;

                                      ucci C., Baier C.N.,
                                      K., Stanaway I.B.,
RA   Nguyen C.P., Gildersleeve H., Cassidy C.M., Johnson E.J., Swanson J.E.,
RA   McFarland I., Yool B., Park C., Nickerson D.A.;
RT   ;
RL   Submitted (22-NOV-2005) to the EMBL/GenBank/DDBJ databases.
RL   Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA
RL   98195, USA
XX
CC   To cite this work please use:  NIEHS-SNPs, Environmental Genome
CC   Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA
CC   (URL: http://egp.gs.washington.edu).
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..6585
```

ID (Identification): Περιέχει τον κωδικό κάθε εγγραφής και γενικές πληροφορίες για την ακολουθία.

ID <1>; SV <2>;, <3>; <4>; <5>; <6>; <7> BP

1. Αρχικός κωδικός εγγραφής
2. Εκδοχή αλληλουχίας (Sequence version number)
3. Γραμμικό ή κυκλικό DNA
4. Τύπος μορίου που έχει αποθηκευτεί
5. Η μεθοδολογία που χρησιμοποιήθηκε, STD: Standard
6. Ταξινόμηση
7. Μήκος αλληλουχίας σε ζεύγη βάσεων

EMBL - Bank format (DNA database) (2/4)

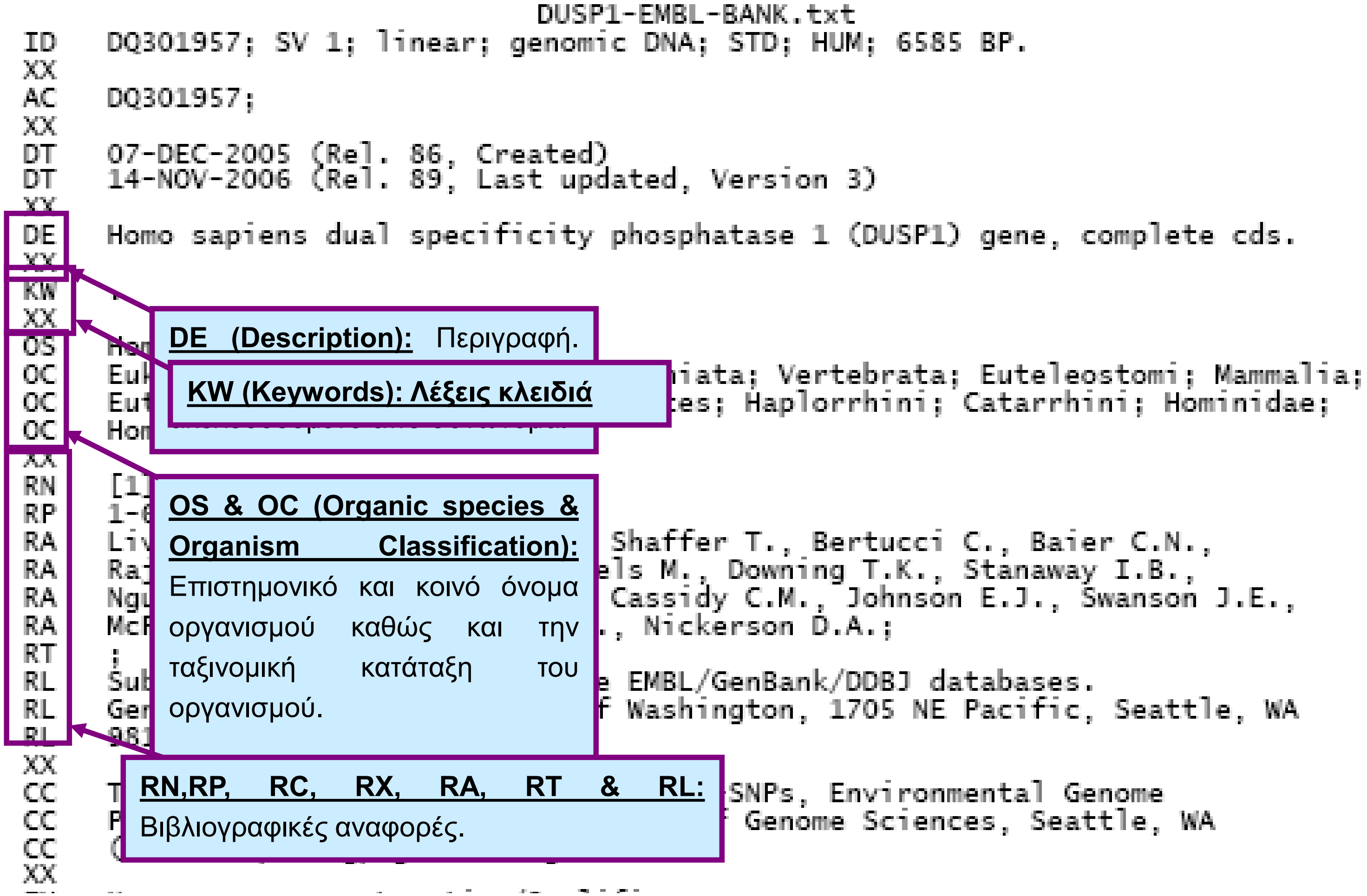


```
ID   DQ301957; S                                               85 BP.
XX
AC   DQ301957;
XX
DT   07-DEC-200
DT   14-NOV-200
XX
                   ecificity phosphatase 1 (DUSP1) gene, complete cds.



                   Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
                   glires; Primates; Haplorrhini; Catarrhini; Hominidae;

XX
RN   [1]
RP   1-6585
RA   Livingston R.J., Rieder M.J., Shaffer T., Bertucci C., Baier C.N.,
RA   Rajkumar N., Willa H.T., Daniels M., Downing T.K., Stanaway I.B.,
RA   Nguyen C.P., Gildersleeve H., Cassidy C.M., Johnson E.J., Swanson J.E.,
RA   McFarland I., Yool B., Park C., Nickerson D.A.;
RT   ;
RL   Submitted (22-NOV-2005) to the EMBL/GenBank/DDBJ databases.
RL   Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA
RL   98195, USA
XX
CC   To cite this work please use:  NIEHS-SNPs, Environmental Genome
CC   Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA
CC   (URL: http://egp.gs.washington.edu).
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..6585
```

AC: Accession number. Αριθμός υποβολής της εγγραφής, ίδιος για τις 3 κύριες βάσεις δεδομένων DNA ακολουθιών

DT (DaTe): Ημερομηνία που πρωτοεμφανίστηκε και ημερομηνία κατά την οποία ενημερώθηκε τελευταία φορά.
Μορφή καταχώρησης:
DT DD-MON-YYYY (Release #, Created)
DT DD-MON-YYYY (Release #, Last updated, Version #)

EMBL - Bank format (DNA database) (3/4)



DUSP1-EMBL-BANK.txt

```
ID   DQ301957; SV 1; linear; genomic DNA; STD; HUM; 6585 BP.
XX
AC   DQ301957;
XX
DT   07-DEC-2005 (Rel. 86, Created)
DT   14-NOV-2006 (Rel. 89, Last updated, Version 3)
XX
DE   Homo sapiens dual specificity phosphatase 1 (DUSP1) gene, complete cds.
XX
KW   .
XX
OS   Hom
OC   Euk                                  iata; Vertebrata; Euteleostomi; Mammalia;
OC   Eut                                 tes; Haplorrhini; Catarrhini; Hominidae;
OC   Hom
XX
RN   [1]
RP   1-6
RA   Liv                         Shaffer T., Bertucci C., Baier C.N.,
RA   Raj                        els M., Downing T.K., Stanaway I.B.,
RA   Ngu                        Cassidy C.M., Johnson E.J., Swanson J.E.,
RA   McP                        ., Nickerson D.A.;
RT   ;
RL   Sub                        e EMBL/GenBank/DDBJ databases.
RL   Ger                        f Washington, 1705 NE Pacific, Seattle, WA
RL   981
XX
CC   T                                SNPs, Environmental Genome
CC   P                               f Genome Sciences, Seattle, WA
CC   (
XX
```

DE (Description): Περιγραφή.

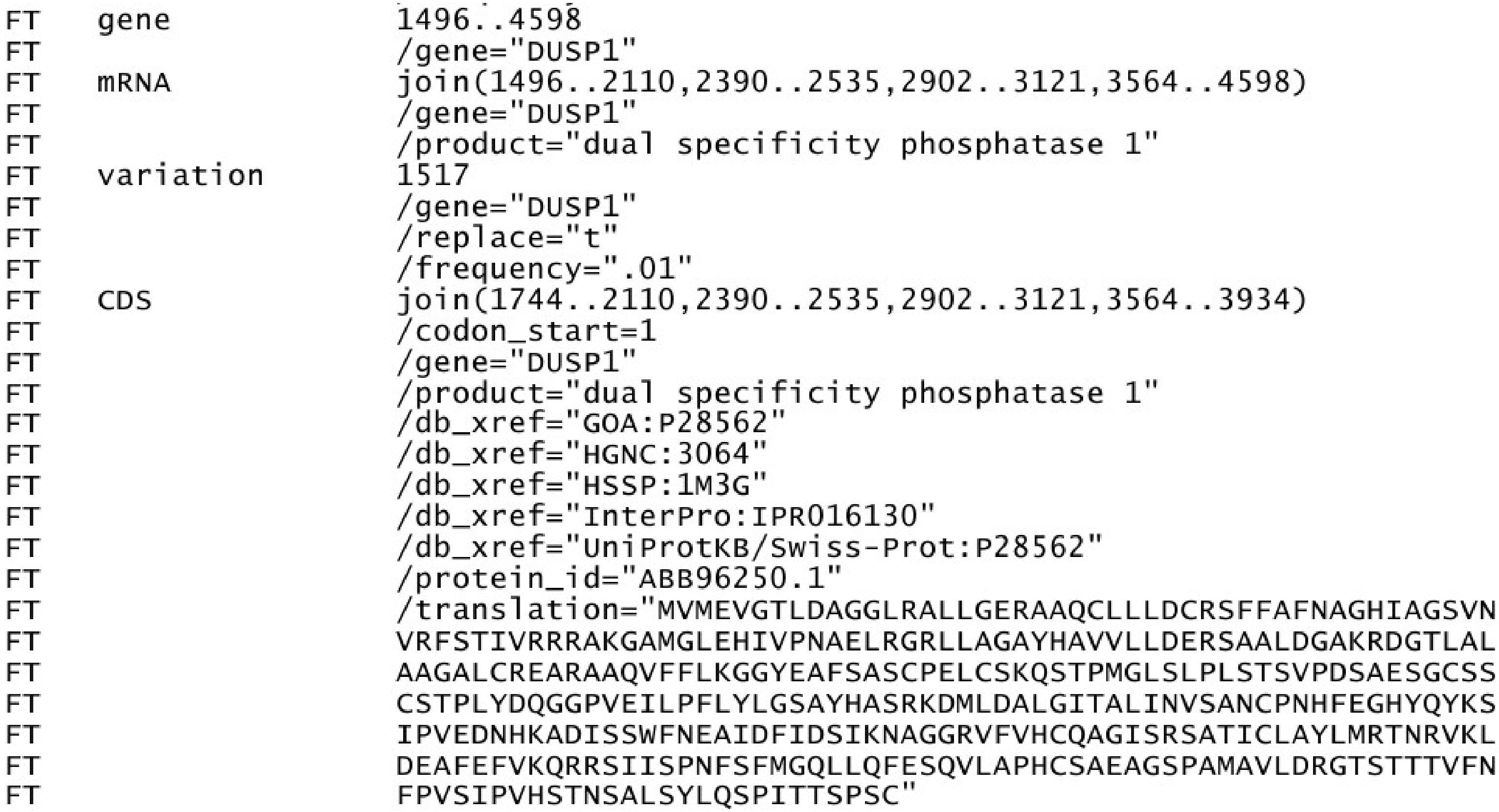
KW (Keywords): Λέξεις κλειδιά

OS & OC (Organic species & Organism Classification): Επιστημονικό και κοινό όνομα οργανισμού καθώς και την ταξινομική κατάταξη του οργανισμού.

RN,RP, RC, RX, RA, RT & RL: Βιβλιογραφικές αναφορές.

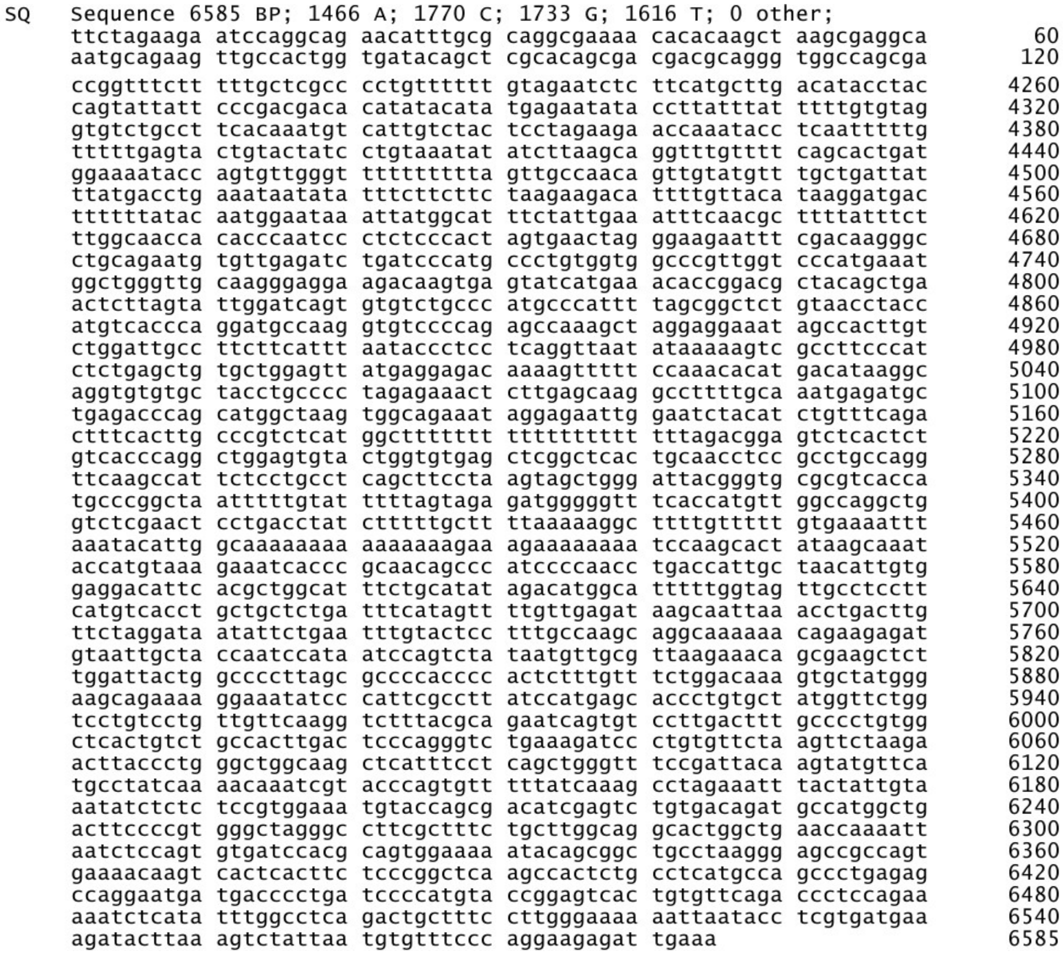
EMBL - Bank format (DNA database) (4/4)

FT Line: Feature table. Παρέχει μηχανισμό για σχολιασμό (annotation) των δεδομένων της ακολουθίας



