



Βιοπληροφορική

Ενότητα 3:

Βάσεις Δεδομένων (2/3), 1 ΔΩ

Τμήμα: **Βιοτεχνολογίας**

Όνομα καθηγητή: **Τ. Θηραίου**



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης





Μαθησιακοί Στόχοι

- Παρουσίαση των κατηγοριών των βιολογικών ΒΔ.
- Αναφορά στις χαρακτηριστικότερες ΒΔ
 - νουκλεοτιδικών ακολουθιών.
 - αμινοξικών ακολουθιών.
 - δομών μακρομορίων.



Λέξεις Κλειδιά

- Λέξεις κλειδιά: ΒΔ νουκλεοτιδικών ακολουθιών, μορφοποίηση FASTA, ΒΔ αμινοξικών ακολουθιών, ΒΔ πρωτεϊνικών δομών.
- Key words: Nucleotide sequence database, FASTA format, Protein sequence database, Protein structure database.



Κατηγορίες ΒΔ

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



Πρωτογενείς ΒΔ 1/13

- **ΒΔ Νουκλεοτιδικών ακολουθιών:**
 - DNA Data Bank of Japan (DDBJ)
 - Center for Information Biology (CIB).
<http://www.ddbj.nig.ac.jp/>
 - GenBank
 - National Center for Biotechnology Information (NCBI).
<http://www.ncbi.nlm.nih.gov/genbank/>
 - European Nucleotide Archive (ENA)
 - European Bioinformatics Institute (EBI).
<http://www.ebi.ac.uk/embl/index.html>



Πρωτογενείς ΒΔ 2/13

- ΒΔ Νουκλεοτιδικών ακολουθιών:
 - International Nucleotide Sequence Database Collaboration. <http://www.insdc.org/>
 - συνεργασία των DDBJ, GenBank, ENA.
 - κοινοί κανόνες για την ταξινόμηση και το σχολιασμό των δεδομένων.
 - καθημερινή ανταλλαγή εγγραφών.



Πρωτογενείς ΒΔ 3/13

- ΒΔ Νουκλεοτιδικών ακολουθιών:
 - μορφοποίηση GenBank.

LOCUS: X64011 756 bp DNA linear BCT 26-SEP-2006
DEFINITION: *Listeria ivanovii* sod gene for superoxide dismutase.
ACCESSION: X64011 S78972
VERSION: X64011.1 GI:44010
KEYWORDS: sod gene; superoxide dismutase.
SOURCE: *Listeria ivanovii*
ORGANISM: *Listeria ivanovii*
Bacteria; Firmicutes; Bacillales; Listeriaceae; *Listeria*.
REFERENCE: 1
AUTHORS: Haas,A. and Goebel,W.
TITLE: Cloning of a superoxide dismutase gene from *Listeria ivanovii* by functional complementation in *Escherichia coli* and characterization of the gene product.
JOURNAL: Mol. Gen. Genet. 231 (2), 313-322 (1992).
PUBMED: 1736100.



Πρωτογενείς ΒΔ 4/13

- ΒΔ Νουκλεοτιδικών ακολουθιών:
 - μορφοποίηση GenBank.

```
FEATURES             Location/Qualifiers
    source             1..756
                       /organism="Listeria ivanovii"
                       /mol_type="genomic DNA"
                       /db_xref="taxon:1638"
    CDS                109..717
                       /gene="sod"
                       /EC_number="1.15.1.1"
                       /codon_start=1
                       /transl_table=11
                       /product="superoxide dismutase"
                       /protein_id="CAA45406.1"
                       /db_xref="InterPro:IPR001189"
                       /db_xref="UniProtKB/Swiss-Prot:P28763"
                       /translation="MTYELPKLPYTYDALEPNFDKETMEIHYTEKHHNIYVTKLNEAVS
GHAEELASKPGEELVANLDSVP EEIRGAVRNHGGGHANHTLFWSSLS PNGGGAP
```




Πρωτογενείς ΒΔ 5/13

- ΒΔ Νουκλεοτιδικών ακολουθιών:
 - μορφοποίηση GenBank.

