

Βιοπληροφορική

Ενότητα 1: Εισαγωγή στη Βιοπληροφορική

Αν. καθηγήτης Αγγελίδης Παντελής

e-mail: paggelidis@uowm.gr

ΕΕΔΙΠ Μπέλλου Σοφία

e-mail: sbellou@uowm.gr

Τμήμα Μηχανικών Πληροφορικής και Τηλεπικοινωνιών



Άδειες Χρήσης

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



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



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



Δομή μαθήματος

1. Εισαγωγή – Τι είναι η βιοπληροφορική.
2. Βασικές έννοιες βιολογίας – Νουκλεοτιδικές αλληλουχίες.
3. Βασικές έννοιες βιολογίας – Πρωτεϊνικές αλληλουχίες.
4. Βιβλιογραφικές βάσεις δεδομένων (Scholar Google, Scopus, PubMed).
5. Βιολογικές βάσεις δεδομένων.
6. Ανάλυση ομοιότητας αλληλουχιών.
7. Πρότυπα, μοτίβα και αποτυπώματα στις αλληλουχίες των βιολογικών μακρομορίων.
8. Φυλογενετική ανάλυση - Φυλογενετικά δέντρα.
9. Βασικές αρχές εξέλιξης.
10. Πρωτεωμική.
11. Ανάλυση των πρωτεϊνικών αλληλεπιδράσεων με τη βιοπληροφορική.
12. Ιατρική Πληροφορική και Βιοπληροφορική.

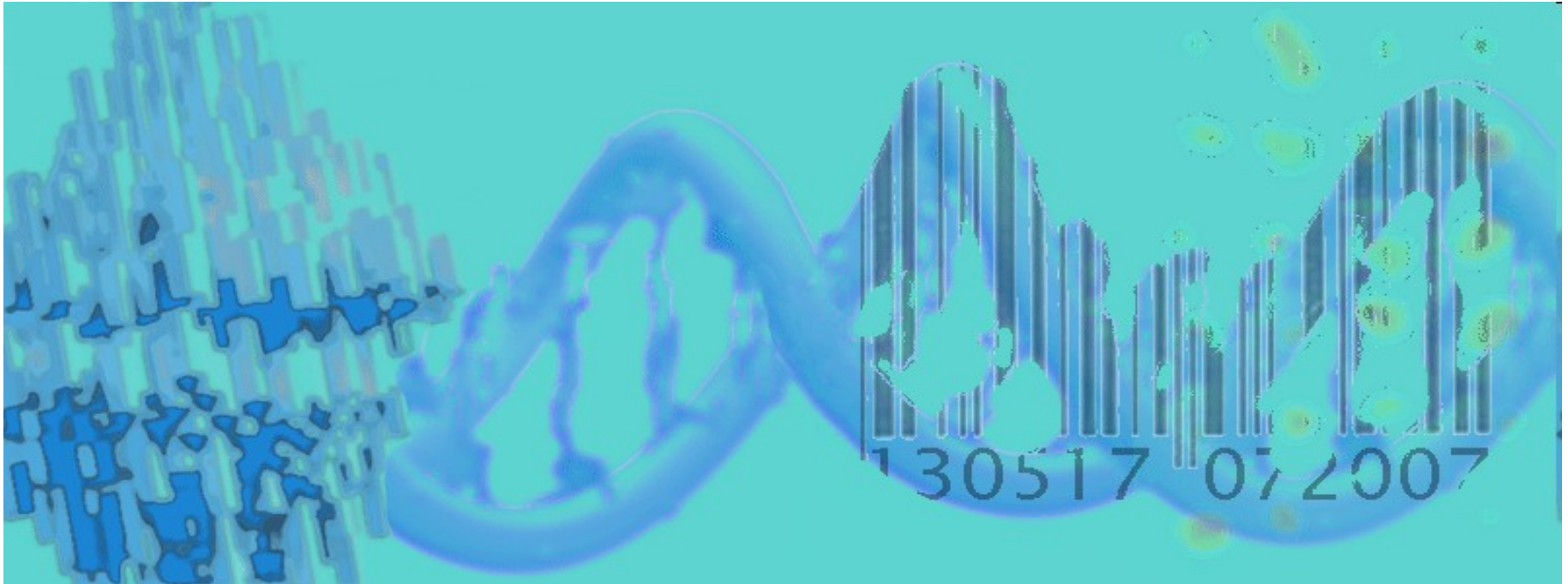


Διδάσκοντες - Ώρες μαθήματος

- Διδάσκων:
Παντελής Αγγελίδης.
- Εισηγήσεις Θεωρίας – Εργαστήριο:
Σοφία Μπέλλου.
- Θεωρία: Τετάρτη 18:00-20:00.
- Εργαστήριο: Δευτέρα 18:00-20:00 (ανακοίνωση).



Τι είναι η βιοπληροφορική;



Πηγή: <http://blog.com.mk/send/121903>



Επιστήμη των Υπολογιστών και Βιολογία: Μια αμφίδρομη σχέση

- Υπολογιστές και πληροφοριακά συστήματα χρησιμοποιούνται για να συγκεντρώσουν, να αναλύσουν και να επιλύσουν βιολογικά προβλήματα.
 - Large database and data storage.
 - Data mining.
- Βιολογικά συστήματα χρησιμοποιούνται ως πρότυπα με σκοπό τη βελτίωση των υπολογιστών και των πληροφοριακών συστημάτων:
 - Genetic algorithms.
 - Neural nets.



(Μοριακή) Βιο-πληροφορική

- Η βιοπληροφορική εφαρμόζει «τεχνικές πληροφορικής» (προερχόμενες από επιστημονικά πεδία όπως τα **εφαρμοσμένα μαθηματικά, η πληροφορική και η στατιστική**) για να κατανοήσει και να οργανώσει τις πληροφορίες που συνδέονται με τη **μοριακή βιολογία**, σε μία ευρεία κλίμακα.

Ειδικότερα...

Χρησιμοποιεί υπολογιστικές μεθόδους για να μελετήσει και να οργανώσει τα βιολογικά δεδομένα που υπάρχουν διαθέσιμα.

Συγκεντρώνει, αποθηκεύει και αναλύει βιολογικά δεδομένα.

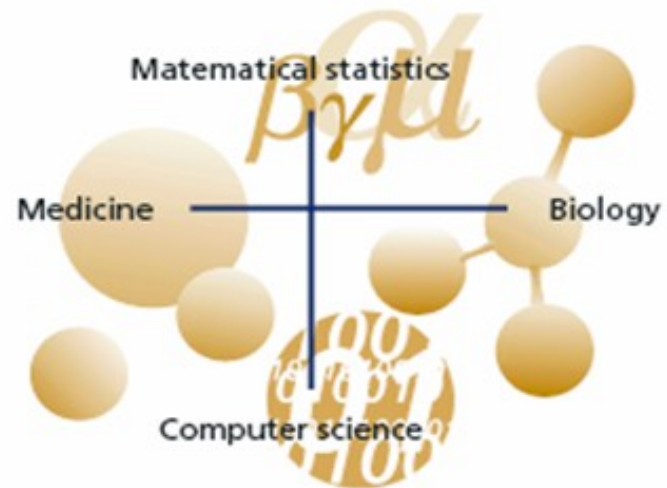
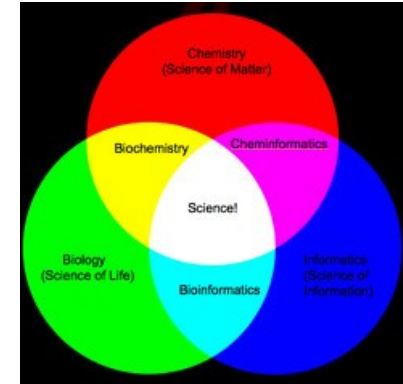


Βιοπληροφορική και συναφείς επιστήμες

- Επιστήμη των υπολογιστών.
- Μαθηματικά και Στατιστική.
- Βιολογία.
- Ιατρική.
- Χημεία.
- Φυσική.