```
FT   gene            1496..4598
FT                   /gene="DUSP1"
FT   mRNA            join(1496..2110,2390..2535,2902..3121,3564..4598)
FT                   /gene="DUSP1"
FT                   /product="dual specificity phosphatase 1"
FT   variation       1517
FT                   /gene="DUSP1"
FT                   /replace="t"
FT                   /frequency=".01"
FT   CDS             join(1744..2110,2390..2535,2902..3121,3564..3934)
FT                   /codon_start=1
FT                   /gene="DUSP1"
FT                   /product="dual specificity phosphatase 1"
FT                   /db_xref="GOA:P28562"
FT                   /db_xref="HGNC:3064"
FT                   /db_xref="HSSP:1M3G"
FT                   /db_xref="InterPro:IPR016130"
FT                   /db_xref="UniProtKB/Swiss-Prot:P28562"
FT                   /protein_id="ABB96250.1"
FT                   /translation="MVMEVGTLDAGGLRALLGERAAQCLLLDCRSFFAFNAGHIAGSVN
FT                   VRFSTIVRRRAKGAMGLEHIVPNAELRGRLLAGAYHAVVLLDERSAALDGAKRDGTLAL
FT                   AAGALCREARAAQVFFLKGGYEAFSASCPELCSKQSTPMGLSLPLSTSVPDSAESGCSS
FT                   CSTPLYDQGGPVEILPFLYLGSAYHASRKDMLDALGITALINVSANCPNHFEGHYQYKS
FT                   IPVEDNHKADISSWFNEAIDFIDSIKNAGGRVFVHCQAGISRSATICLAYLMRTNRVKL
FT                   DEAFEFVKQRRSIISPNFSFMGQLLQFESQVLAPHCSAEAGSPAMAVLDRGTSTTTVFN
FT                   FPVSIPVHSTNSALSYLQSPITTSPSC"
```

EMBL-Bank format (DNA database) - Sequence



```
SQ   Sequence 6585 BP; 1466 A; 1770 C; 1733 G; 1616 T; 0 other;
     ttctagaaga atccaggcag aacatttgcg caggcgaaaa cacacaagct aagcgaggca        60
     aatgcagaag ttgccactgg tgatacagct cgcacagcga cgacgcaggg tggccagcga       120
     ccggtttctt tttgctcgcc cctgtttttt gtagaatctc ttcatgcttg acatacctac      4260
     cagtattatt cccgacgaca catatacata tgagaatata ccttatttat ttttgtgtag      4320
     gtgtctgcct tcacaaatgt cattgtctac tcctagaaga accaaatacc tcaatttttg      4380
     tttttgagta ctgtactatc ctgtaaatat atcttaagca ggtttgtttt cagcactgat      4440
     ggaaaatacc agtgttgggt tttttttta gttgccaaca gttgtatgtt tgctgattat      4500
     ttatgacctg aaataatata tttcttcttc taagaagaca ttttgttaca taaggatgac      4560
     tttttatac aatggaataa attatggcat ttctattgaa atttcaacgc ttttatttct      4620
     ttggcaacca cacccaatcc ctctcccact agtgaactag ggaagaattt cgacaagggc      4680
     ctgcagaatg tgttgagatc tgatcccatg ccctgtggtg gcccgttggt cccatgaaat      4740
     ggctgggttg caagggagga agacaagtga gtatcatgaa acaccggacg ctacagctga      4800
     actcttagta ttggatcagt gtgtctgccc atgcccattt tagcggctct gtaacctacc      4860
     atgtcaccca ggatgccaag gtgtccccag agccaaagct aggaggaaat agccacttgt      4920
     ctggattgcc ttcttcattt aataccctcc tcaggttaat ataaaaagtc gccttcccat      4980
     ctctgagctg tgctggagtt atgaggagac aaaagttttt ccaaacacat gacataaggc      5040
     aggtgtgtgc tacctgcccc tagagaaact cttgagcaag gccttttgca aatgagatgc      5100
     tgagacccag catggctaag tggcagaaat aggagaattg gaatctacat ctgtttcaga      5160
     ctttcacttg cccgtctcat ggcttttttt tttttttttt tttagacgga gtctcactct      5220
     gtcacccagg ctggagtgta ctggtgtgag ctcggctcac tgcaacctcc gcctgccagg      5280
     ttcaagccat tctcctgcct cagcttccta agtagctggg attacgggtg cgcgtcacca      5340
     tgcccggcta atttttgtat ttttagtaga gatgggggtt tcaccatgtt ggccaggctg      5400
     gtctcgaact cctgacctat ctttttgctt ttaaaaaggc ttttgttttt gtgaaaattt      5460
     aaatacattg gcaaaaaaaa aaaaaaagaa agaaaaaaaa tccaagcact ataagcaaat      5520
     accatgtaaa gaaatcaccc gcaacagccc atccccaacc tgaccattgc taacattgtg      5580
     gaggacattc acgctggcat ttctgcatat agacatggca tttttggtag ttgcctcctt      5640
     catgtcacct gctgctctga tttcatagtt ttgttgagat aagcaattaa acctgacttg      5700
     ttctaggata atattctgaa tttgtactcc tttgccaagc aggcaaaaaa cagaagagat      5760
     gtaattgcta ccaatccata atccagtcta taatgttgcg ttaagaaaca gcgaagctct      5820
     tggattactg gccccttagc gccccacccc actctttgtt tctggacaaa gtgctatggg      5880
     aagcagaaaa ggaaatatcc cattcgcctt atccatgagc accctgtgct atggttctgg      5940
     tcctgtcctg ttgttcaagg tctttacgca gaatcagtgt ccttgacttt gcccctgtgg      6000
     ctcactgtct gccacttgac tcccagggtc tgaaagatcc ctgtgttcta agttctaaga      6060
     acttaccctg ggctggcaag ctcatttcct cagctgggtt tccgattaca agtatgttca      6120
     tgcctatcaa aacaaatcgt acccagtgtt tttatcaaag cctagaaatt tactattgta      6180
     aatatctctc tccgtggaaa tgtaccagcg acatcgagtc tgtgacagat gccatggctg      6240
     acttccccgt gggctagggc cttcgctttc tgcttggcag gcactggctg aaccaaaatt      6300
     aatctccagt gtgatccacg cagtggaaaa atacagcggc tgcctaaggg agccgccagt      6360
     gaaaacaagt cactcacttc tcccggctca agccactctg cctcatgcca gccctgagag      6420
     ccaggaatga tgacccctga tccccatgta ccggagtcac tgtgttcaga ccctccagaa      6480
     aaatctcata tttggcctca gactgctttc cttgggaaaa aattaatacc tcgtgatgaa      6540
     agatacttaa agtctattaa tgtgtttccc aggaagagat tgaaa                      6585
```

Πρωτεϊνική βάση δεδομένων - UniProt

- Το 2002, οι τρεις μεγάλες πρωτεϊνικές βάσεις δεδομένων, Swiss-Prot, TrEMBL και PIR-PSD, ενώθηκαν με αποτέλεσμα τη δημιουργία της μεγάλης παγκοσμιοποιημένης βάσης δεδομένων, **Universal Protein Resource, UniProt.**
- **Περιέχει αναφορές από περισσότερες από 50 πρωτεϊνικές βάσεις δεδομένων.**

Protein sequence database – UniProt

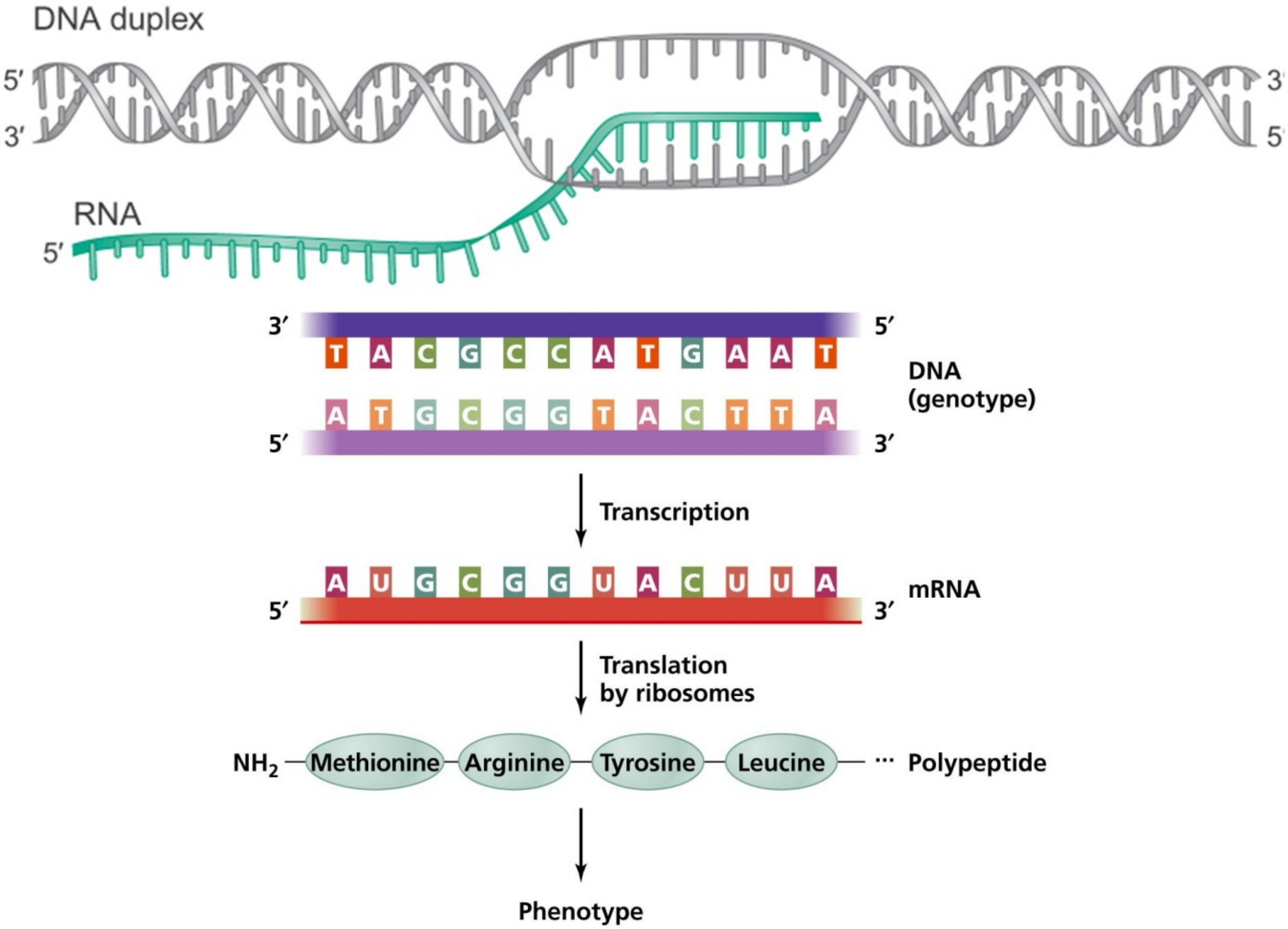
Consists of 3 layers:

i. Knowledgebase (UniProt): A comprehensive catalogue of data on protein sequence and function. Contains only the active entries

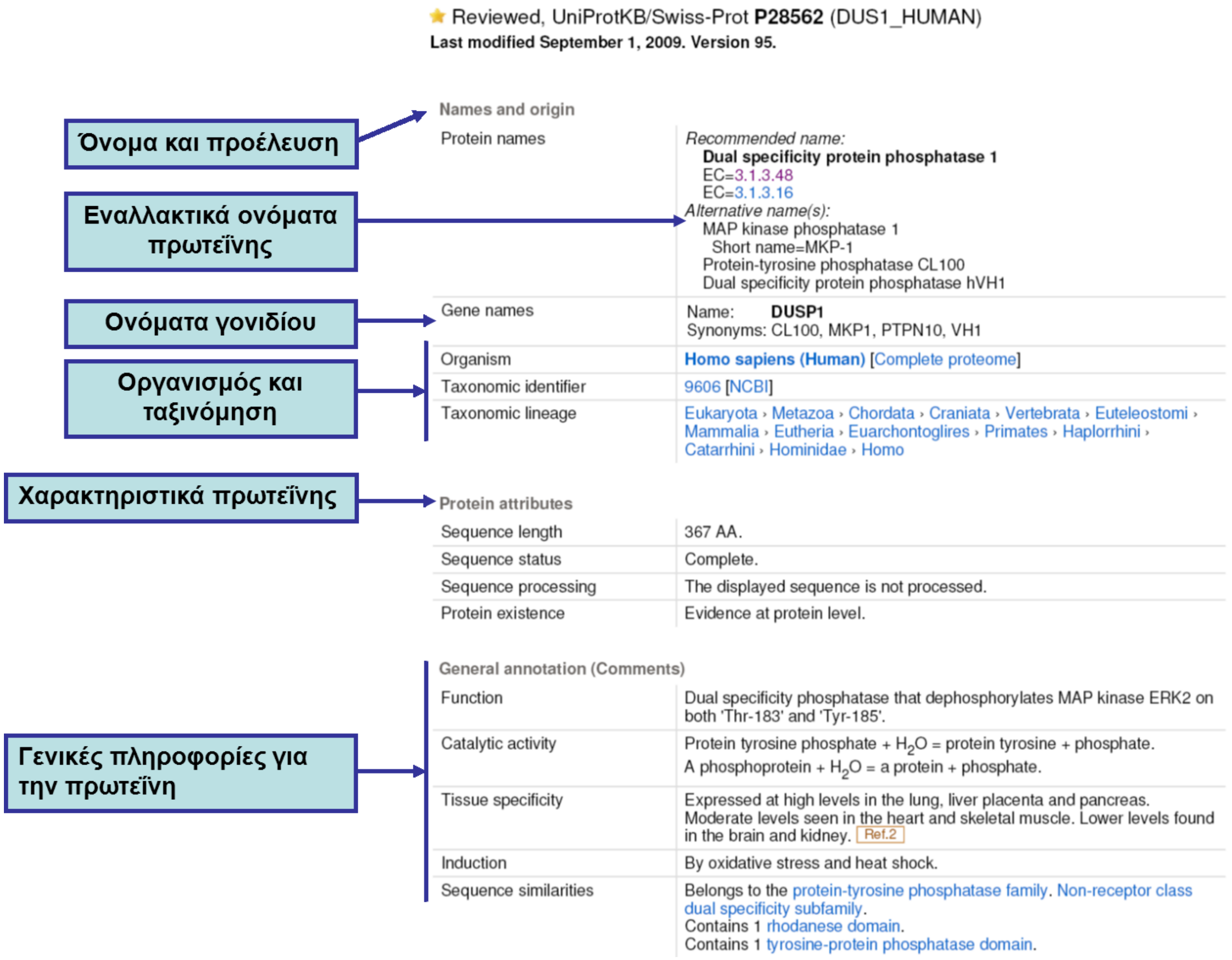
ii. Archive (UniParc): Repository for all protein sequences, providing a mechanism by which the historical association of database records and protein sequences can be tracked. Contains both active and inactive entries.

iii. Non-redundant database (UniRef): Sequence clusters, used to speed up sequence similarity searches.

DNA duplex


DNA duplex
5′
3′
3′
5′
RNA
5′
3′
5′
T A C G C C A T G A A T
DNA
(genotype)
A T G C G G T A C T T A
5′
3′
Transcription
A U G C G G U A C U U A
mRNA
5′
3′
Translation
by ribosomes
NH$_2$
Methionine
Arginine
Tyrosine
Leucine
Polypeptide
Phenotype

UniProt Format (1/3)



Όνομα και προέλευση

Εναλλακτικά ονόματα πρωτεΐνης

Ονόματα γονιδίου

Οργανισμός και ταξινόμηση

Χαρακτηριστικά πρωτεΐνης

Γενικές πληροφορίες για την πρωτεΐνη

★ Reviewed, UniProtKB/Swiss-Prot **P28562** (DUS1_HUMAN)
Last modified September 1, 2009. Version 95.

Names and origin

Protein names	*Recommended name:* **Dual specificity protein phosphatase 1** EC=3.1.3.48 EC=3.1.3.16 *Alternative name(s):* MAP kinase phosphatase 1 Short name=MKP-1 Protein-tyrosine phosphatase CL100 Dual specificity protein phosphatase hVH1
Gene names	Name: **DUSP1** Synonyms: CL100, MKP1, PTPN10, VH1
Organism	**Homo sapiens (Human)** [Complete proteome]
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo

Protein attributes

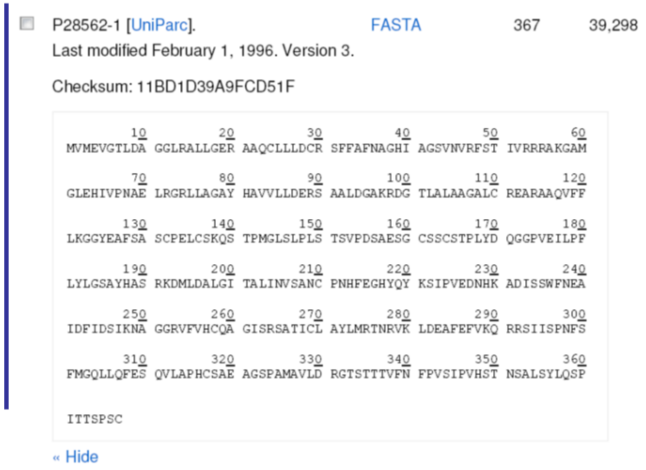
Sequence length	367 AA.
Sequence status	Complete.
Sequence processing	The displayed sequence is not processed.
Protein existence	Evidence at protein level.

General annotation (Comments)

Function	Dual specificity phosphatase that dephosphorylates MAP kinase ERK2 on both 'Thr-183' and 'Tyr-185'.
Catalytic activity	Protein tyrosine phosphate + H_2O = protein tyrosine + phosphate. A phosphoprotein + H_2O = a protein + phosphate.
Tissue specificity	Expressed at high levels in the lung, liver placenta and pancreas. Moderate levels seen in the heart and skeletal muscle. Lower levels found in the brain and kidney. Ref.2
Induction	By oxidative stress and heat shock.
Sequence similarities	Belongs to the protein-tyrosine phosphatase family. Non-receptor class dual specificity subfamily. Contains 1 rhodanese domain. Contains 1 tyrosine-protein phosphatase domain.

UniProt Format (2/3)

Πληροφορίες για την αλληλουχία



☐ P28562-1 [UniParc]. FASTA 367 39,298

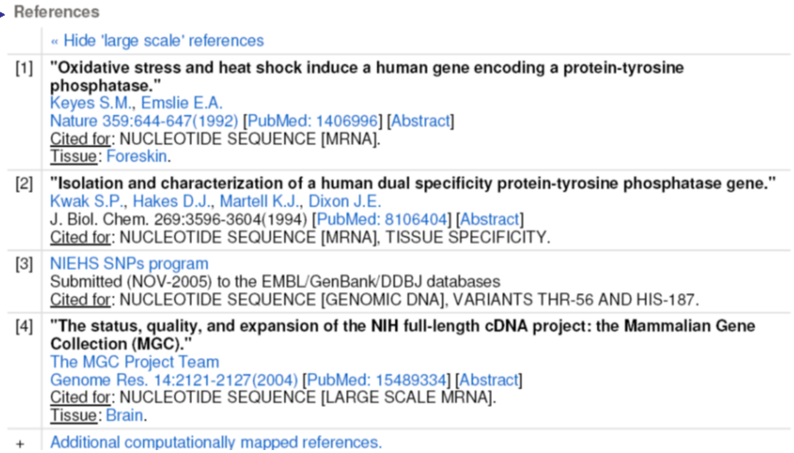
Last modified February 1, 1996. Version 3.

Checksum: 11BD1D39A9FCD51F