ORIGIN

```
1  cgttatttaa ggtgttacat agttctatgg aataggggtc tataccttc gccttacaat
61  gtaatttctt ttcacataaa taataaaca tccgaggagg aattttaat gacttacgaa
121 ttaccaaagt taccttatac ttatgatgct ttggagccga atttgataa agaaacaatg
181 gaaattcact atacaaagca ccacaatatt tatgtaacaa aactaaatga agcagtctca
241 ggacacgcag aacttgcaag taaacctggg gaagaattag ttgctaact agatagcggt
301 cctgaagaaa ttcgtggcgc agtacgtaac cacgggtggg gacatgctaa ccatacttta
361 ttctgggtcta gtcttagccc aatgggtggg ggtgctccaa ctggtaactt aaaagcagca
421 atcgaaagcg aattcggcac atttgatgaa ttcaaagaaa aattcaatgc ggcagctgcg
481 gctcgttttg gttcaggatg ggcatggcta gtagtgaaca atggtaaact agaaattggt
541 tccactgcta accaagattc tccacttagc gaaggtaaaa ctccagtctt tggcttagat
601 gtttgggaac atgcttatta tcttaaattc caaaaccgct gtcctgaata cattgacaca
661 ttttggaatg taattaactg ggatgaacga aataaacgct ttgacgcagc aaaataatta
721 tcgaaaggct cacttaggtg ggtcttttta ttctca
```

//



Πρωτογενείς ΒΔ 6/13

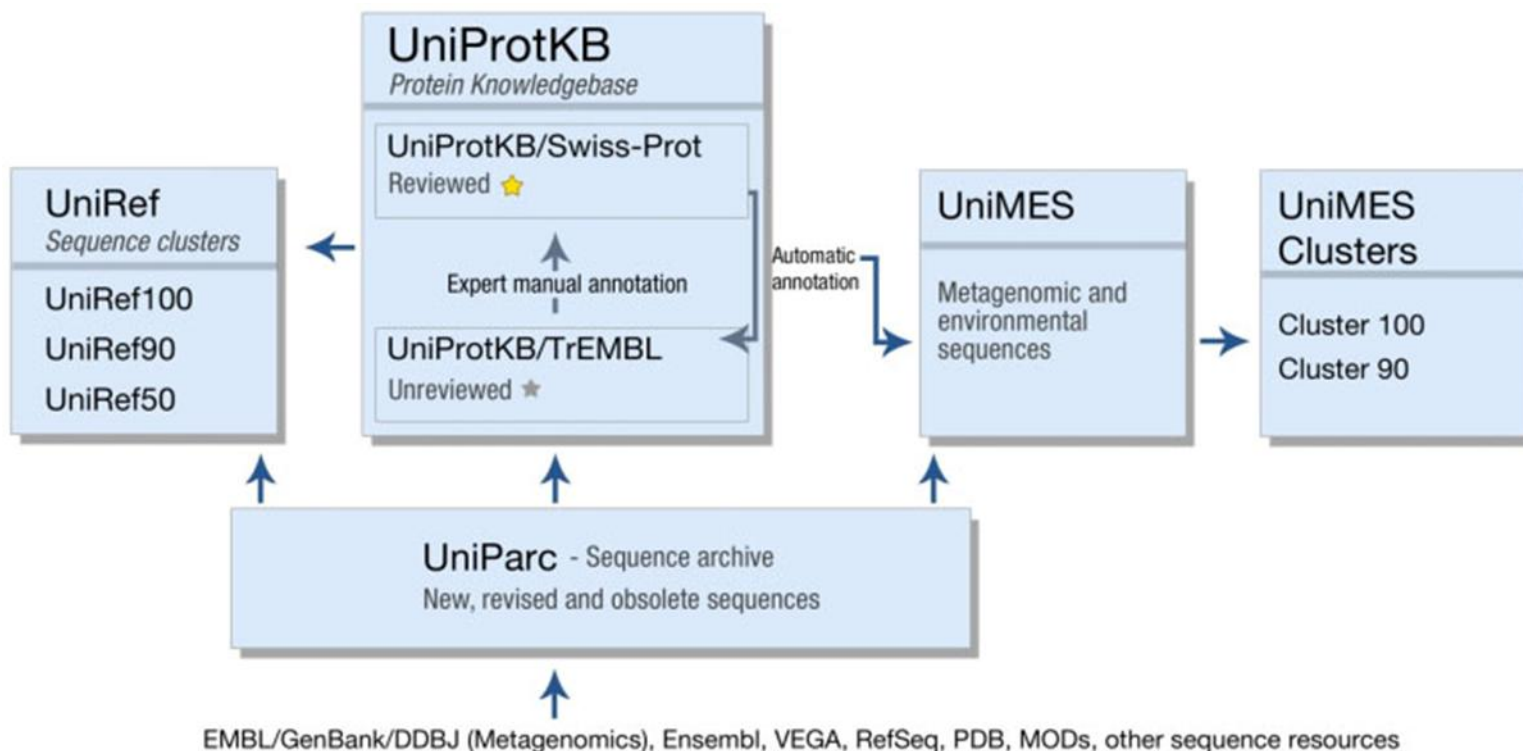
- ΒΔ Νουκλεοτιδικών ακολουθιών:
 - **μορφοποίηση FASTA**
 - πρώτη γραμμή:
 - **ταυτότητα ακολουθίας** (επικεφαλίδα).
 - επόμενες γραμμές:
ακολουθία βάσεων ή αμινοξέων (με τους κωδικούς του ενός γράμματος).

```
>gi|44010|emb|X64011.1| Listeria ivanovii sod gene for superoxide dismutase  
CGTTATTTAAGGTGTTACATAGTTCTATGGAAATAGGGTCTATACCTTTTCGC  
CTTACAATGTAATTTCTTTTCACATAAATAATAACAATCCGAGGAGGAATT  
TTAATGACTTACGAATTACCAAAAATTACCTTATAC
```