Πηγή εικόνων:

<http://foter.com/Cmyk/>



Ένα μοριακό αλφάβητο

- Τα περισσότερα **βιολογικά μόρια** είναι **πολυμερή**, διατεταγμένες αλυσίδες απλών μορίων που ονομάζονται **μονομερή**.
- Όλα τα μονομερή ανήκουν στην ίδια γενική κατηγορία, αλλά υπάρχουν διάφοροι τύποι με ξεχωριστά και συγκεκριμένα χαρακτηριστικά.
- Πολλά μονομερή μπορούν να ενωθούν μεταξύ τους και να σχηματίσουν ένα μοναδικό και μεγάλο μακρομόριο.
- Η σειρά τοποθέτησης των μονομερών στο μακρομόριο κωδικοποιεί πληροφορία, όπως ακριβώς τα γράμματα σε ένα αλφάβητο.



Βασικοί ορισμοί...

- **Γονίδιο:** Συγκεκριμένα μέρη του DNA με αποθηκευμένη πληροφορία.
- **Γονιδίωμα:** Το σύνολο των γονιδίων ενός οργανισμού.
- **Πρωτεΐνη:** Βιολογικό μακρομόριο με συγκεκριμένη δομή και λειτουργία. Προκύπτει από το αντίστοιχο γονίδιο.

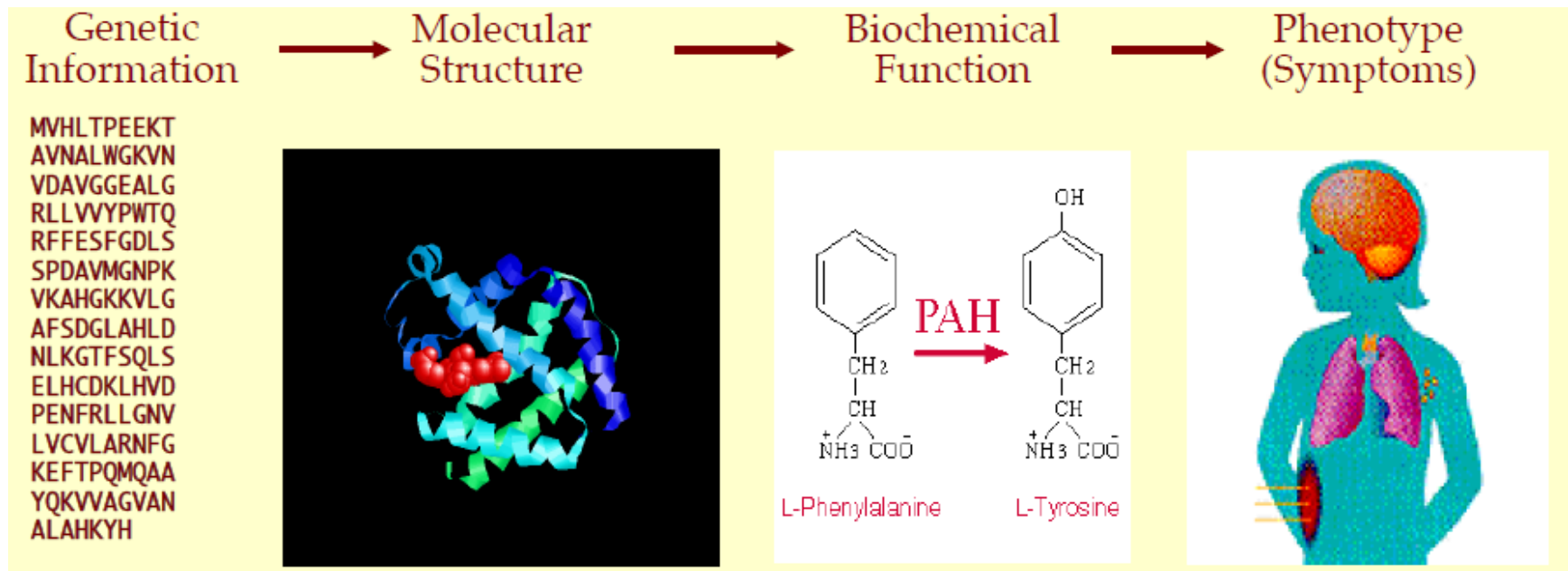


Σύγχρονοι στόχοι της Μοριακής Βιολογίας

- **Αναγνώριση γονιδίων και καθορισμός των λειτουργιών που ρυθμίζουν:** Ταυτοποίηση γονιδίων, αναγνώριση λειτουργιών και ιδιοτήτων τους.
- **Ακολουθοποίηση και σύγκριση των γονιδιωμάτων διαφορετικών οργανισμών:** Πληροφορίες για την εξελεγκτική πορεία των οργανισμών.
- **Κατανόηση της γονιδιακής έκφρασης:** Κατανόηση βασικών κυτταρικών λειτουργιών, όπως η αναπνοή και ο μεταβολισμός.
- **Κατανόηση γενετικών ασθενειών:** Στόχος η αντιμετώπιση ασθενειών με γενετικό υπόβαθρο.
- Η χρήση εργαλείων Πληροφορικής επιλύει πολλά υπολογιστικά προβλήματα, όπως:



Κεντρικό παράδειγμα Βιοπληροφορικής



Επίλυση προβλημάτων με τη χρήση εργαλείων Πληροφορικής

1. Ανάγνωση ακολουθιών DNA (Sequencing & sequence assembly).
2. Σύγκριση ακολουθιών DNA (Sequence alignment).
3. Κατηγοριοποίηση πρωτεϊνών.
4. Εξαγωγή πληροφοριών σχετικά με τη δομή και λειτουργία πρωτεϊνών από τις αντίστοιχες ακολουθίες.
5. Μοντελοποίηση ενός κυττάρου για καλύτερη κατανόηση συγκεκριμένων διεργασιών.



Ομαδοποίηση τύπων δεδομένων

Πηγή Δεδομένων	Μέγεθος Δεδομένων	Εφαρμογές Βιοπληροφορικής
Ακολουθίες DNA	11.5 εκατ. Ακολουθίες (12.5 δις. Βάσεις)	<ul style="list-style-type: none"> - Αναγνώριση introns και exons - Διαχωρισμός coding & non-coding περιοχών
Ακολουθίες Πρωτεϊνών	400.000 ακολουθίες (~300 αμινοξέα για καθένα)	<ul style="list-style-type: none"> - Αλγόριθμοι σύγκρισης ακολουθιών - Ανακάλυψη σημαντικών μοτίβων
Δομές Μακρομορίων	15.000 δομές (~1000 ατομικές συντεταγμένες)	<ul style="list-style-type: none"> - Καθορισμός Δευτερεύουσας δομής - Αλγόριθμοι ταξινόμησης προστάσεων
Άλλα δεδομένα		
Βιβλιογραφία	11 εκατ. αναφορές	<ul style="list-style-type: none"> - Ψηφιακές βιβλιοθήκες για την αυτόματη ανάκτηση γνώσης - Text Mining Techniques για ανακάλυψη γνώσης
Γονιδιώματα	300 πλήρη γονιδιώματα (1.6 εκατ-3 δις βάσεις το καθένα)	<ul style="list-style-type: none"> - Ανακάλυψη περιοδικοτήτων - Φυλογενετική Ανάλυση - Αντιστοίχιση γονιδίων σε αρρώστιες
Εκφράσεις Γονιδίων	~20 μετρήσεις σημείων για ~6000 γονίδια	<ul style="list-style-type: none"> - Σύγκριση εκφράσεων γονιδίων - Αντιστοίχιση εκφράσεων γονιδίων σε ακολουθιακά, δομικά και βιοχημικά δεδομένα.