```
        10         20         30         40         50         60
MVMEVGTLDA GGLRALLGER AAQCLLLDCR SFFAFNAGHI AGSVNVRFST IVRRRAKGAM

        70         80         90        100        110        120
GLEHIVPNAE LRGRLLAGAY HAVVLLDERS AALDGAKRDG TLALAAGALC REARAAQVFF

       130        140        150        160        170        180
LKGGYEAFSA SCPELCSKQS TPMGLSLPLS TSVPDSAESG CSSCSTPLYD QGGPVEILPF

       190        200        210        220        230        240
LYLGSAYHAS RKDMLDALGI TALINVSANC PNHFEGHYQY KSIPVEDNHK ADISSWFNEA

       250        260        270        280        290        300
IDFIDSIKNA GGRVFVHCQA GISRSATICL AYLMRTNRVK LDEAFEFVKQ RRSIISPNFS

       310        320        330        340        350        360
FMGQLLQFES QVLAPHCSAE AGSPAMAVLD RGTSTTTVFN FPVSIPVHST NSALSYLQSP


ITTSPSC
```

« Hide

Βιβλιογραφικές αναφορές



References

« Hide 'large scale' references

[1] **"Oxidative stress and heat shock induce a human gene encoding a protein-tyrosine phosphatase."**
Keyes S.M., Emslie E.A.
Nature 359:644-647(1992) [PubMed: 1406996] [Abstract]
Cited for: NUCLEOTIDE SEQUENCE [MRNA].
Tissue: Foreskin.

[2] **"Isolation and characterization of a human dual specificity protein-tyrosine phosphatase gene."**
Kwak S.P., Hakes D.J., Martell K.J., Dixon J.E.
J. Biol. Chem. 269:3596-3604(1994) [PubMed: 8106404] [Abstract]
Cited for: NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY.

[3] NIEHS SNPs program
Submitted (NOV-2005) to the EMBL/GenBank/DDBJ databases
Cited for: NUCLEOTIDE SEQUENCE [GENOMIC DNA], VARIANTS THR-56 AND HIS-187.

[4] **"The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC)."**
The MGC Project Team
Genome Res. 14:2121-2127(2004) [PubMed: 15489334] [Abstract]
Cited for: NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
Tissue: Brain.

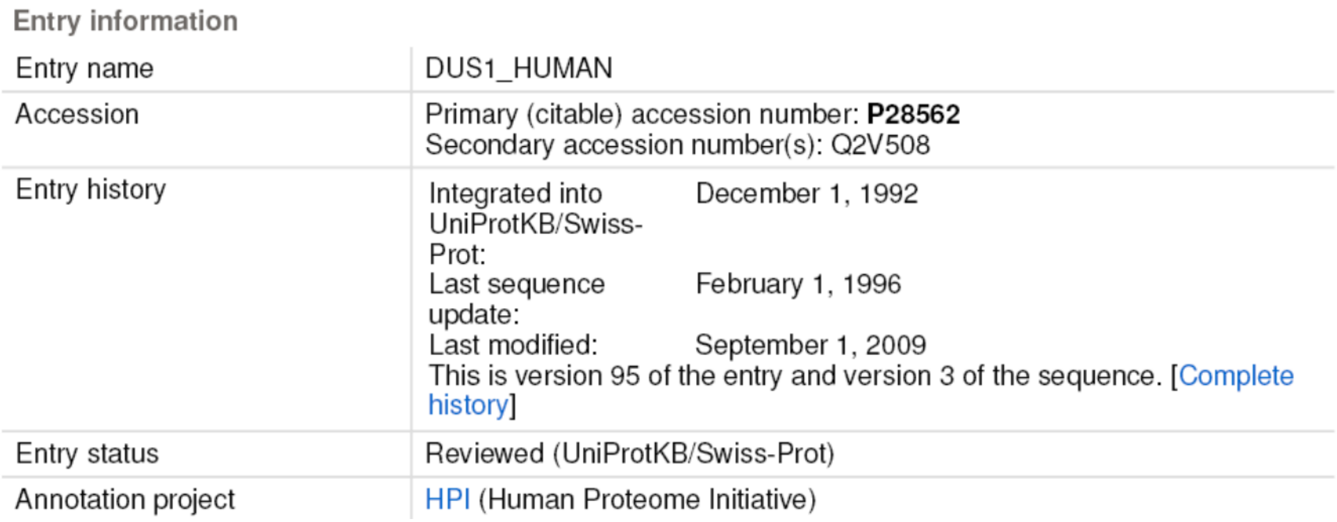
\+ Additional computationally mapped references.

UniProt Format (3/3)

Πληροφορίες για την τριτοταγή δομή της πρωτεΐνης

3D structure databases	
HSSP	HSSP built from PDB template 1M3G based on UniProtKB Q05923.
SMR	P28562. Positions 171-313.
Protein-protein interaction databases	
IntAct	P28562. 7 interactions.
STRING	P28562.

Γενικές πληροφορίες για την καταχώρηση

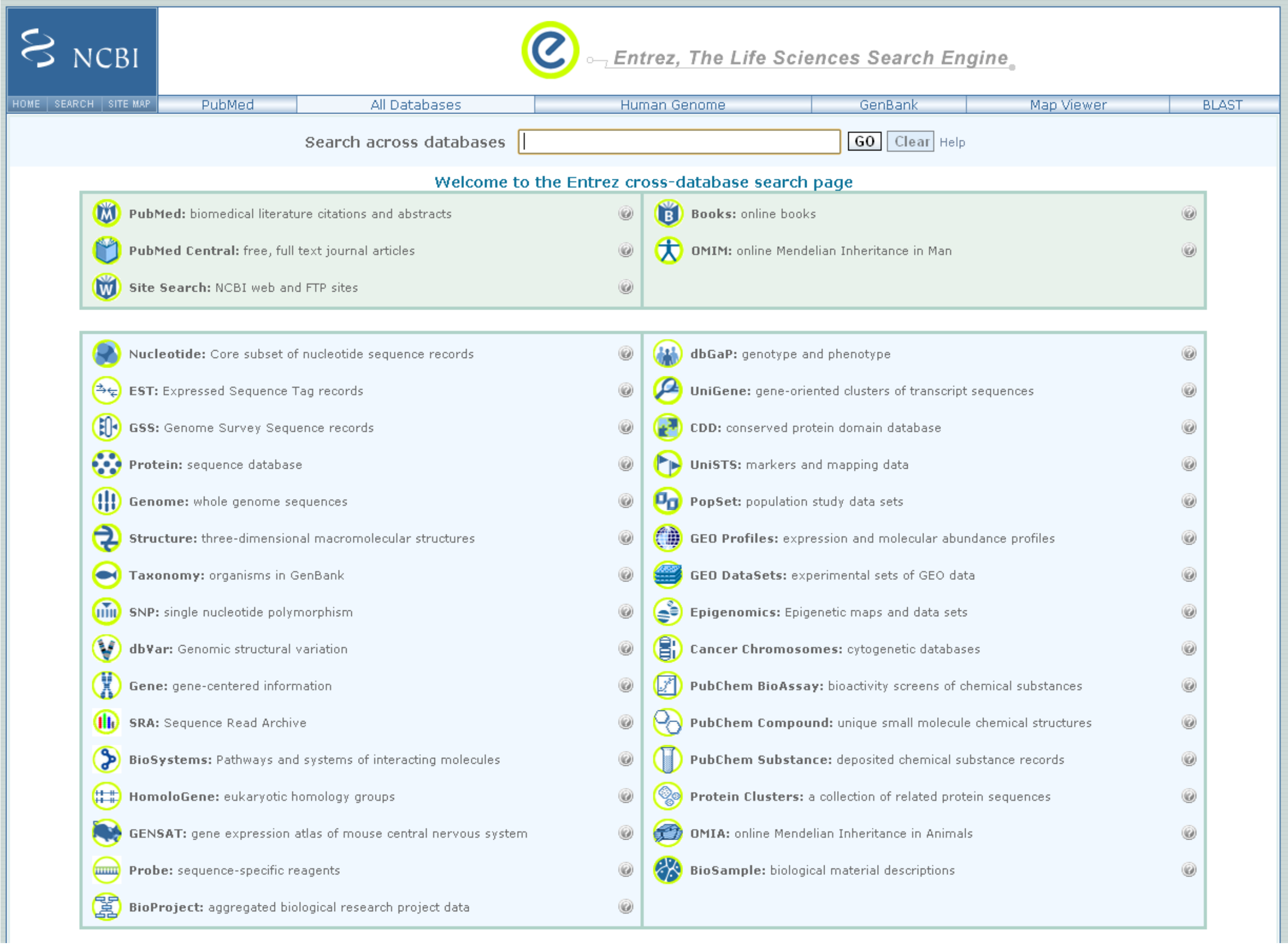


Entry information	
Entry name	DUS1_HUMAN
Accession	Primary (citable) accession number: **P28562** Secondary accession number(s): Q2V508
Entry history	Integrated into UniProtKB/Swiss-Prot: December 1, 1992 Last sequence update: February 1, 1996 Last modified: September 1, 2009 This is version 95 of the entry and version 3 of the sequence. [Complete history]
Entry status	Reviewed (UniProtKB/Swiss-Prot)
Annotation project	HPI (Human Proteome Initiative)

DDBJ – DNA Data Banf of Japan (Japan)

- Η μοναδική νουκλεοτιδική βάση δεδομένων στην Ασία.
- Λειτουργεί υπό την αιγίδα του Υπουργείου Παιδείας Επιστημών και Αθλητισμού της Ιαπωνίας.
- Πηγή των δεδομένων της είναι, κυρίως, Ιάπωνες ερευνητές.
- Η τελευταία έκδοσή της περιέχει:
  - 87,903,140 εγγραφές.
  - 91,294,770,939 βάσεις.

Μηχανές αναζήτησης βιολογικών δεδομένων – NCBI, Entrez



Welcome to the Entrez cross-database search page

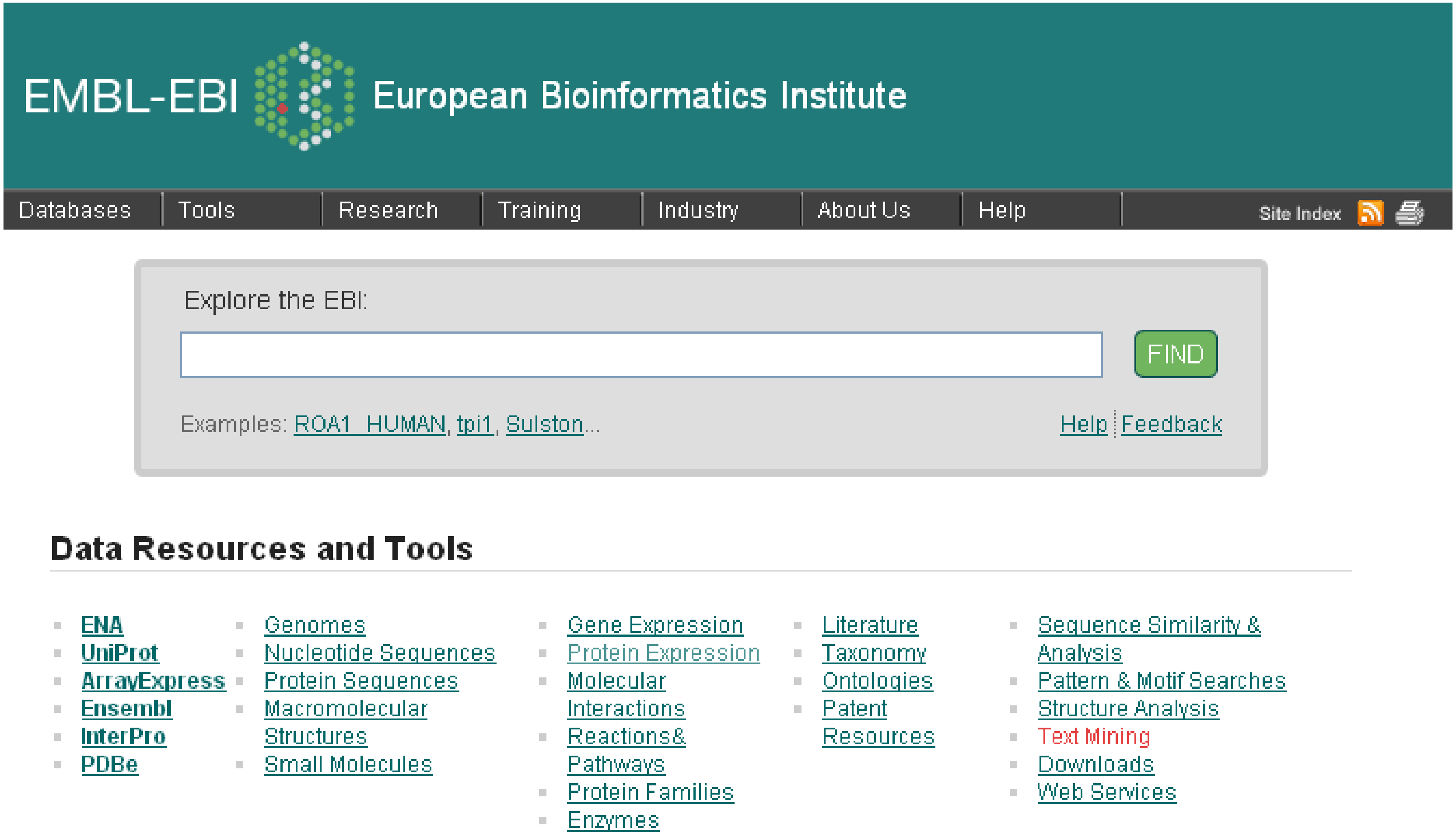
PubMed: biomedical literature citations and abstracts
PubMed Central: free, full text journal articles
Site Search: NCBI web and FTP sites

Books: online books
OMIM: online Mendelian Inheritance in Man

Nucleotide: Core subset of nucleotide sequence records
EST: Expressed Sequence Tag records
GSS: Genome Survey Sequence records
Protein: sequence database
Genome: whole genome sequences
Structure: three-dimensional macromolecular structures
Taxonomy: organisms in GenBank
SNP: single nucleotide polymorphism
dbVar: Genomic structural variation
Gene: gene-centered information
SRA: Sequence Read Archive
BioSystems: Pathways and systems of interacting molecules
HomoloGene: eukaryotic homology groups
GENSAT: gene expression atlas of mouse central nervous system
Probe: sequence-specific reagents
BioProject: aggregated biological research project data