Πρωτογενείς ΒΔ 7/13

- ΒΔ Αμινοξικών ακολουθιών:
 - UniProt. <http://www.uniprot.org/>
 - Swiss-Prot / TrEMBL / PIR-PSD.

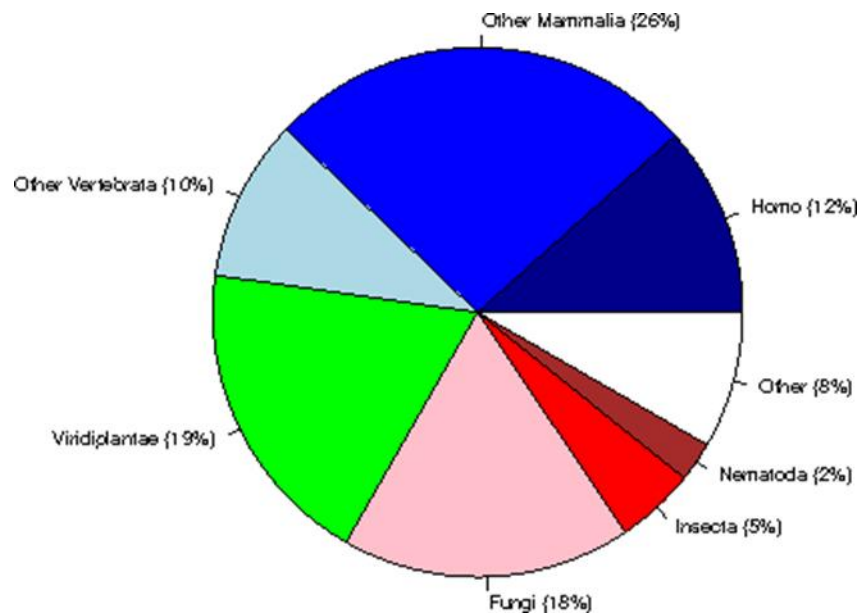
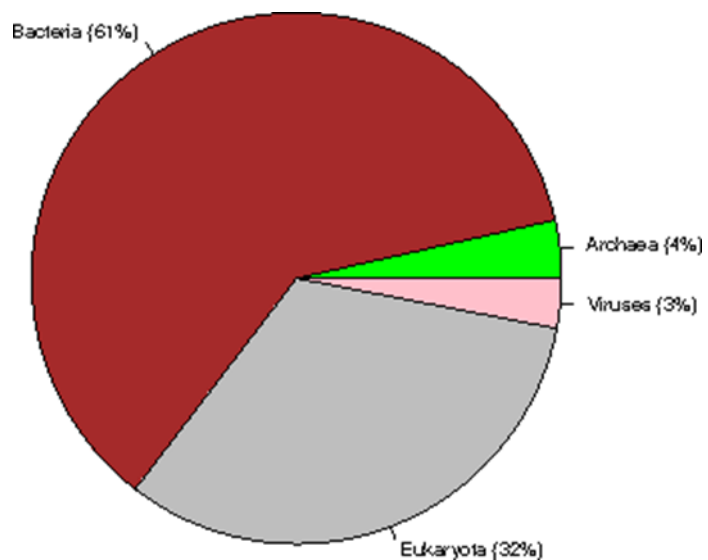




Πρωτογενείς ΒΔ 8/13

● ΒΔ Αμινοξικών ακολουθιών:

– UniProt. <http://www.uniprot.org/>



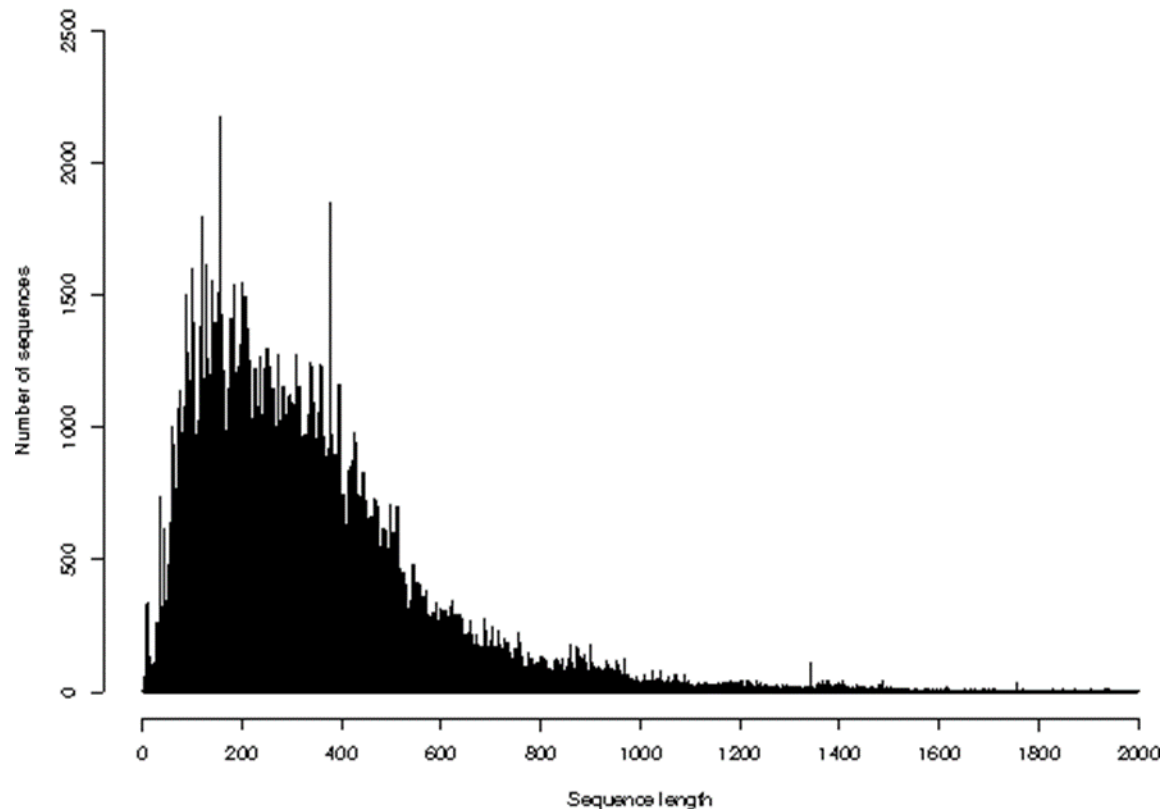


Πρωτογενείς ΒΔ 9/13

- ΒΔ Αμινοξικών ακολουθιών:

- UniProt. <http://www.uniprot.org/>

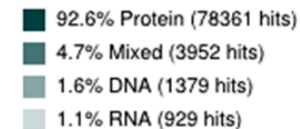
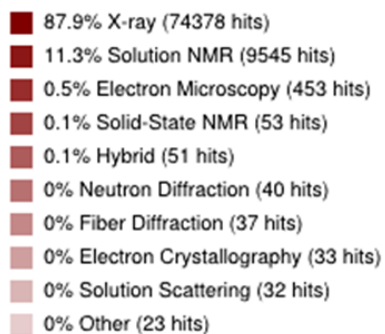
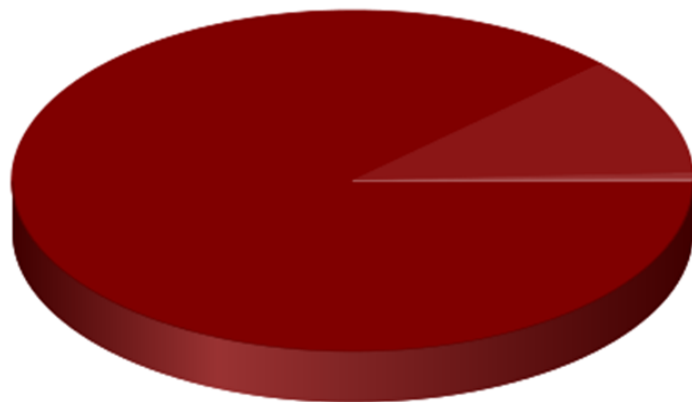
Length distribution of the sequences