Τομείς έρευνας στη Βιοπληροφορική (1/3)

1. Αποδοτική οργάνωση των δεδομένων ώστε να είναι δυνατή η αποθήκευση, ανάκτηση και ενημέρωσή τους.
2. Ανάπτυξη εργαλείων που επιτρέπουν την ανάλυση των βιολογικών δεδομένων.
3. Ανάπτυξη εργαλείων που επιτρέπουν την ερμηνεία των αποτελεσμάτων βιολογικής σημασίας.



Κατηγοριοποίηση/ανάλυση βιολογικών δεδομένων

- Ομαδοποίηση βάσει κοινών ομοιοτήτων:
 - Επαναλαμβανόμενα τμήματα στις ακολουθίες DNA/πρωτεϊνών
 - Παρόμοια δράση πρωτεϊνών
 - Ομοιότητες στην τρισδιάστατη δομή πρωτεϊνών (Protein Data Bank)
- Κύριο ρόλο έχουν οι αλγόριθμοι που χρησιμοποιούνται για την εύρεση ομοιοτήτων μεταξύ μακρομορίων.
- Εφαρμογές ολοκλήρωσης διαφορετικών δεδομένων (data integration) από το χώρο της Βιολογίας:
 - Παράδειγμα: Οι τρισδιάστατες συντεταγμένες μιας πρωτεΐνης είναι περισσότερο χρήσιμες αν συνδυαστούν με πληροφορίες που αφορούν τη λειτουργία των πρωτεϊνών και τις αλληλεπιδράσεις με άλλα μόρια.



Υλοποίηση και σχεδιασμός υπολογιστών εργαλείων για αυτόματη ανάκτηση γνώσης από βάσεις βιολογικών δεδομένων

- **Αναγκαιότητα:** Επεξεργασία και ανάλυση βιολογικών δεδομένων
- Ανάπτυξη κατάλληλων εργαλείων διαχείρισης και ανάλυσης των αποθηκευμένων πληροφοριών.
- **Πρόβλημα:** Η διαχείριση δεδομένων Μοριακής Βιολογίας παρουσιάζει αυξημένες απαιτήσεις.
- **ΣΤΟΧΟΣ:**
 - Σχεδιασμός και υλοποίηση ενός μοντέλου που να ικανοποιεί τις απαιτήσεις της έρευνας και, κυρίως, την αυτόματη ανάκτηση γνώσης (automated knowledge discovery) από μεγάλο πλήθος πληροφοριών χρησιμοποιώντας τεχνικές ομαδοποίησης των δεδομένων.
 - Σχεδιασμός αποδοτικών τεχνικών και μεθόδων που ανιχνεύουν την ομοιότητα μεταξύ **2D** ή **3D** σχημάτων.



Τομείς έρευνας στη Βιοπληροφορική (2/3)

1. Αποδοτική οργάνωση των δεδομένων ώστε να είναι δυνατή η αποθήκευση, ανάκτηση και ενημέρωσή τους.
2. Ανάπτυξη εργαλείων που επιτρέπουν την ανάλυση των βιολογικών δεδομένων.
3. Ανάπτυξη εργαλείων που επιτρέπουν την ερμηνεία των αποτελεσμάτων βιολογικής σημασίας.



Ανάλυση ακολουθιών βιολογικών δεδομένων (1/2)

- **Αντικείμενο:** Διαχείριση και ανάλυση ακολουθιών βιολογικών δεδομένων.
- **Σκοπός:** Αναγνώριση δομικών χαρακτηριστικών κρίσιμων για τη λειτουργία των ζωντανών οργανισμών.
- Κάθε μόριο DNA ή πρωτεΐνης μπορεί να θεωρηθεί ως μια συμβολοσειρά από ένα αλφάβητο:
 - 4 χαρακτήρων (DNA).
 - 20 χαρακτήρων (πρωτεΐνη).
- Περιοδικές επαναλήψεις συμβολοσειρών-μοτίβων (μοτίβο: σύνολο χαρακτήρων που εμφανίζεται παραπάνω από μία φορά σε μία ακολουθία).

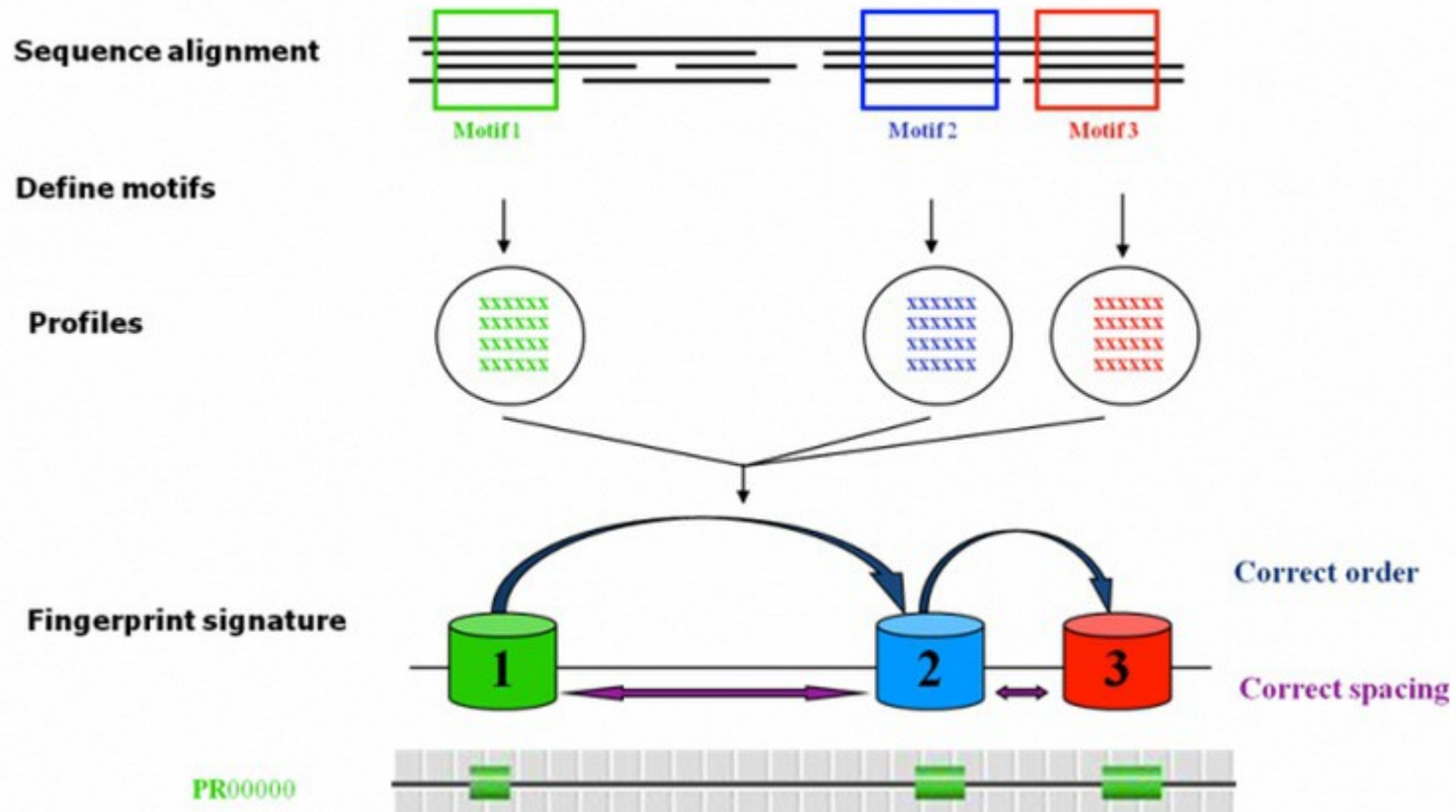


Ανάλυση ακολουθιών βιολογικών δεδομένων (2/2)

- 2 κατηγορίες προβλημάτων προς λύση:
 - **Ακριβής επανάληψη μοτίβων:** Εντοπισμός των θέσεων σε μία ακολουθία που ένα μοτίβο εμφανίζεται. Δεν επιτρέπεται καμία διαφορά μεταξύ των δύο ακολουθιών.
 - **Προσεγγιστική επανάληψη μοτίβων:** Επιτρέπεται η ύπαρξη διαφορών ανάμεσα στις συγκρινόμενες ακολουθίες, που προκύπτουν από την αντικατάσταση, προσθήκη ή διαγραφή συμβόλων.
- Συχνά, ευθυγραμμίζουμε τις αλληλουχίες (alignment) με σκοπό να τις συγκρίνουμε.
- **Ολική ευθυγράμμιση (Global alignment):** Μελετάμε ολόκληρες τις ακολουθίες.
- **Τοπική ευθυγράμμιση (Local alignment):** Μελετάμε μόνο ένα κομμάτι των αλληλουχιών.