dbGaP: genotype and phenotype
UniGene: gene-oriented clusters of transcript sequences
CDD: conserved protein domain database
UniSTS: markers and mapping data
PopSet: population study data sets
GEO Profiles: expression and molecular abundance profiles
GEO DataSets: experimental sets of GEO data
Epigenomics: Epigenetic maps and data sets
Cancer Chromosomes: cytogenetic databases
PubChem BioAssay: bioactivity screens of chemical substances
PubChem Compound: unique small molecule chemical structures
PubChem Substance: deposited chemical substance records
Protein Clusters: a collection of related protein sequences
OMIA: online Mendelian Inheritance in Animals
BioSample: biological material descriptions

Μηχανές αναζήτησης βιολογικών δεδομένων - EBI



Explore the EBI:

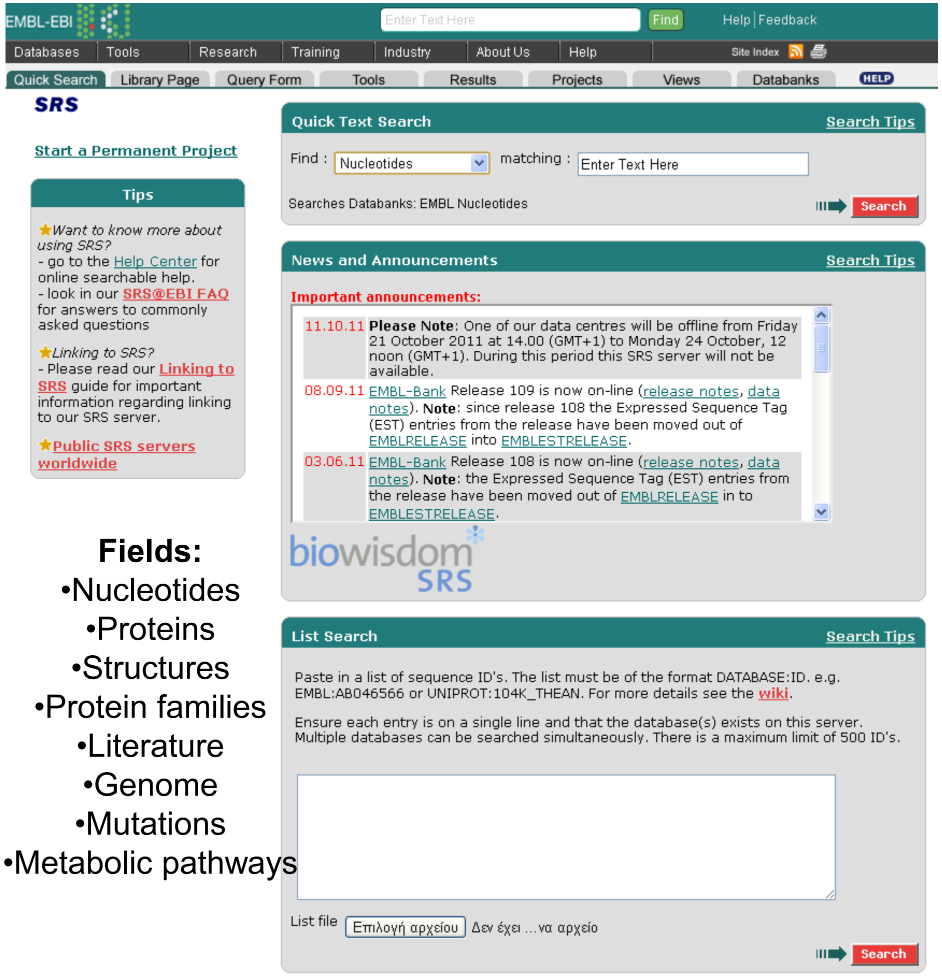
Examples: ROA1_HUMAN, tpi1, Sulston...

Help | Feedback

Data Resources and Tools

- **ENA**
- **UniProt**
- **ArrayExpress**
- **Ensembl**
- **InterPro**
- **PDBe**
- Genomes
- Nucleotide Sequences
- Protein Sequences
- Macromolecular Structures
- Small Molecules
- Gene Expression
- Protein Expression
- Molecular Interactions
- Reactions& Pathways
- Protein Families
- Enzymes
- Literature
- Taxonomy
- Ontologies
- Patent Resources
- Sequence Similarity & Analysis
- Pattern & Motif Searches
- Structure Analysis
- Text Mining
- Downloads
- Web Services

Μηχανές αναζήτησης βιολογικών δεδομένων - SRS



Tips

★*Want to know more about using SRS?*
- go to the Help Center for online searchable help.
- look in our SRS@EBI FAQ for answers to commonly asked questions

★*Linking to SRS?*
- Please read our Linking to SRS guide for important information regarding linking to our SRS server.

★Public SRS servers worldwide

Quick Text Search

Find : Nucleotides matching : Enter Text Here

Searches Databanks: EMBL Nucleotides

News and Announcements

Important announcements:

11.10.11 **Please Note**: One of our data centres will be offline from Friday 21 October 2011 at 14.00 (GMT+1) to Monday 24 October, 12 noon (GMT+1). During this period this SRS server will not be available.

08.09.11 EMBL-Bank Release 109 is now on-line (release notes, data notes). **Note**: since release 108 the Expressed Sequence Tag (EST) entries from the release have been moved out of EMBLRELEASE into EMBLESTRELEASE.

03.06.11 EMBL-Bank Release 108 is now on-line (release notes, data notes). **Note**: the Expressed Sequence Tag (EST) entries from the release have been moved out of EMBLRELEASE in to EMBLESTRELEASE.

List Search

Paste in a list of sequence ID's. The list must be of the format DATABASE:ID. e.g. EMBL:AB046566 or UNIPROT:104K_THEAN. For more details see the wiki.

Ensure each entry is on a single line and that the database(s) exists on this server. Multiple databases can be searched simultaneously. There is a maximum limit of 500 ID's.

List file Επιλογή αρχείου Δεν έχει ...να αρχείο

Fields:
- Nucleotides
- Proteins
- Structures
- Protein families
- Literature
- Genome
- Mutations
- Metabolic pathways

Μηχανές αναζήτησης δεδομένων στις βιολογικές βάσεις δεδομένων

- Επιλογές για αναζήτηση αλληλουχίας DNA:
  - EBI, http://www.ebi.ac.uk/
  - GENBANK, http://www.ncbi.nlm.nih.gov/Genbank/
- Επιλογές για αναζήτηση πρωτεϊνικής αλληλουχίας:
  - UniProt, http://www.uniprot.org/

Βάσεις δεδομένων πρωτεϊνικών δομών

- Βάσεις δεδομένων για την τρισδιάστατη δομή πρωτεϊνών, των οποίων η δομή έχει λυθεί με τη χρήση διαφόρων τεχνικών όπως κρυσταλλογραφία ακτίνων Χ ή NMR.
- **Protein Databases:**
  - **PDB.**
  - SCOP.
  - Swiss-Prot.
  - PIR.

RasMol:

- Molecular Visualization Freeware for proteins, DNA and macromolecules.

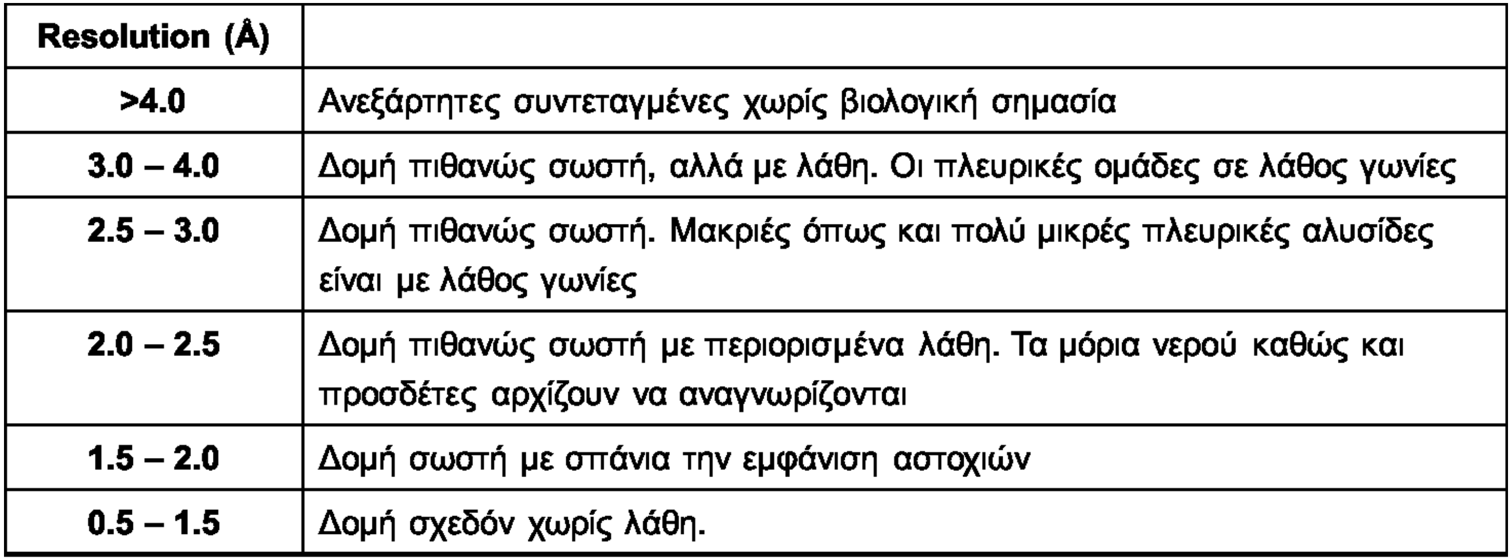
Protein Structure Databases

- Η πιο εκτεταμένη βάση δεδομένων τρισδιάστατων πρωτεϊνικών δομών είναι η **Protein Data Bank (PDB).**
- January 18 , 2010: 54,039 structures.
- October 30, 2012: 85,848 structures.

RCSB PDB
PROTEIN DATA BANK

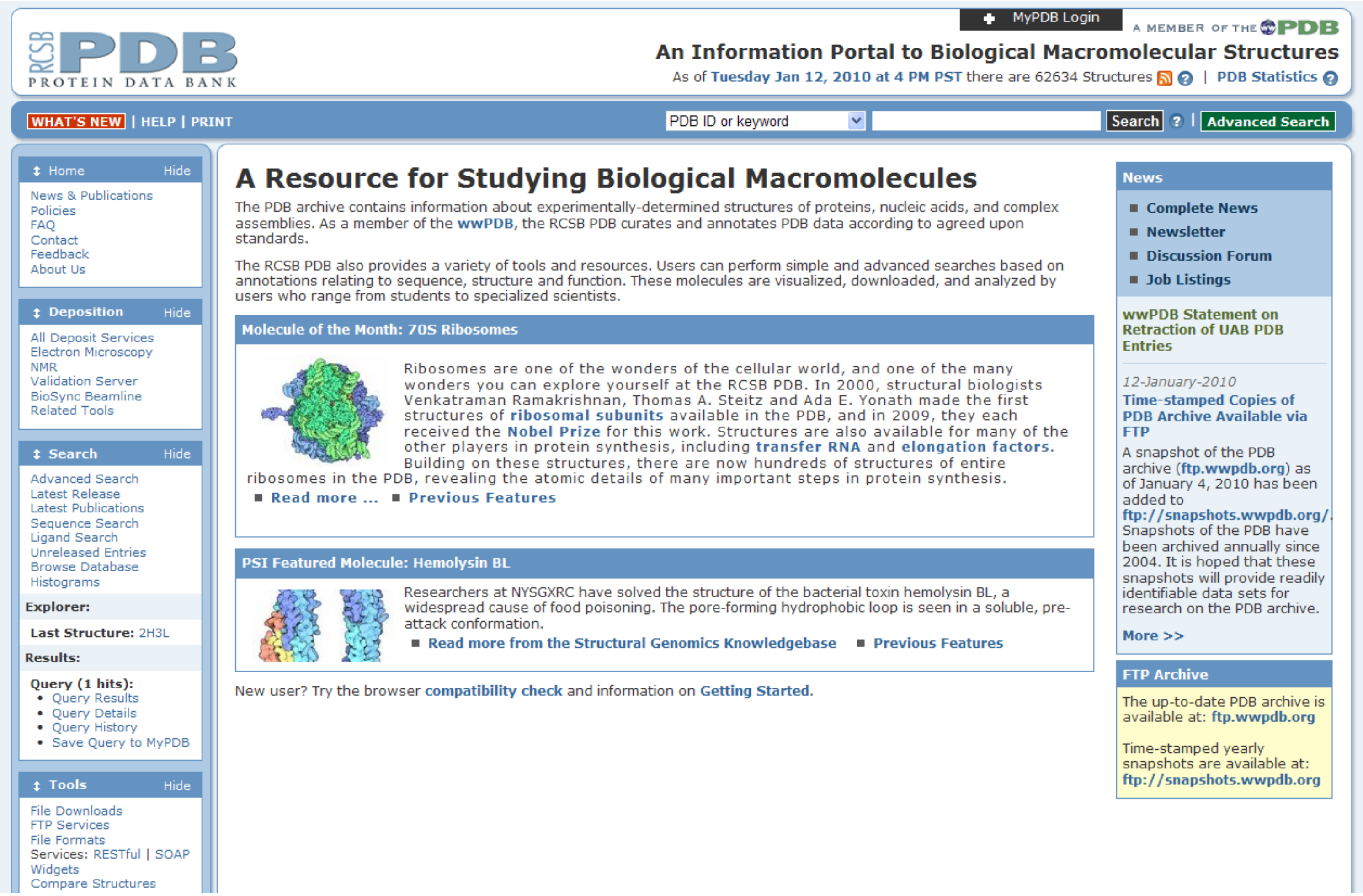
Protein Data Bank (PDB) (1/4)

- 90% των πρωτεϊνικών δομών: X-ray crystallography.