Πρωτογενείς ΒΔ 10/13

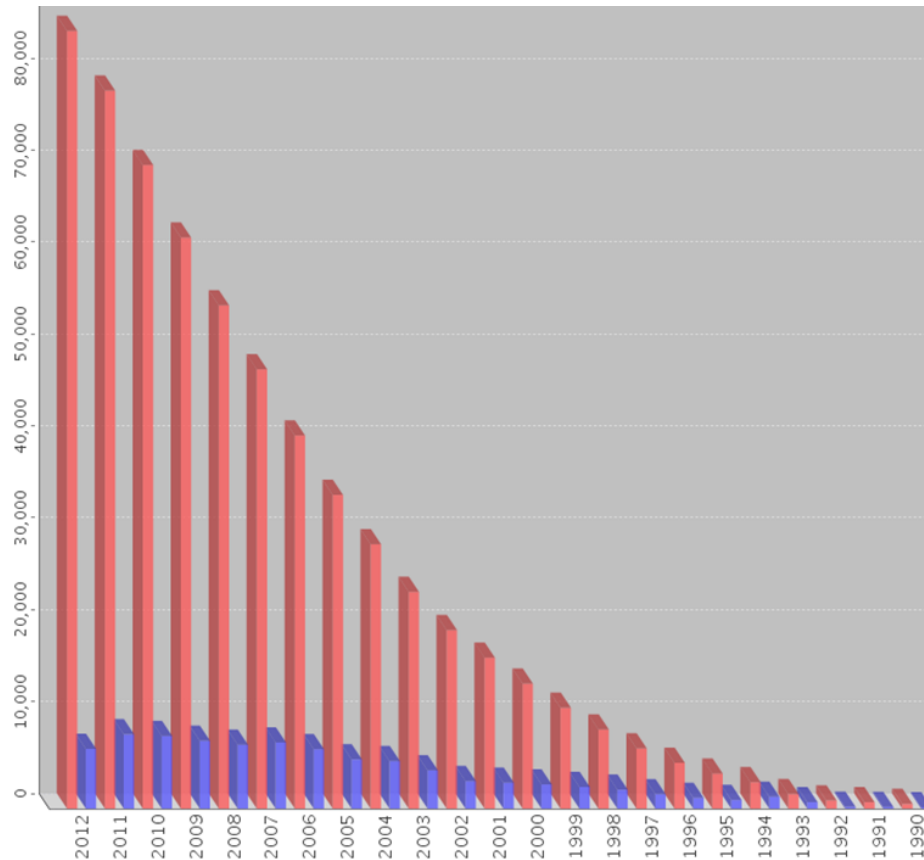
- ΒΔ Δομών μακρομορίων και μικρών μορίων:
 - PDB. <http://www.rcsb.org/pdb/home/home.do>





Πρωτογενείς ΒΔ 11/13

- ΒΔ Δομών μακρομορίων και μικρών μορίων:
 - PDB. <http://www.rcsb.org/pdb/home/home.do>