Παράδειγμα πολλαπλής σύγκρισης αλληλουχιών



Πηγή: <http://www.ebi.ac.uk/training/online/course/introduction-protein-classification-ebi/what-are-protein-signatures/signature-types/what-ar-0>



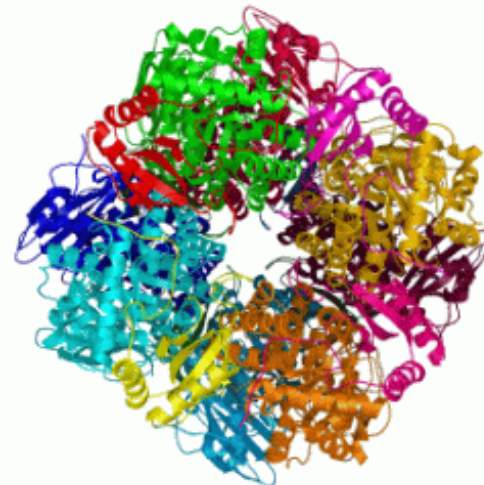
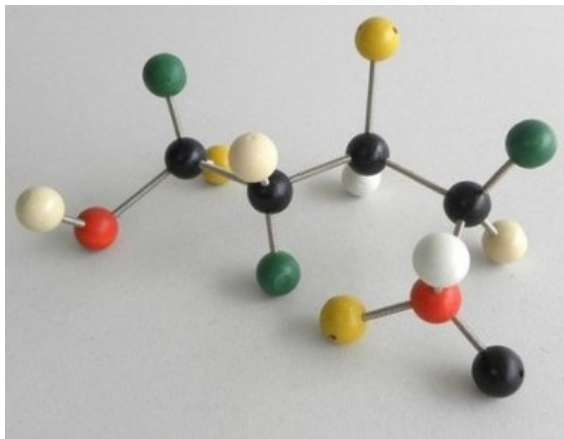
Τομείς έρευνας στη Βιοπληροφορική (3/3)

1. Αποδοτική οργάνωση των δεδομένων ώστε να είναι δυνατή η αποθήκευση, ανάκτηση και ενημέρωσή τους.
2. Ανάπτυξη εργαλείων που επιτρέπουν την ανάλυση των βιολογικών δεδομένων.
3. Ανάπτυξη εργαλείων που επιτρέπουν την ερμηνεία των αποτελεσμάτων βιολογικής σημασίας.



Μοριακή μοντελοποίηση – Molecular modelling

- Αναπτυσσόμενος επιστημονικός κλάδος που συνδυάζει τις επιστήμες της Βιολογίας και της Πληροφορικής
- **Σκοπός:** Προσπαθεί να μιμηθεί τη συμπεριφορά των μοριακών συστημάτων, βασιζόμενη σε μεγάλο βαθμό στη σχεδίαση μοντέλων μορίων με τη βοήθεια του Η/Υ.



Μοριακή μοντελοποίηση

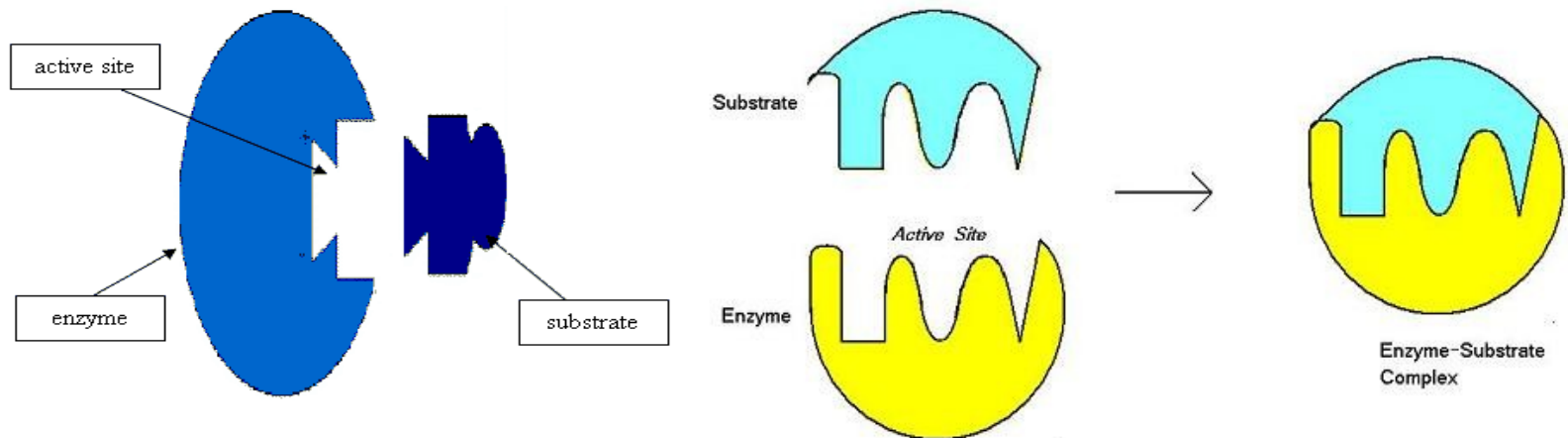
– 3 βασικά βήματα

- 1. Επιλογή του κατάλληλου μοντέλου.** Περιγράφει ικανοποιητικά τις ενδομοριακές και εσωμοριακές συσχετίσεις του μορίου.
- 2. Υπολογισμός της ενεργειακής κατάστασης του συστήματος και ελαχιστοποίηση της.**
- 3. Ανάλυση των παραπάνω υπολογισμών και έλεγχος της τελικής διαμόρφωσης.** Ικανοποίηση όλων των συνθηκών και περιορισμών που ο σχεδιαστής έχει θέσει.



Ανάλυση πρωτεϊνών (1/3)

- Οι πρωτεΐνες περιγράφονται από την αμινοξική τους ακολουθία. Ωστόσο, οι ειδικές λειτουργίες τους εξαρτώνται από την τρισδιάστατη δομή τους. Είναι αυτό αρκετό;
- Πολλές πρωτεΐνες-ένζυμα μετατρέπονται σε ενεργά μόρια με βάση μία μικρή περιοχή στην επιφάνειά τους, την ενεργό περιοχή (active site), ή ενεργό κέντρο του ενζύμου.
- Το ενεργό κέντρο έχει γεωμετρικά και φυσικοχημικά χαρακτηριστικά που είναι συμπληρωματικά ενός άλλου μορίου, του υποστρώματος (substrate).