- 9% των πρωτεϊνικών δομών: Nuclear Magnetic Resonance techniques (NMR).



Resolution (Å)	
>4.0	Ανεξάρτητες συντεταγμένες χωρίς βιολογική σημασία
3.0 – 4.0	Δομή πιθανώς σωστή, αλλά με λάθη. Οι πλευρικές ομάδες σε λάθος γωνίες
2.5 – 3.0	Δομή πιθανώς σωστή. Μακριές όπως και πολύ μικρές πλευρικές αλυσίδες είναι με λάθος γωνίες
2.0 – 2.5	Δομή πιθανώς σωστή με περιορισμένα λάθη. Τα μόρια νερού καθώς και προσδέτες αρχίζουν να αναγνωρίζονται
1.5 – 2.0	Δομή σωστή με σπάνια την εμφάνιση αστοχιών
0.5 – 1.5	Δομή σχεδόν χωρίς λάθη.

Protein Data Bank (PDB) (2/4)



An Information Portal to Biological Macromolecular Structures

As of **Tuesday Jan 12, 2010 at 4 PM PST** there are 62634 Structures | **PDB Statistics**

A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the **wwPDB**, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Molecule of the Month: 70S Ribosomes

Ribosomes are one of the wonders of the cellular world, and one of the many wonders you can explore yourself at the RCSB PDB. In 2000, structural biologists Venkatraman Ramakrishnan, Thomas A. Steitz and Ada E. Yonath made the first structures of **ribosomal subunits** available in the PDB, and in 2009, they each received the **Nobel Prize** for this work. Structures are also available for many of the other players in protein synthesis, including **transfer RNA** and **elongation factors**. Building on these structures, there are now hundreds of structures of entire ribosomes in the PDB, revealing the atomic details of many important steps in protein synthesis.

- **Read more ...**
- **Previous Features**

PSI Featured Molecule: Hemolysin BL

Researchers at NYSGXRC have solved the structure of the bacterial toxin hemolysin BL, a widespread cause of food poisoning. The pore-forming hydrophobic loop is seen in a soluble, pre-attack conformation.

- **Read more from the Structural Genomics Knowledgebase**
- **Previous Features**

New user? Try the browser **compatibility check** and information on **Getting Started**.

wwPDB Statement on Retraction of UAB PDB Entries

12-January-2010

Time-stamped Copies of PDB Archive Available via FTP

A snapshot of the PDB archive (**ftp.wwpdb.org**) as of January 4, 2010 has been added to **ftp://snapshots.wwpdb.org/**. Snapshots of the PDB have been archived annually since 2004. It is hoped that these snapshots will provide readily identifiable data sets for research on the PDB archive.

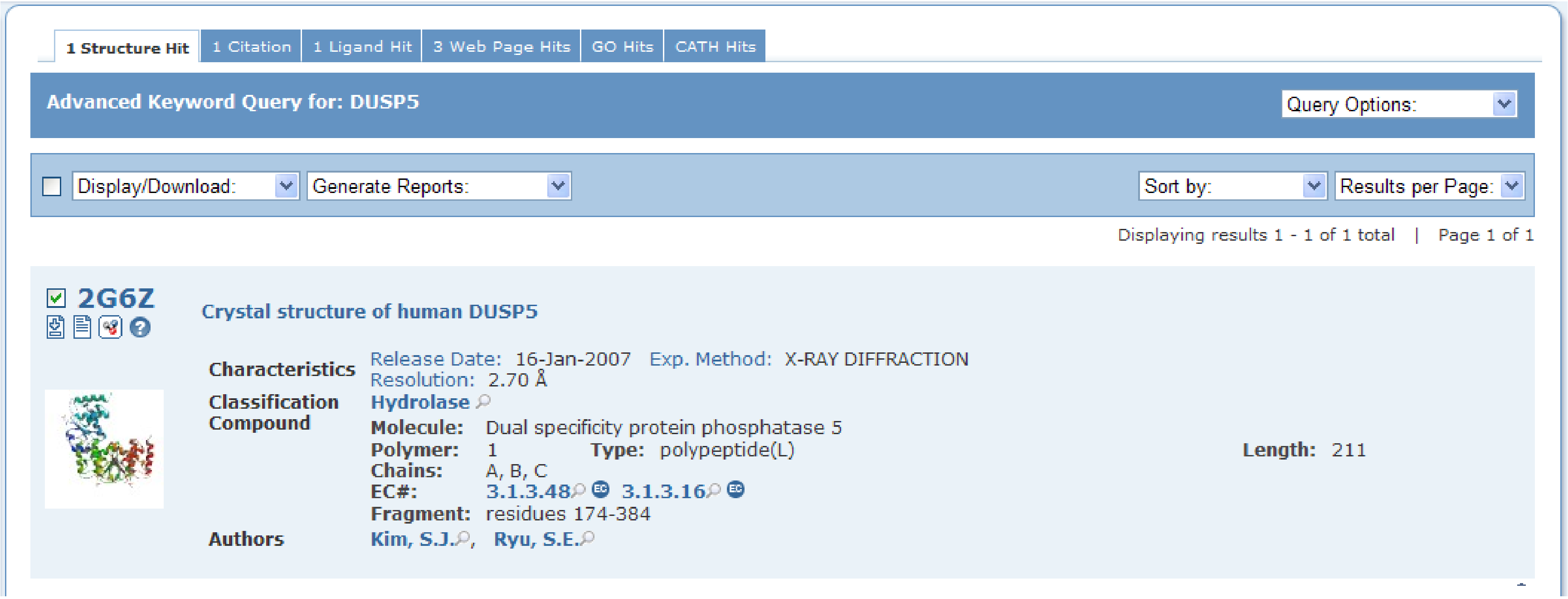
More >>

FTP Archive

The up-to-date PDB archive is available at: **ftp.wwpdb.org**

Time-stamped yearly snapshots are available at: **ftp://snapshots.wwpdb.org**

Protein Data Bank (PDB) (3/4)



Advanced Keyword Query for: DUSP5

☐

Displaying results 1 - 1 of 1 total | Page 1 of 1

☑ **2G6Z**

Crystal structure of human DUSP5

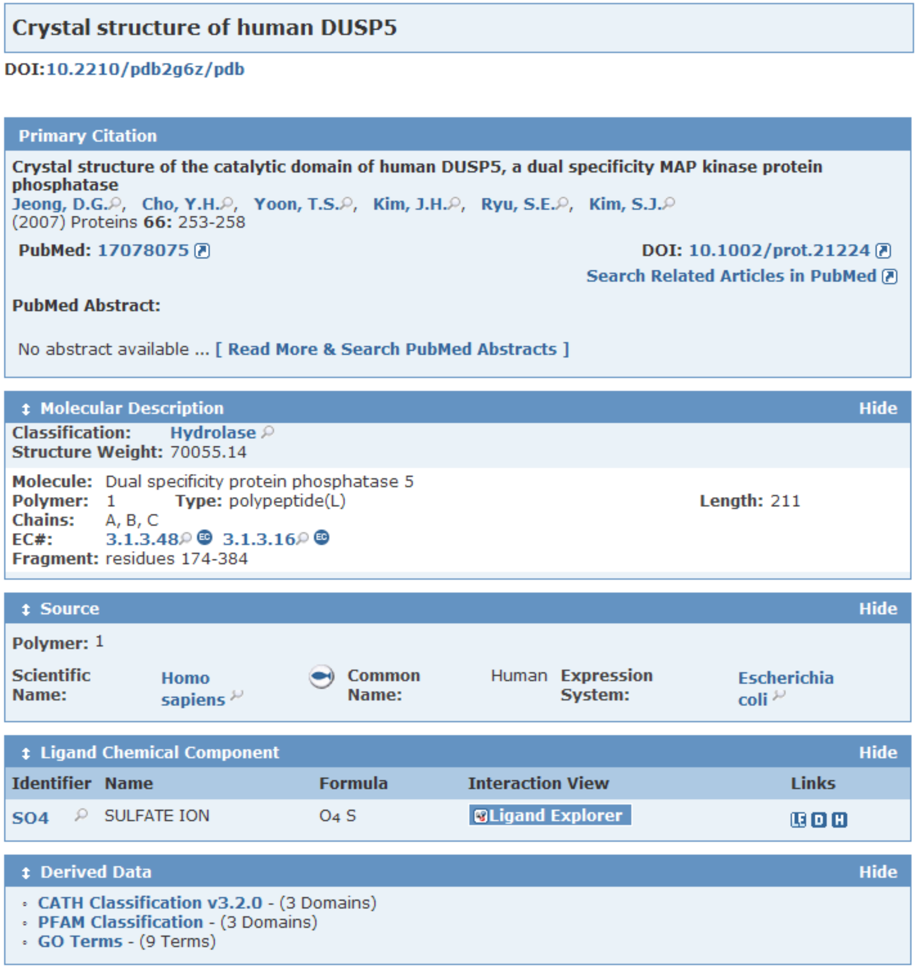
Characteristics Release Date: 16-Jan-2007 Exp. Method: X-RAY DIFFRACTION
Resolution: 2.70 Å

Classification **Hydrolase**

Compound
Molecule: Dual specificity protein phosphatase 5
Polymer: 1 **Type:** polypeptide(L) **Length:** 211
Chains: A, B, C
EC#: 3.1.3.48 3.1.3.16
Fragment: residues 174-384

Authors **Kim, S.J.**, **Ryu, S.E.**

Protein Data Bank (PDB) (4/4)



Crystal structure of human DUSP5

DOI:10.2210/pdb2g6z/pdb

Primary Citation

Crystal structure of the catalytic domain of human DUSP5, a dual specificity MAP kinase protein phosphatase
Jeong, D.G., **Cho, Y.H.**, **Yoon, T.S.**, **Kim, J.H.**, **Ryu, S.E.**, **Kim, S.J.**
(2007) Proteins **66:** 253-258

PubMed: 17078075 **DOI:** 10.1002/prot.21224
Search Related Articles in PubMed

PubMed Abstract:

No abstract available ... [Read More & Search PubMed Abstracts]

Molecular Description Hide

Classification: Hydrolase
Structure Weight: 70055.14

Molecule: Dual specificity protein phosphatase 5
Polymer: 1 **Type:** polypeptide(L) **Length:** 211
Chains: A, B, C
EC#: 3.1.3.48 3.1.3.16
Fragment: residues 174-384

Source Hide

Polymer: 1

Scientific Name: Homo sapiens **Common Name:** Human **Expression System:** Escherichia coli

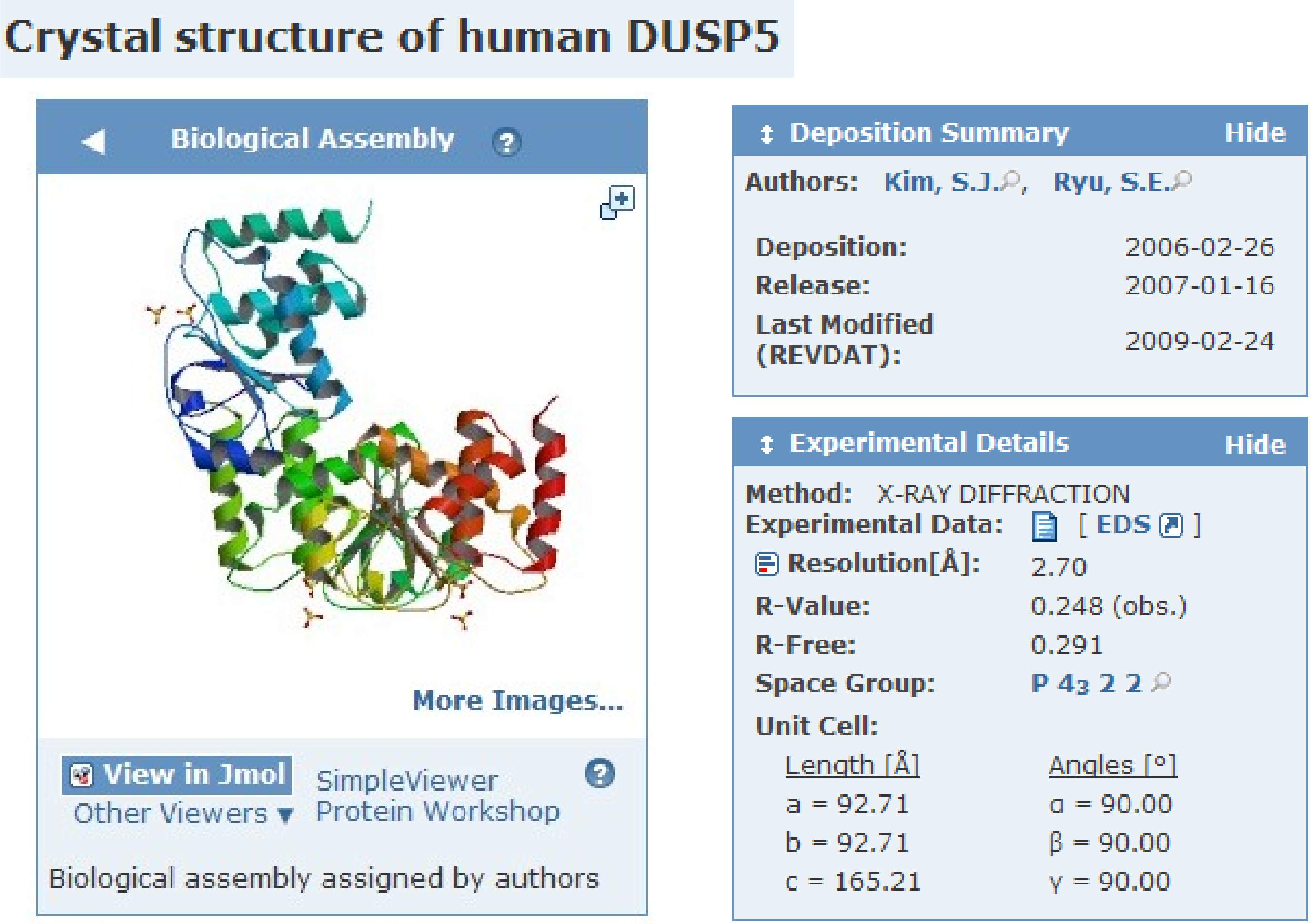