Πρωτογενείς ΒΔ 12/13

- ΒΔ Δομών μακρομορίων και μικρών μορίων:
 - **μορφοποίηση PDB.**

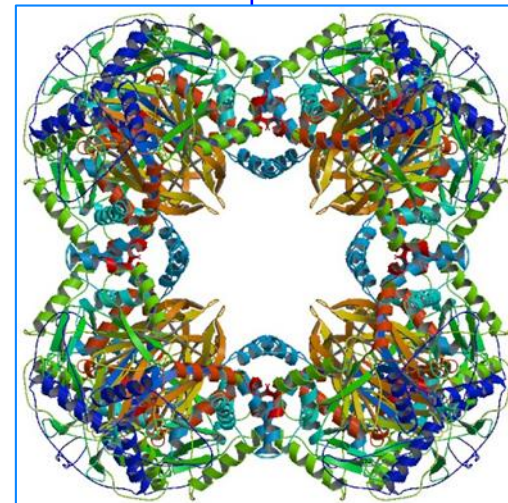
```
HEADER    DIHYDROLIPOAMIDE ACETYLTRANSFERASE      16-DEC-92    1EAA
TITLE     ATOMIC STRUCTURE OF THE CUBIC CORE OF THE PYRUVATE DEHYDROGENASE
TITLE     2 MULTIENZYME COMPLEX
COMPND    MOL_ID: 1;
COMPND    2 MOLECULE: DIHYDROLIPOYL-TRANSACETYLASE;
COMPND    3 CHAIN: A;
COMPND    4 EC: 2.3.1.12;
SOURCE    MOL_ID: 1;
SOURCE    2 ORGANISM_SCIENTIFIC: AZOTOBACTER VINELANDII;
SOURCE    3 ORGANISM_TAXID: 354
KEYWDS    DIHYDROLIPOAMIDE ACETYLTRANSFERASE
EXPDTA    X-RAY DIFFRACTION
AUTHOR    A.MATTEVI,W.G.J.HOL
REVDAT    4 13-JUL-11 1EAA 1 VERSN
REVDAT    3 24-FEB-09 1EAA 1 VERSN
REVDAT    2 01-APR-03 1EAA 1 JRNL
REVDAT    1 31-OCT-93 1EAA 0
JRNL      AUTH    A.MATTEVI,G.OBMOLOVA,K.H.KALK,A.TEPLYAKOV,W.G.HOL
JRNL      TITL    CRYSTALLOGRAPHIC ANALYSIS OF SUBSTRATE BINDING AND CATALYSIS
JRNL      TITL 2  IN DIHYDROLIPOYL TRANSACETYLASE (E2P).
JRNL      REF     BIOCHEMISTRY                      V. 32 3887 1993
```




Πρωτογενείς ΒΔ 13/13

- ΒΔ Δομών μακρομορίων και μικρών μορίων:
 - μορφοποίηση PDB.

ATOM	1	N	ILE	A	395	44.038	49.117	47.927	1.00	38.09	N
ATOM	2	CA	ILE	A	395	42.911	48.771	47.065	1.00	37.62	C
ATOM	3	C	ILE	A	395	41.806	49.715	47.553	1.00	38.16	C
ATOM	4	O	ILE	A	395	41.597	49.893	48.754	1.00	38.35	O
ATOM	5	CB	ILE	A	395	42.420	47.243	47.174	1.00	37.56	C
ATOM	6	CG1	ILE	A	395	41.141	47.074	47.963	1.00	38.29	
ATOM	7	CG2	ILE	A	395	43.486	46.413	47.845	1.00	39.24	
ATOM	8	CD1	ILE	A	395	40.674	45.622	47.993	1.00	40.41	
ATOM	9	N	PRO	A	396	41.197	50.465	46.656	1.00	39.07	
ATOM	10	CA	PRO	A	396	40.317	51.575	46.973	1.00	38.84	
ATOM	11	C	PRO	A	396	39.120	50.925	47.584	1.00	38.21	
ATOM	12	O	PRO	A	396	38.734	49.837	47.193	1.00	39.18	
ATOM	13	CB	PRO	A	396	39.949	52.236	45.689	1.00	39.42	
ATOM	14	CG	PRO	A	396	40.778	51.481	44.659	1.00	41.29	
ATOM	15	CD	PRO	A	396	41.011	50.108	45.262	1.00	40.06	
ATOM	16	N	PRO	A	397	38.536	51.543	48.551	1.00	37.94	
ATOM	17	CA	PRO	A	397	37.376	50.997	49.185	1.00	38.05	
ATOM	18	C	PRO	A	397	36.138	51.535	48.488	1.00	37.27	
ATOM	19	O	PRO	A	397	36.120	52.568	47.798	1.00	37.03	O
ATOM	20	CB	PRO	A	397	37.570	51.444	50.585	1.00	38.31	C
ATOM	21	CG	PRO	A	397	38.038	52.882	50.351	1.00	38.61	C
ATOM	22	CD	PRO	A	397	39.013	52.751	49.196	1.00	38.71	C





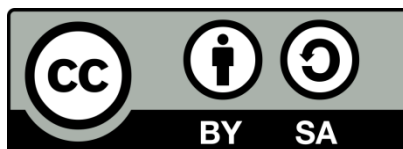
Βιβλιογραφία

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



Άδειες Χρήσης

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





Χρηματοδότηση

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



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



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

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



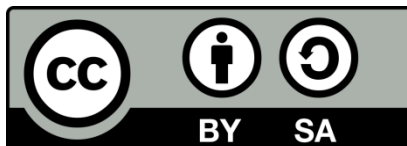
Σημείωμα Αναφοράς

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



Σημείωμα Αδειοδότησης

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



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

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

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

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

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



Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

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