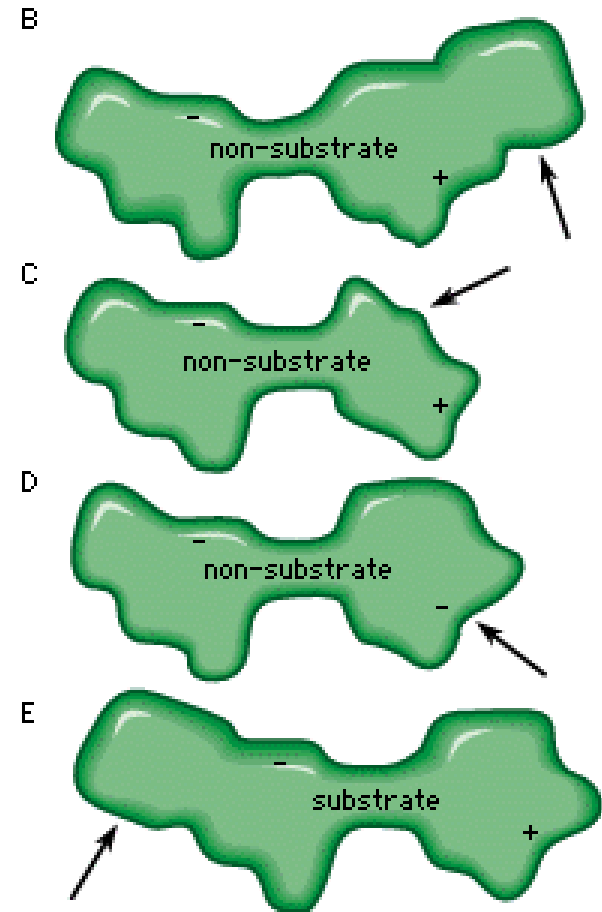
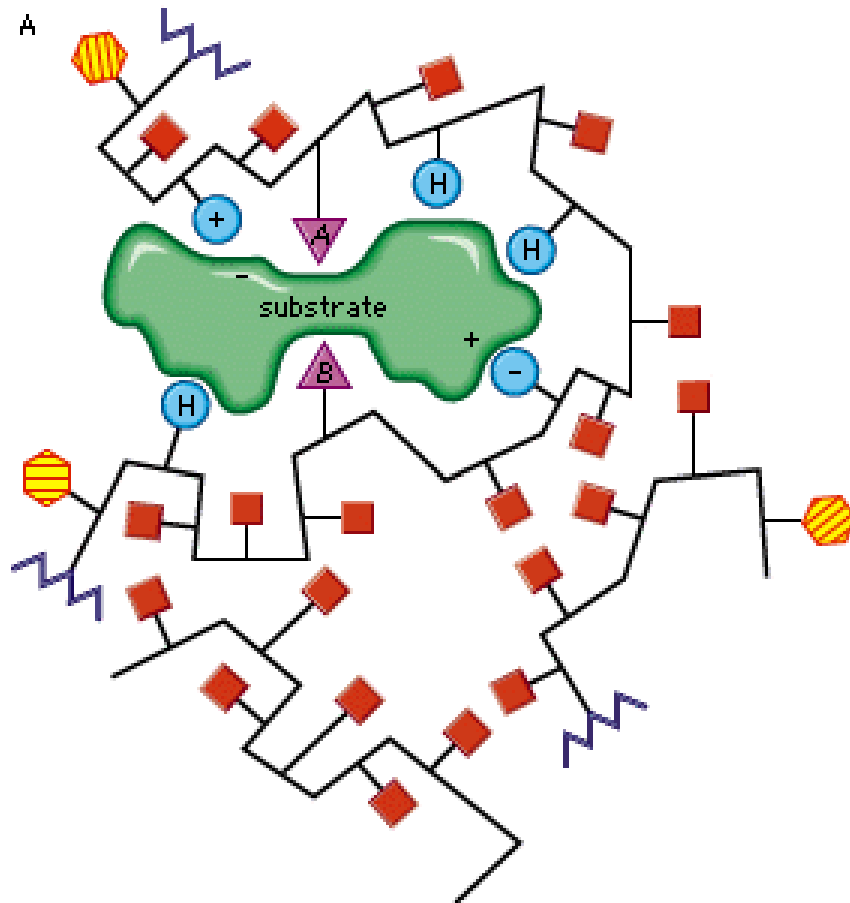


Ανάλυση πρωτεϊνών (2/3)

- Ένζυμο και υπόστρωμα για να ενωθούν πρέπει να βρεθούν στην ευνοϊκότερη διαμόρφωση.
- Προκειμένου να εντοπίσουμε την ευνοϊκότερη διαμόρφωση απαιτείται να μελετήσουμε τις επιφάνειες επαφής.
- Οι περισσότεροι αλγόριθμοι βασίζονται στην:
 - Αρχή συμπληρωματικότητας ως προς τη δομή.
 - Αρχή συμπληρωματικότητας ως προς την ηλεκτρονική διαμόρφωση.
- Προκειμένου να βρούμε τα συμπληρωματικά μόρια που ανήκουν σε δύο πρωτεΐνες A & B, έχουμε να επιλύσουμε ένα πρόβλημα ταιριάσματος σε 3 διαστάσεις (**3D matching problem**).



Ανάλυση πρωτεϊνών (3/3)



©1997 Encyclopaedia Britannica, Inc.



Σχετικά πεδία (1): Υπολογιστική Βιολογία (Computational Biology)

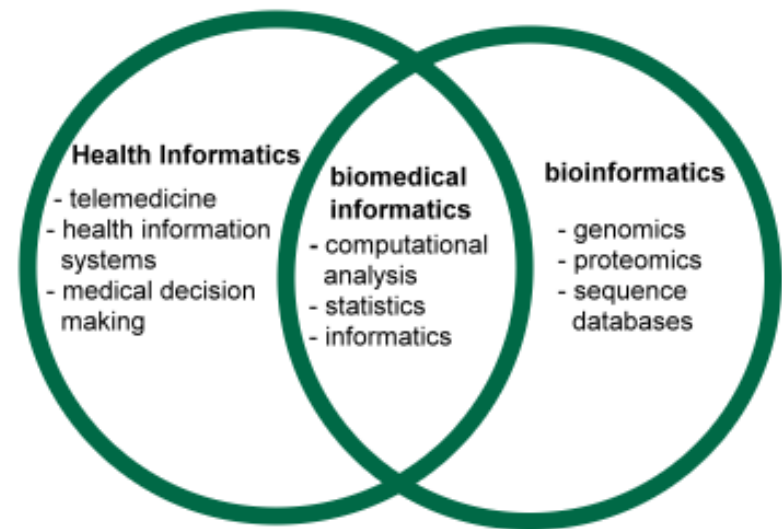
- Η μελέτη και η εφαρμογή υπολογιστικών μεθόδων στην κλασική βιολογία.
- Κύριο αντικείμενο μελέτης η εξέλιξη των ειδών (πληθυσμιακή και θεωρητική βιολογία) και όχι η μελέτη σε κυτταρικό και μοριακό επίπεδο.

Πηγή εικόνας: <http://contentinacottage.blogspot.ca/2012/01/29/archive.html>



Σχετικά πεδία (2): Ιατρική πληροφορική (Medical Informatics)

- Η μελέτη και η εφαρμογή υπολογιστικών μεθόδων για τη βελτίωση της επικοινωνίας, της κατανόησης και της διαχείρισης ιατρικών δεδομένων.
- Κύριο αντικείμενο μελέτης ο χειρισμός/αξιοποίηση της πληροφορίας και όχι η ίδια η πληροφορία.



Σχετικά πεδία (3): Cheminformatics

- Η μελέτη και η εφαρμογή υπολογιστικών μεθόδων, σε συνδυασμό με χημικές και βιολογικές τεχνικές, για τον σχεδιασμό και ανάπτυξη φαρμάκων.



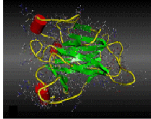
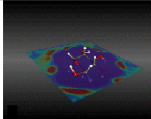
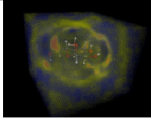
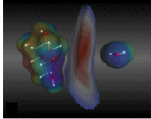
Σχεδιασμός φαρμάκων με τη βοήθεια Η/Υ

Μεθοδολογία:

1. Αποτελεσματική απεικόνιση των δομών κανονικών και παθολογικών μορίων.
2. Σύγκριση με μόρια-ενεργοποιητές, για την αναγνώριση περιοχών που είναι υπεύθυνες για τη δράση των μορίων.
3. Προσομοίωση της ενεργοποίησης του παθολογικού μορίου από το μόριο-ενεργοποιητή στην οθόνη του Η/Υ.
4. Σχεδίαση και ανάπτυξη μορίων-φαρμάκων με δυνατότητα πρόσδεσης στο παθολογικό μόριο, με σκοπό την αδυναμία ενεργοποίησης του.
5. Επιβεβαίωση προβλέψεων με στοχευμένα βιολογικά πειράματα.