Ligand Chemical Component Hide

Identifier	Name	Formula	Interaction View	Links
SO4	SULFATE ION	$O_4 S$	Ligand Explorer	L D H

Derived Data Hide

- CATH Classification v3.2.0 - (3 Domains)
- PFAM Classification - (3 Domains)
- GO Terms - (9 Terms)

Crystal Structure



Crystal structure of human DUSP5

Biological assembly assigned by authors

Deposition Summary

Authors: Kim, S.J., Ryu, S.E.

Deposition:	2006-02-26
Release:	2007-01-16
Last Modified (REVDAT):	2009-02-24

Experimental Details

Method: X-RAY DIFFRACTION

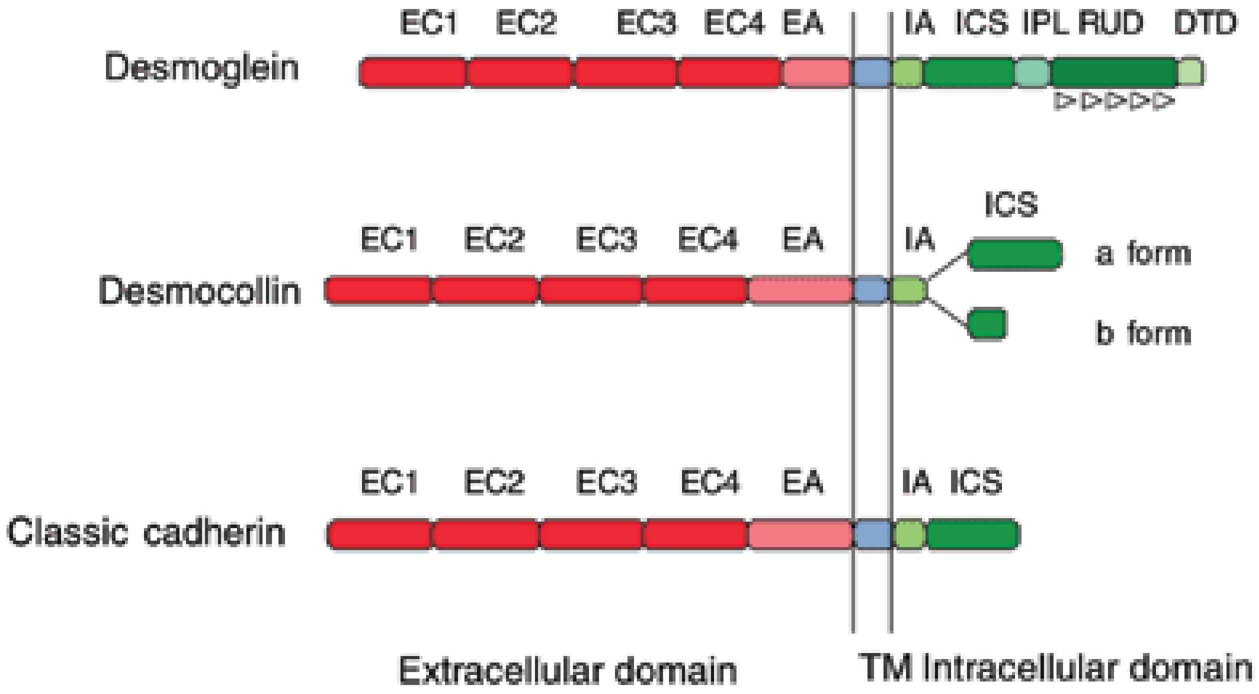
Experimental Data: [EDS]

Resolution[Å]:	2.70
R-Value:	0.248 (obs.)
R-Free:	0.291
Space Group:	$P\ 4_3\ 2\ 2$

Unit Cell:

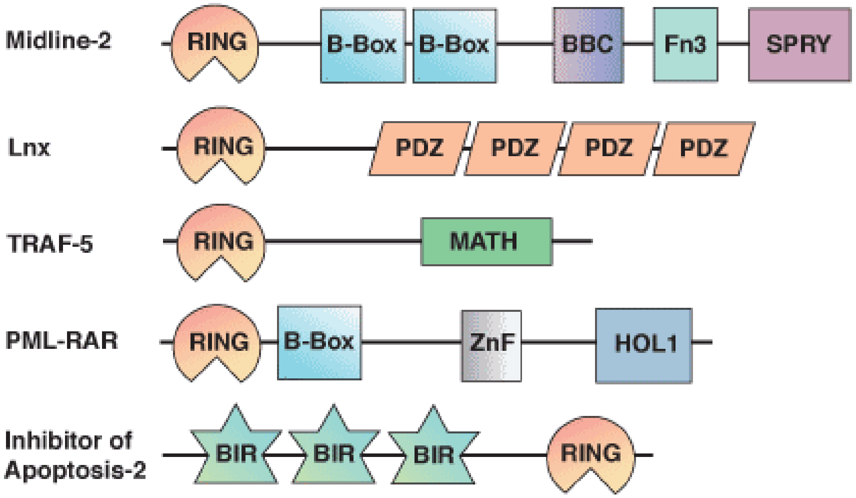
Length [Å]	Angles [°]
a = 92.71	α = 90.00
b = 92.71	β = 90.00
c = 165.21	γ = 90.00

Protein **domain / motif** databases (1/2)



Protein **domain / motif** databases (2/2)

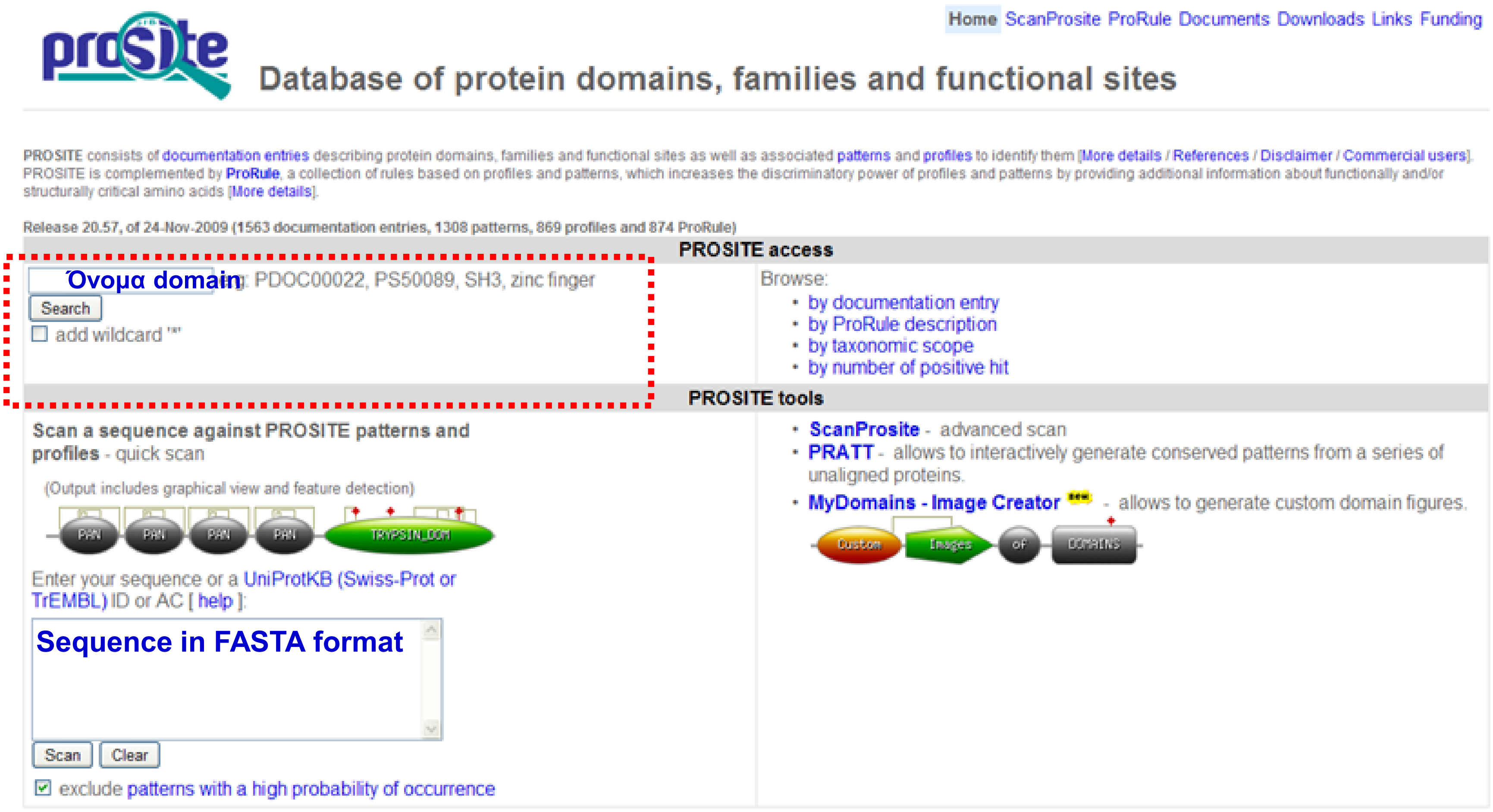
- **Motif:** A small structural element that is recognizable in several proteins. Length of 6-10 a.a.
- **Domain:** A part of a protein with specific structure which correspond to the family that belongs to the protein. Is related to the function of localization of the protein.



PROSITE (1/4)

- Αρχικά: Περιείχε απλά μοτίβα πρωτεϊνών.
- Σήμερα περιλαμβάνει αρχεία που περιγράφουν:
  - Οικογένειες πρωτεϊνών.
  - Λειτουργικές θέσεις πρωτεϊνών.
  - Προφίλ πρωτεϊνών.
- Τρέχουσα έκδοση:
  - 1650 εγγραφές με βιβλιογραφικά δεδομένα για τα,
  - 1308 πρότυπα και τα,
  - 1039 προφίλ που είναι καταγεγραμμένα σε αυτή.

PROSITE (2/4)



Database of protein domains, families and functional sites

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More details / References / Disclaimer / Commercial users]. PROSITE is complemented by ProRule, a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More details].

Release 20.57, of 24-Nov-2009 (1563 documentation entries, 1308 patterns, 869 profiles and 874 ProRule)

PROSITE access

Όνομα domain : PDOC00022, PS50089, SH3, zinc finger

Search

☐ add wildcard '*'

Browse:
- by documentation entry
- by ProRule description
- by taxonomic scope
- by number of positive hit

PROSITE tools

Scan a sequence against PROSITE patterns and profiles - quick scan

(Output includes graphical view and feature detection)

Enter your sequence or a UniProtKB (Swiss-Prot or TrEMBL) ID or AC [help]:

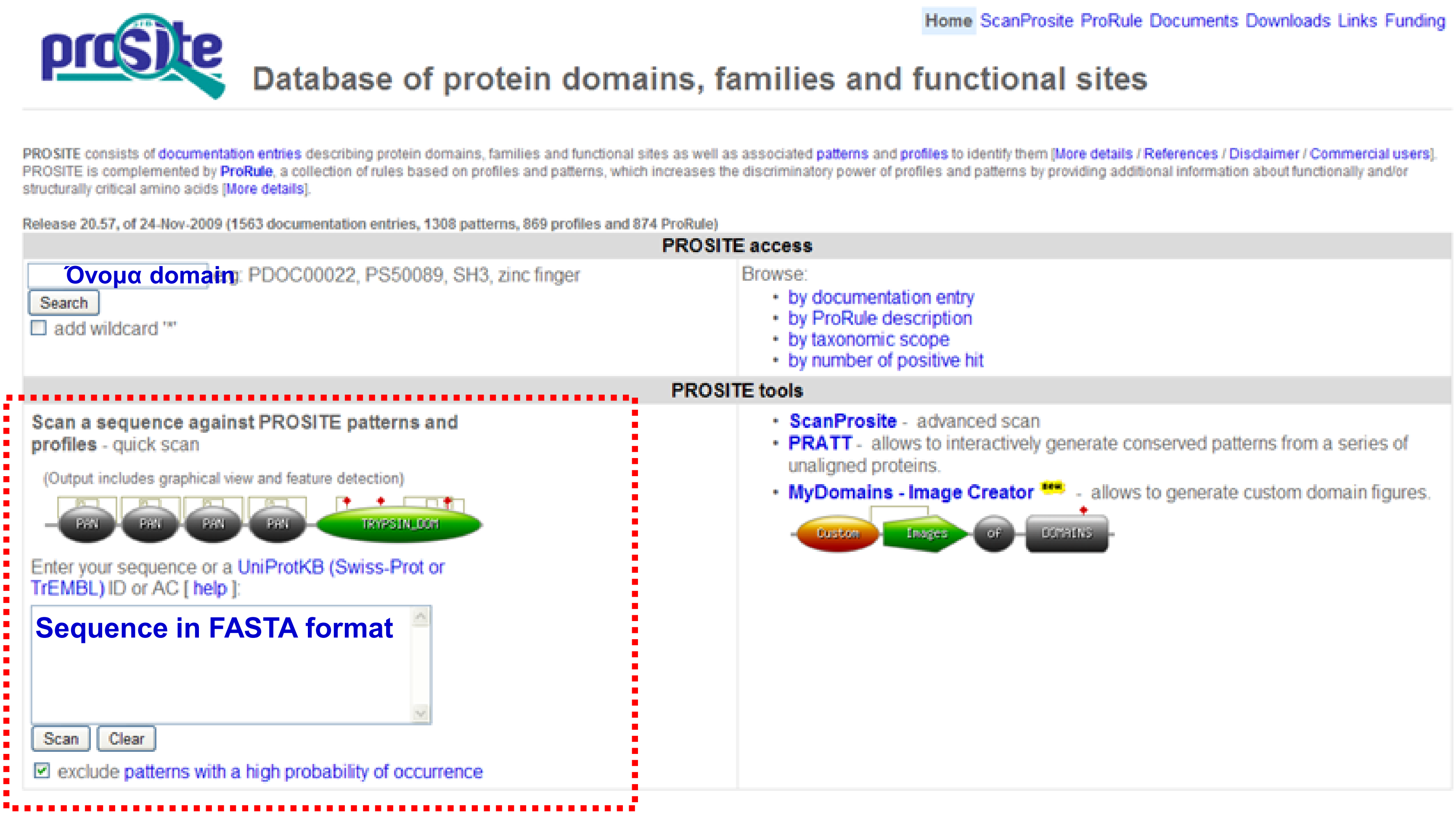
Sequence in FASTA format

Scan Clear

☑ exclude patterns with a high probability of occurrence

- **ScanProsite** - advanced scan
- **PRATT** - allows to interactively generate conserved patterns from a series of unaligned proteins.
- **MyDomains - Image Creator** new - allows to generate custom domain figures.

PROSITE (3/4)



Database of protein domains, families and functional sites

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More details / References / Disclaimer / Commercial users]. PROSITE is complemented by ProRule, a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More details].

Release 20.57, of 24-Nov-2009 (1563 documentation entries, 1308 patterns, 869 profiles and 874 ProRule)

PROSITE access

Όνομα domain: PDOC00022, PS50089, SH3, zinc finger

Search

☐ add wildcard '*'

Browse:

- by documentation entry
- by ProRule description
- by taxonomic scope
- by number of positive hit

PROSITE tools

Scan a sequence against PROSITE patterns and profiles - quick scan

(Output includes graphical view and feature detection)

Enter your sequence or a UniProtKB (Swiss-Prot or TrEMBL) ID or AC [help]:

Sequence in FASTA format

Scan Clear

☑ exclude patterns with a high probability of occurrence

- **ScanProsite** - advanced scan
- **PRATT** - allows to interactively generate conserved patterns from a series of unaligned proteins.
- **MyDomains - Image Creator** new - allows to generate custom domain figures.

PROSITE (4/4)

hits by profiles: [5 hits (by 1 profile) on 1 sequence]

Hits by PS50268 **CADHERIN_2** *Cadherins domain profile :*

USERSEQ1 (882 aa)

180 - 262: score = 20.152