Αλληλουχία DNA

Εύρεση Ομολόγων						
Ορθολογικός Σχεδιασμός Φαρμάκων			Σύγκριση ζεύγους ακολουθιών	Αλγόριθμοι σύγκρισης πολλαπλών ακολουθιών	Εύρεση κοινών σε βάσεις δεδομένων ακολουθιών	
			1	2	3 -100	100+
		Ακολουθίες Γονιδιωμάτων	atcgatcgatattgg	atcgatcgatattgg atcgatcgatattgg	atcgatcgatattgg atcgatcgatattgg atcgatcgatattgg	atcgatcgatattgg atcgatcgatattgg atcgatcgatattgg atcgatcgatattgg atcgatcgatattgg
	Εύρεση γονιδίων	↓				
		Ακολουθίες Πρωτεϊνών	ALMNAKKKRQ	ALMNAKKKRQ ALMNAKKKRQ	ALMNAKKKRQ ALMNAKKKRQ ALMNAKKKRQ ALMNAKKKRQ	ALMNAKKKRQ ALMNAKKKRQ ALMNAKKKRQ ALMNAKKKRQ ALMNAKKKRQ
	Πρόβλεψη Δομής	↓				
		Δομή Πρωτεϊνών				Ο κάθετος άξονας παρουσιάζει την πορεία που ακολουθεί ο ορθολογικός σχεδιασμός φαρμάκων. Αντίστοιχα ο οριζόντιος άξονας παρουσιάζει τη συμβολή των υπολογιστικών τεχνικών στη διαδικασία σύγκρισης των ακολουθιών πρωτεϊνών και του καθορισμού της δομής σχετικών πρωτεϊνών.
	Υπολογισμός Γεωμετρίας	↓				
		Επιφάνεια Πρωτεϊνών				
	Μοριακές Προσομοιώσεις	↓				
	Δυνάμεις Πεδίου					
Προσάραξη μορίων	↓					
	Πρόσραξη προσδέματος					



Σχετικά πεδία (4): Genomics

- Γονιδίωμα ονομάζεται το σύνολο των γονιδίων ενός οργανισμού.
- Ανάλυση και σύγκριση γονιδιωμάτων ενός ή πολλών ειδών μεταξύ τους.
- Το πεδίο έρευνας «Genomics» υπήρχε και πριν από την αποκρυπτογράφηση γονιδιωμάτων διάφορων οργανισμών (π.χ. ανθρώπου, κουνουπιού), αλλά σε πολύ αρχικό στάδιο.



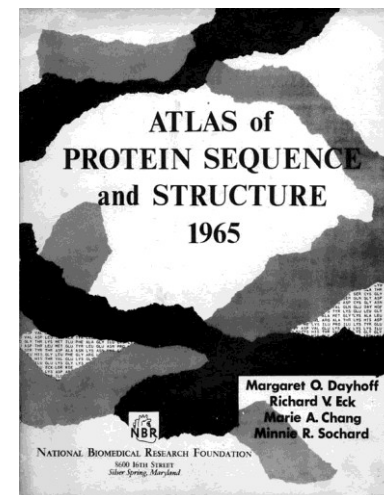
Σχετικά πεδία (5): Proteomics

- Μελέτη των πρωτεϊνών που παράγονται από γονιδίωμα.
- Τρόποι λειτουργίας και αλληλεπίδρασης των πρωτεϊνών.
- Ενδιαφέρον για την πραγματική κατάσταση των πρωτεϊνών στα κύτταρα και όχι για το σύνολο των πιθανών καταστάσεων που περιγράφεται από το γονιδίωμα.



Και μία ιστορική αναδρομή...(1/2)

- **1965:**
 - Κυκλοφόρησε σε έντυπη μορφή η 1^η έκδοση του Atlas of Protein Sequence and Structure, από τη Margaret Dayhoff. Θεμέλιο για τη δημιουργία της βάσης δεδομένων πρωτεϊνικών ακολουθιών PIR (**Protein Information Resource**).
 - Αναφέρεται η **πρώτη ακολουθία νευκλεοτιδίων (DNA) μήκους 77 βάσεων**.
- **1970:** Δημοσιεύεται ο πρώτος αλγόριθμος για τη σύγκριση ακολουθιών.
- **1971:** Δημιουργείται η πρωτεϊνική βάση δεδομένων Protein Data Bank (PDB).
- **1988:** Δημοσιεύεται ο αλγόριθμος FASTA για αναζήτηση των βάσεων δεδομένων.
- **1990:** Δημοσιεύεται ο πιο γρήγορος αλγόριθμος BLAST για αναζήτηση των βάσεων δεδομένων.



Και μία ιστορική αναδρομή...(2/2)

- 1990-2003: Human Genome Project
- **Σκοπός:**
 - Να προσδιορίσει τα περίπου 20.000-25.000 γονίδια στο ανθρώπινο DNA.
 - Να καθορίσει τις ακολουθίες των 3 δισεκατομμυρίων ζευγών βάσεων που αποτελούν το ανθρώπινο DNA.
 - Να καταχωρήσει όλη την πληροφορία σε βάσεις δεδομένων.
 - Να βελτιώσει τα εργαλεία για την ανάλυση των συγκεκριμένων δεδομένων.
 - Να διευθετήσει να ηθικά, νομικά και κοινωνικά ζητήματα που πιθανώς θα ανερχόταν κατά τη διάρκεια του προγράμματος.

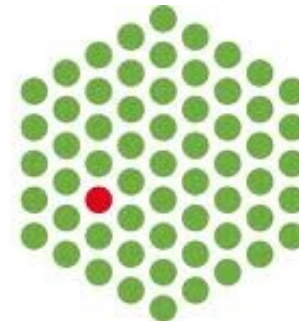
Human Genome Project
Information



Η έρευνα στον τομέα της Βιοπληροφορικής

- **EMBL** – European Molecular Biology Laboratory.
- **NCBI** – National Center for Biothechnology.

EMBL



Βιβλιογραφικές και βιολογικές βάσεις δεδομένων

- Περιγράφονται οι βιβλιογραφικές βάσεις δεδομένων και οι μηχανές αναζήτησής τους (Scholar Google, Scopus, PubMed).
- Περιγράφονται οι βιολογικές βάσεις δεδομένων και οι μηχανές αναζήτησής τους.
- Ανάλυση της χρησιμότητας των βιολογικών βάσεων δεδομένων στην ενίσχυση των επιστημών για την κατανόηση και ερμηνεία ενός πλήθους βιολογικών φαινομένων.



Ανάλυση και στοίχιση ακολουθιών

- Πλήθος διαθέσιμων ακολουθιών (κομμάτια DNA ή πρωτεΐνης).
- Αδύνατη ανάλυση με «το χέρι».
- Απαραίτητη η ηλεκτρονική ανάλυση και στοίχιση ακολουθιών.
 - Αλγόριθμοι δυναμικού προγραμματισμού.
 - Ευρετικοί αλγόριθμοι.
- Οι στοιχίσεις μεταξύ ακολουθιών μας δείχνουν ποια κομμάτια είναι μεταβλητά και ποια συντηρούνται.
- Αναγνωρίζονται θέσεις πιθανών μεταλλάξεων, προσθήκης, αφαίρεσης ή και αντικατάστασης μέρους μιας αλληλουχίας.

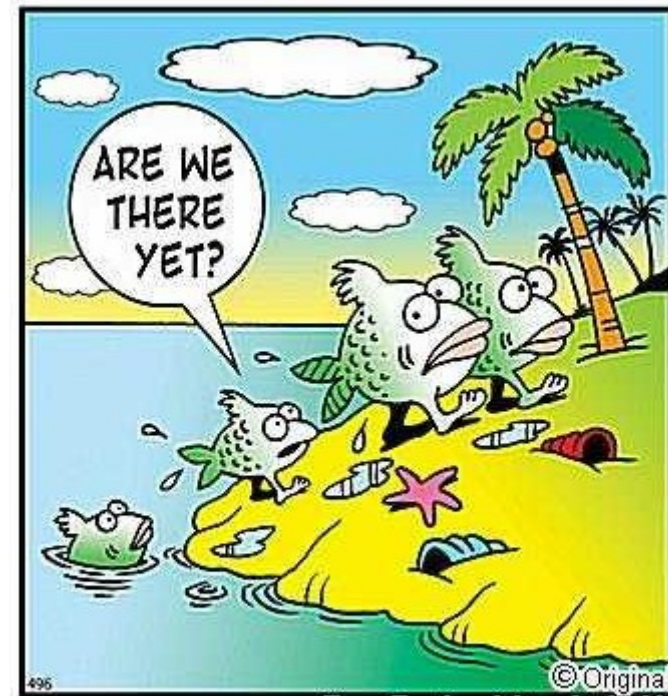
Scarites	C	T	T	A	G	A	T	C	G	T	A	C	C	A	A	-	-	-	A	A	T	A	T	T	A	C
Carenum	C	T	T	A	G	A	T	C	G	T	A	C	C	A	C	A	-	T	A	C	-	T	T	T	A	C
Pasimachus	A	T	T	A	G	A	T	C	G	T	A	C	C	A	C	T	A	T	A	A	G	T	T	T	A	C
Pheropsophus	C	T	T	A	G	A	T	C	G	T	T	C	C	A	C	-	-	-	A	C	A	T	A	T	A	C
Brachinus armiger	A	T	T	A	G	A	T	C	G	T	A	C	C	A	C	-	-	-	A	T	A	T	A	T	T	C
Brachinus hirsutus	A	T	T	A	G	A	T	C	G	T	A	C	C	A	C	-	-	-	A	T	A	T	A	T	A	C
Aptinus	C	T	T	A	G	A	T	C	G	T	A	C	C	A	C	-	-	-	A	C	A	A	T	T	A	C
Pseudomorpha	C	T	T	A	G	A	T	C	G	T	A	C	C	-	-	-	-	-	A	C	A	A	A	T	A	C



Φυλογενετική ανάλυση

– Κατασκευή φυλ. Δέντρων (1/2)

- **Εξελεκτική βιολογία:**
 - Μελέτη προέλευσης και καταγωγής των ειδών.
 - Αλλαγή των έμβιων όντων με το πέρασμα του χρόνου.
- **Βιοπληροφορική:**
 - Αποκρυπτογραφεί τους ρυθμούς εξέλιξης γονιδίων με την καταγραφή των αλλαγών στο DNA
 - Προβλέπει την έκβαση κάθε συστήματος με το πέρασμα του χρόνου.



search ID: mlyn190

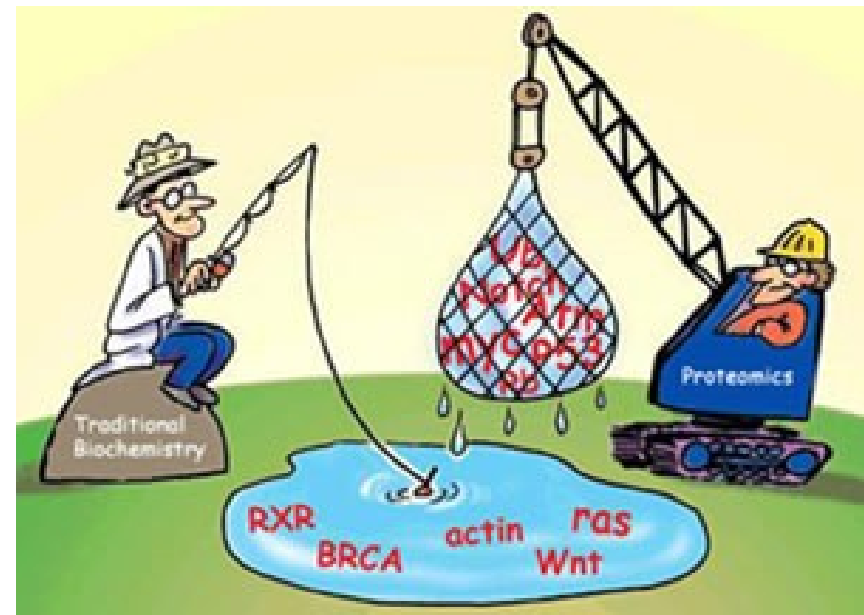
©Original Artist
Reproduction rights obtainable from
www.CartoonStock.com



Φυλογενετική ανάλυση

– Κατασκευή φυλ. Δέντρων (2/2)

- **Πρωτεϊνική ανάλυση:**
 - Πρόγνωση πρωτεϊνικής δομής από την αλληλουχία τους.
 - Οπτικοποίηση της δομής των πρωτεϊνών.
 - Πρόβλεψη πιθανών αλληλεπιδράσεων μεταξύ των πρωτεϊνών.
- **Πρωτεομική (Proteomics):**
 - Ανάλυση της δομής και λειτουργίας των πρωτεϊνών.



Βιβλία...

- **ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ. ΔΥΝΑΤΟΤΗΤΕΣ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ.**

Σοφία Κοσσίδα,

Εκδόσεις: Σοφία Κοσσίδα, Ίδρυμα Ιατροβιολογικών,
Ερευνών της Ακαδημίας Αθηνών,
2008 Σοφία Κοσσίδα.

- **ΕΙΣΑΓΩΓΗ ΣΤΟΥΣ ΑΛΓΟΡΙΘΜΟΥΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ.**

Neil C. Jones, Pavel A. Pevzner,

Μετάφραση: Π. Μπένος,
Εκδοσεις Κλειδαριθμος ΕΠΕ, 2010.

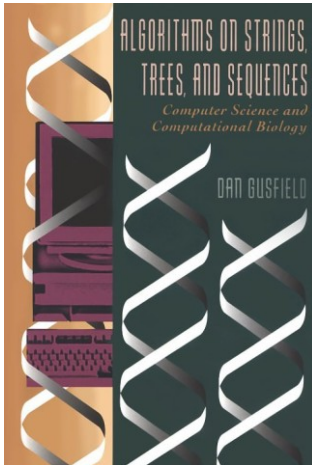
- **ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ. Ένας Πρακτικός Οδηγός για την Ανάλυση Γονιδίων και Πρωτεϊνών.**

A.D. Baxevanis & B.F.F Ouellette,

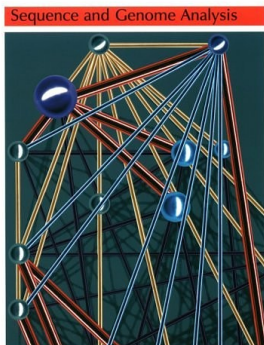
Μετάφραση: Ε. Μουδριανάκη, Σ. Χαμόδρακα,
Επιστημονικές Εκδόσεις ΠΑΡΙΣΙΑΝΟΥ, 2004.