```
------------------------SNKD--KEGKVFYSITGQGADTppvGVFIIERETGW
LKVTEPLDRERIATYTLFSHAVSSNGNAVEDPMEILITVTDQNDNKPEF
```

263 - 375: score = 29.162

```
TQEVFKGSVMEGALPGTSVMEVTATDADDdvnTYNAAIAYTILSQDPELpdkNMFTINRN
TgVISVVTTGLDRESFPTYTLVVQAADLQGEGLSTTATAVITVTDTNDNPPIF
```

376 - 486: score = 21.402

```
NPTTYKGQVPENEANVVI-TTLKVTDADApnTPAWEAVYTILNDDG--GQFVVTTNPvnn
dGILKTAKGLDFEAKQQYILHVAVTNVVpfeVSLTTSTATVTVDVLDVNE-APIF
```

487 - 595: score = 23.682

```
VPPEKRVEVSEDFGVGQEITSYTAQEPDTFMEQKITYRIWRDTA--NWLEINPDTGAIST
RAELDREdfehvKNSTYTALIIATDNGSPVATGTGTLLLILSDVNDNAPIP
```

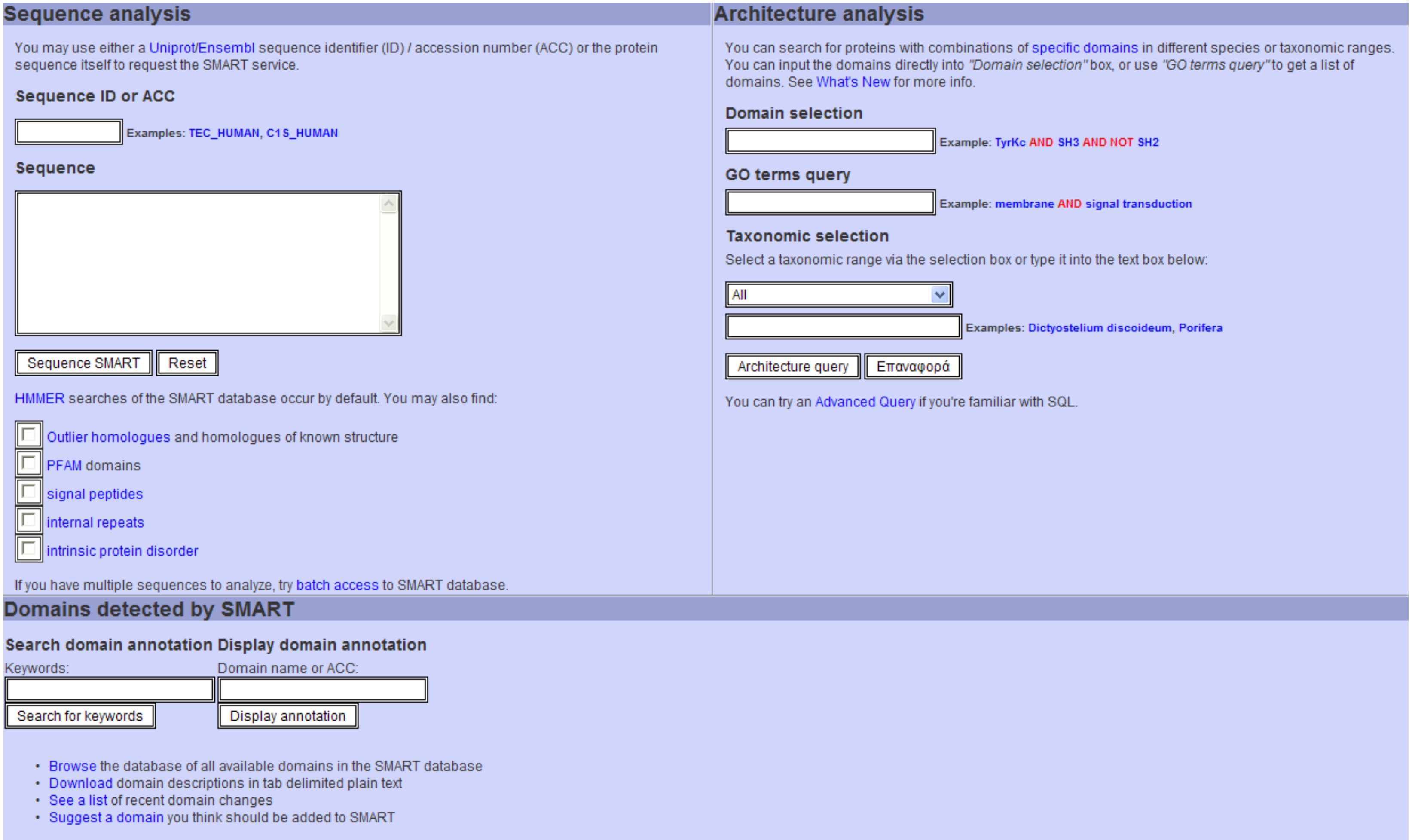
594 - 702: score = 8.840