Βιβλία σχετικά με τη Βιοπληροφορική



Bioinformatics



David W. Mount

COLD SPRING HARBOR LABORATORY PRESS

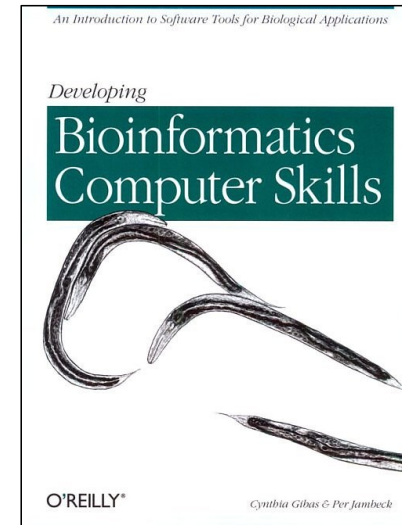
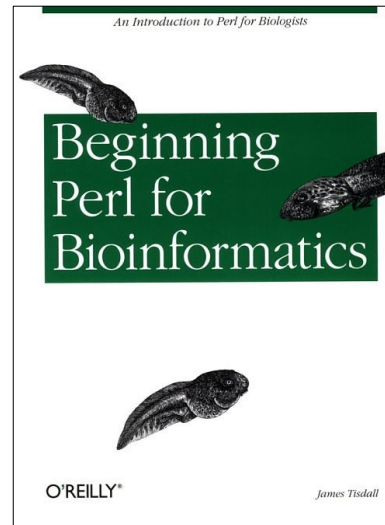
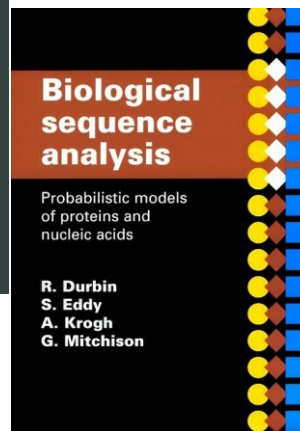


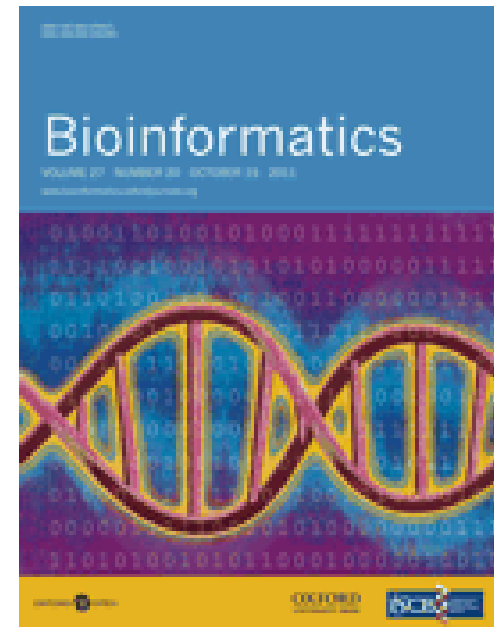
Image Source: <http://www.amazon.com/>



Τα κυριότερα επιστημονικά περιοδικά

- Bioinformatics, Oxford University Press.
- IEEE/ACM transactions on Computational Biology and Bioinformatics (TCBB).
- Journal of Computational Biology).

Science, Nature, Nucleic Acid Research, Journal of Molecular Biology, Proceedings of the National Academy of Sciences (PNAS).



Άλλα χρήσιμα βιβλία...

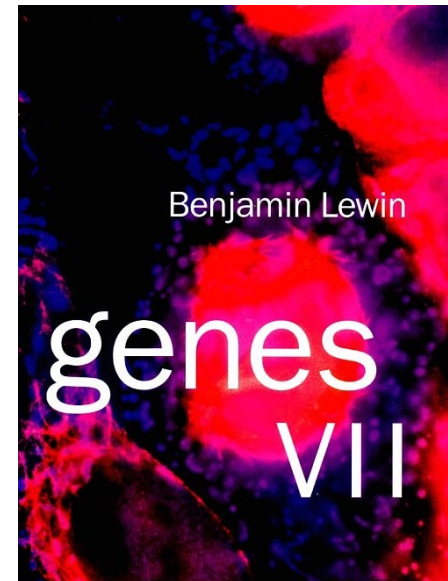
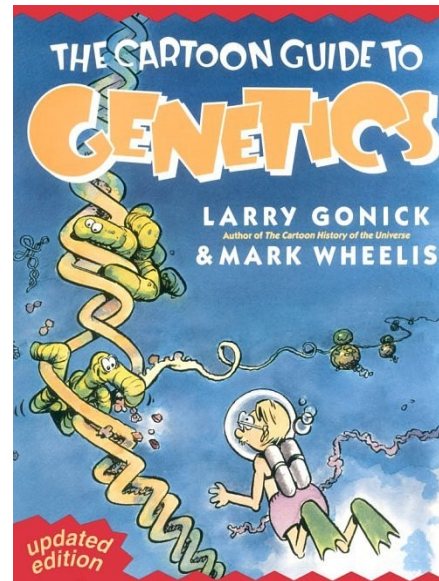
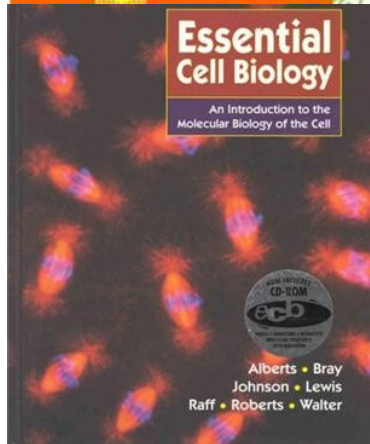
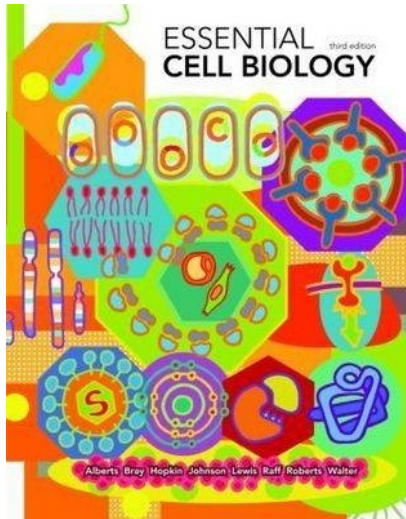


Image Source: <http://www.amazon.com/>



Εξετάσεις - Βαθμολόγηση

- **Εξετάσεις:**

- Θεωρία (2 ώρες).

- Εργαστήριο (2 ώρες).

- **Βαθμολόγηση:**

- Βαθμός θεωρητικής εξέτασης: **60%**

- Βαθμός πρακτικής εξέτασης: **20%**

- Βαθμός εργασιών: **20%**



Πιθανά θέματα εργασιών (1/2)

- **Sequence alignment** (pairwise sequence alignment, spliced alignment and similarity based gene recognition, multiple sequence alignment, parametric sequence alignment).
- **Bioinformatic databases and data mining** (string search in external memory , index structures for approximate matching in sequence databases, algorithms for motif search, data mining in computational biology).
- **Genome scale computational methods** (comparison of long genomic sequence, algorithms and applications, chaining algorithms and applications in comparative genomics, computational analysis of alternative splicing, human genetic linkage analysis, haplotype inference).
- **Phylogenetics** (phylogenetic reconstruction, consensus trees and supertrees, large scale phylogenetic analysis, high performance phylogeny reconstruction).



Πιθανά θέματα εργασιών (2/2)

- **Microarrays and gene expression analysis** (microarray data: annotation retrieval, storage and communication, computational methods for microarray design, clustering algorithms for gene expression analysis, biclustering algorithms, identifying gene regulatory networks from gene expression data, modeling and analysis of gene networks using feedback control analysis).
- **Computational Structural Biology** (predicting protein structure and supersecondary structure, protein structure prediction with lattice models, proteins structure determination via NMR spectral data, geometric and signal processing of reconstructed 3D maps of molecular complexes, in search of remote homologs, biomolecular modeling using parallel supercomputers).



Τέλος Ενότητας



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



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:

- <http://blog.com.mk/send/121903>
- <http://foter.com/Cmyk/>
- <http://www.ebi.ac.uk/training/online/course/introduction-protein-classification-ebi/what-are-protein-signatures/signature-types/what-ar-0>
- http://contentinacottage.blogspot.ca/2012_01_29_archive.html
- <https://www.cartoonstock.com/>