```
IPEPRTIFFCERNPKPQV---INIIDADLpPNTSPFTAELTHGASA--NWTIQYNDptqE
SIILKPKMALE-VGDYKINLKLMDN--QNKDQVTTLEVSVCDCEGAagvcrkaQPVE
```

SMART (**S**imple **M**odular **A**rchitecture **R**esearch **T**ool)

- Αναγνώριση και σχολιασμό domains.
- «Αρχιτεκτονική» των domains.
- Καταγεγραμμένες περισσότερες από 500 οικογένειες domains:
  - Μετάδοση βιολογικού σήματος.
  - Εξωκυττάρια περιοχή πρωτεϊνών.
  - Ενδοκυττάρια περιοχή πρωτεϊνών.
- Καταχώρηση σύμφωνα με:
  - Εξέλιξη, ομοιότητες/διαφορές μεταξύ διαφορετικών ειδών.
  - Λειτουργία – ρόλος.
  - Τριτοταγής δομή (στο χώρο).
  - Κρίσιμα κατάλοιπα (αμινοξέα) για τη λειτουργία της πρωτεΐνης.

SMART Interface



Sequence analysis

You may use either a Uniprot/Ensembl sequence identifier (ID) / accession number (ACC) or the protein sequence itself to request the SMART service.

Sequence ID or ACC

Examples: TEC_HUMAN, C1S_HUMAN

Sequence

HMMER searches of the SMART database occur by default. You may also find:

- ☐ Outlier homologues and homologues of known structure
- ☐ PFAM domains
- ☐ signal peptides
- ☐ internal repeats
- ☐ intrinsic protein disorder

If you have multiple sequences to analyze, try batch access to SMART database.

Architecture analysis

You can search for proteins with combinations of specific domains in different species or taxonomic ranges. You can input the domains directly into *"Domain selection"* box, or use *"GO terms query"* to get a list of domains. See What's New for more info.

Domain selection

Example: TyrKc AND SH3 AND NOT SH2

GO terms query

Example: membrane AND signal transduction

Taxonomic selection

Select a taxonomic range via the selection box or type it into the text box below:

All

Examples: Dictyostelium discoideum, Porifera

You can try an Advanced Query if you're familiar with SQL.

Domains detected by SMART

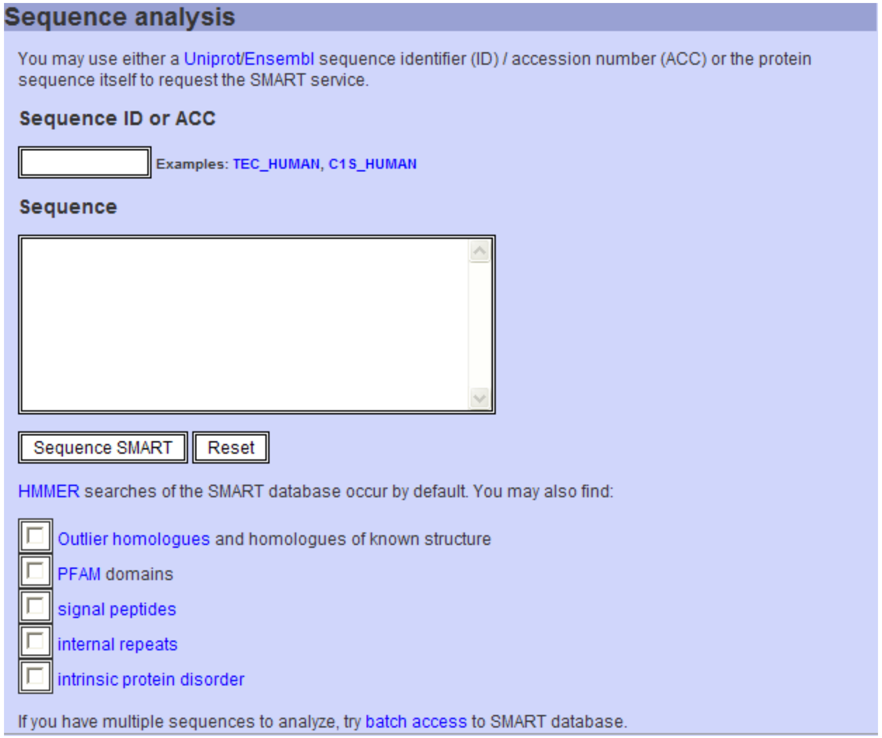
Search domain annotation **Display domain annotation**

Keywords:

Domain name or ACC:

- Browse the database of all available domains in the SMART database
- Download domain descriptions in tab delimited plain text
- See a list of recent domain changes
- Suggest a domain you think should be added to SMART

SMART – Ανάλυση αλληλουχίας



Sequence analysis

You may use either a Uniprot/Ensembl sequence identifier (ID) / accession number (ACC) or the protein sequence itself to request the SMART service.

Sequence ID or ACC

Examples: TEC_HUMAN, C1S_HUMAN

Sequence

HMMER searches of the SMART database occur by default. You may also find:

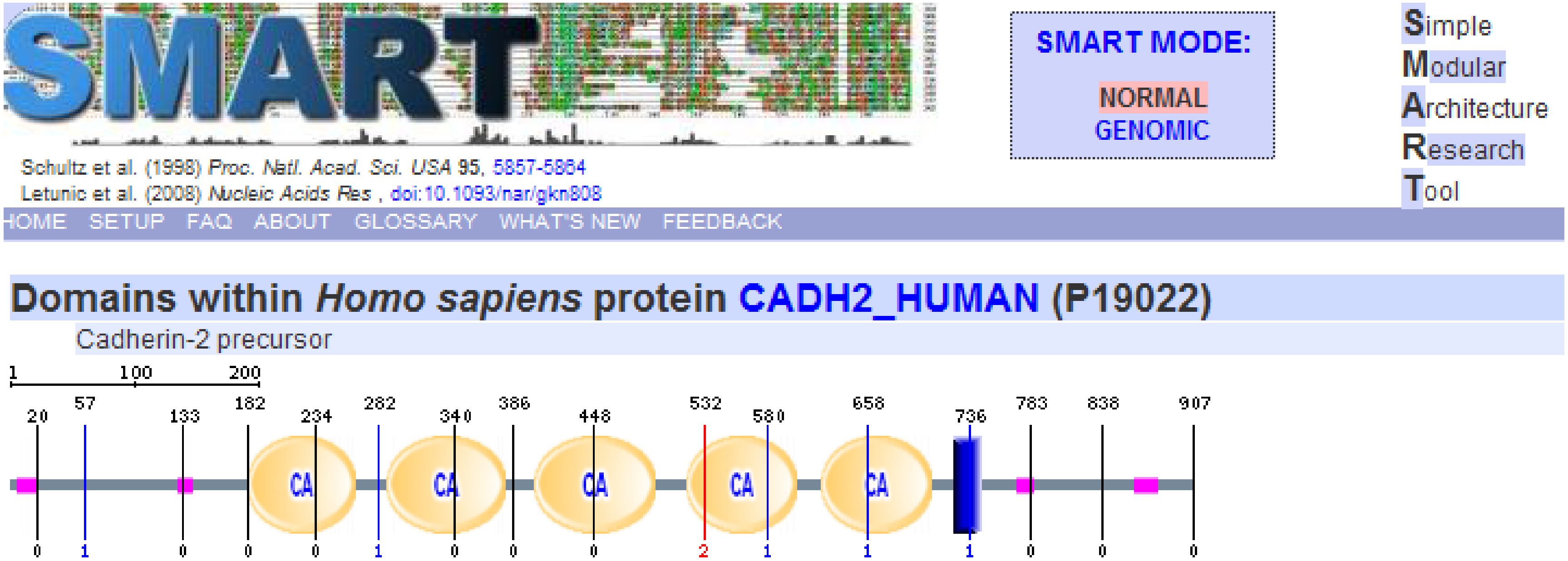
- ☐ Outlier homologues and homologues of known structure
- ☐ PFAM domains
- ☐ signal peptides
- ☐ internal repeats
- ☐ intrinsic protein disorder

If you have multiple sequences to analyze, try batch access to SMART database.

- Εύρεση πιθανών domains / motifs σε συγκεκριμένη αλληλουχία πρωτεΐνης.
- Εισαγωγή αλληλουχίας σε FASTA format.

Εισαγωγή αλληλουχίας καντχερίνης-2 σε FASTA format

SMART – Παράδειγμα cadherin



Schultz et al. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5857-5864

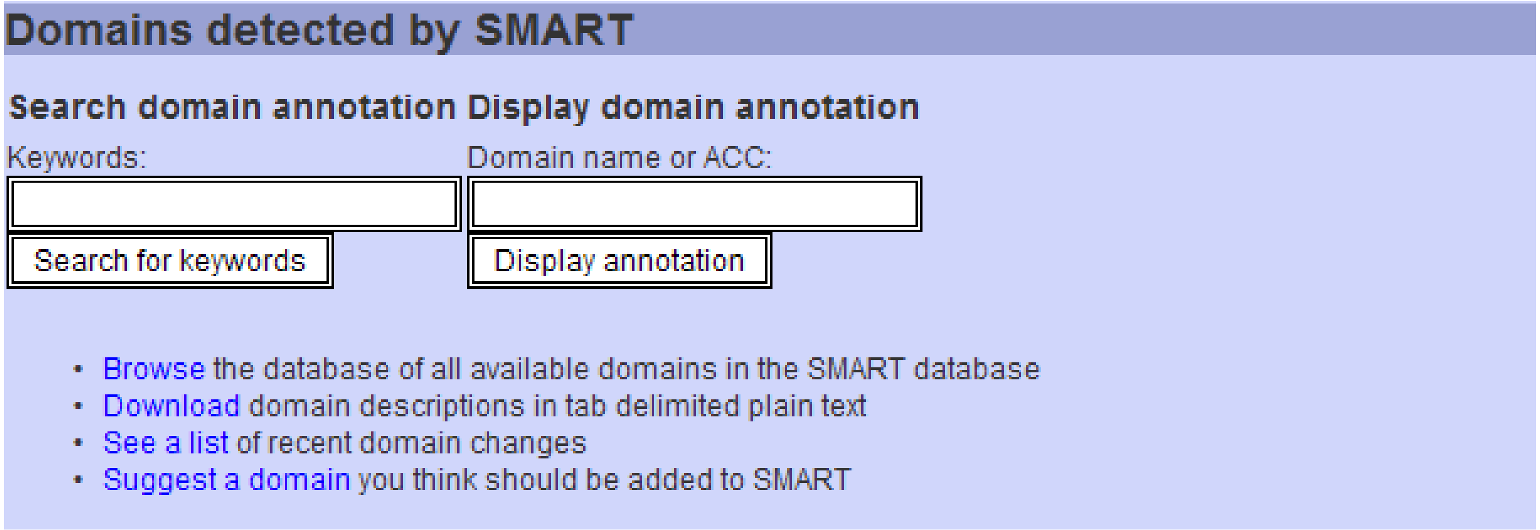
Letunic et al. (2008) *Nucleic Acids Res* , doi:10.1093/nar/gkn808

Domains within *Homo sapiens* protein CADH2_HUMAN (P19022)

Cadherin-2 precursor

SMART – Βάση τομέων (domains) πρωτεΐνης

- Εύρεση πληροφοριών για συγκεκριμένα domains.
- Λίστα με όλα τα καταχωρημένα domains.



Domains detected by SMART

Search domain annotation **Display domain annotation**

Keywords:

Domain name or ACC:

- Browse the database of all available domains in the SMART database
- Download domain descriptions in tab delimited plain text
- See a list of recent domain changes
- Suggest a domain you think should be added to SMART

“Ten Important Bioinformatics Databases”

- GenBank http://www.ncbi.nlm.nih.gov nucleotide sequences.
- Ensembl http://www.ensembl.org human/mouse genome (and others).
- PubMed http://www.ncbi.nlm.nih.gov literature references.
- NR http://www.ncbi.nlm.nih.gov protein sequences.
- SWISS-PROT http://www.expasy.ch protein sequences.
- InterPro http://www.ebi.ac.uk protein domains.
- OMIM http://www.ncbi.nlm.nih.gov genetic diseases.
- Enzymes http://www.chem.qmul.ac.uk enzymes.
- PDB http://www.rcsb.org/pdb/ protein structures.
- KEGG http://www.genome.ad.jp metabolic pathways.

Source: *Bioinformatics for Dummies*

Χρήσιμες πηγές

- **Gene cards (only human):**

http://www.genecards.org/index.shtml

  - Χαρακτηριστικά και πολλαπλές ονομασίες γονιδίου.
  - Γενική λειτουργία πρωτεΐνης που εκφράζεται από γονίδιο.
  - Πληροφορίες για την αλληλουχία πρωτεΐνης.
  - Δομή πρωτεΐνης.
  - Ρόλος πρωτεΐνης στη φυσιολογία του κυττάρου.
  - Επίπεδα έκφρασης σε διάφορους ιστούς.
  - Βιβλιογραφία.
- http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html
  - Επεξεργασία DNA και πρωτεϊνικών αλληλουχιών, π.χ. **FASTA format.**

Μεταλλάξεις - Mutations

Normal DNA sequence	**ATC-CCT-AGT-AAA**
Mutated DNA sequence	**ATC-CTT-AGT-AAG**
Normal protein sequence	**Isoleucine – Proline – Serine – Lysine**
Mutated protein sequence	**Isoleucine – Leucine – Serine – Lysine**

Επίσης…

- **a.a.:** Amino acids = Αμινοξέα (Πρωτεΐνη).
- **bp:** Base pairs = Ζεύγη βάσεων, νουκλεοτιδίων (DNA).
- **mRNA:** messenger RNA, αλληλουχία νουκλεοτιδίων, η θυμίνη (T) έχει αντικατασταθεί από ουρακίλη (U).
- **Partial, fragment:** Μέρος της αλληλουχίας (πρωτεΐνη ή DNA).
- **Gene:** Γονίδιο, κομμάτι DNA από το οποίο προκύπτει μία πρωτεΐνη.
- **Start codon:** ATG. Κωδικοποιεί τη μεθειονίνη.
- **Stop codon:** TAA (UAA), TAG (UAG), TGA (UGA). Δεν κωδικοποιεί αμινοξύ.
- **CDS:** Coding sequence, Κωδικοποιούσα αλληλουχία, Γονίδιο. Αρχίζει με ATG και τελειώνει με ένα κωδικόνιο τερματισμού.

Τέλος Ενότητας

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημείωμα Αναφοράς

- Copyright Πανεπιστήμιο Δυτικής Μακεδονίας, Τμήμα Μηχανικών Πληροφορικής και Τηλεπικοινωνιών, Αγγελίδης Παντελής. «**Βιοπληροφορική**». Έκδοση: 1.0. Κοζάνη 2015. Διαθέσιμο από τη δικτυακή διεύθυνση:

  https: //eclass.uowm.gr/courses/ICTE102/

Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Όχι Παράγωγα Έργα Μη Εμπορική Χρήση 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».

[1] http://creativecommons.org/licenses/by-nc-nd/4.0/

Ως Μη Εμπορική ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό

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.

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Εικόνες:

- https: //www.cartoonstock.com/
- Bioinformatics Computing, Book by Bryan Bergeron