

Εισαγωγή στη Βιοπληροφορική

Διδάσκων: Μακρής Χρήστος, Μεγαλοοικονόμου Βασίλειος
Καλογερόπουλος Νικήτας

e-mail: makri@ceid.upatras.gr

Διδασκαλία: Δευτέρα 17:00-19:00 (αίθουσα Δ1)

Τετάρτη 16:00-18:00 (αίθουσα Δ1)

Μερικές διαφάνειες από διαφάνειες βιβλίου [Neil C. Jones, Pavel A. Pevzner, Εισαγωγή στους Αλγορίθμους Βιοπληροφορική, οι συγκεκριμένες είναι εισαγωγικές με συμμετοχή των Angela Brooks, Raymond Brown, Calvin Chen, Mike Daly, Hoa Dinh, Erinn Hama, Robert Hinman, Julio Ng, Michael Sneddon, Hoa Troung, Jerry Wang, Che Fung Yung](#)

Σελίδα μαθήματος

✓ e_class: <https://eclass.upatras.gr/courses/CEID1047/>

Πρέπει να εγγραφείτε αν έχετε 'δηλώσει το μάθημα

Εισαγωγικά

Η Βιοπληροφορική μπορεί να οριστεί ως: “η εφαρμογή υπολογιστικών τεχνικών και μεθόδων στην προσπάθεια κατανόησης και οργάνωσης των δεδομένων και πληροφοριών που σχετίζονται με τα βιολογικά μακρομόρια...”.

Bioinformatics-Bioinformatics is conceptualizing biology in terms of molecules (in the sense of Physical Chemistry) and applying "informatics techniques" (derived from disciplines such as applied maths, computer science and statistics) to understand and organise the information associated with these molecules, on a large scale. In short bioinformatics is a management information system for molecular biology and has many practical applications. (N. Luscombe, D. Greenbaum, M. Gerstein, What is Bioinformatics? A proposed definition and overview of the field, Method Inform Med, 2001)

Τομείς Έρευνας

- ✓ Αποδοτική αποθήκευση και οργάνωση των βιολογικών δεδομένων
- ✓ Ανάπτυξη εργαλείων που επιτρέπουν την αποδοτική ανάλυση των βιολογικών δεδομένων
- ✓ Ανάπτυξη εργαλείων για ερμηνεία αποτελεσμάτων

Διάγραμμα Ύλης

✓ Γενική Εισαγωγή

- ✓ Εισαγωγή στη χρήση αλγορίθμων για αποτελεσματική διαχείριση και αποθήκευση συμβολοσειρών (strings) και ακολουθιών βιολογικών δεδομένων.
- ✓ Αλγόριθμοι ακριβούς ταιριάσματος προτύπου (Boyer-Moore, Knuth-Morris-Pratt, Shift-Or, Πολλαπλών Προτύπων).
- ✓ Δέντρο επιθεμάτων (suffix tree) και εφαρμογές του
- ✓ Αλγόριθμοι προσεγγιστικού ταιριάσματος προτύπου και στοίχισης συμβολοσειρών/ ακολουθιών (Sequence Alignment)
- ✓ Αλγόριθμοι αναζήτησης σε Βάσεις Δεδομένων ακολουθιών (FASTA, BLAST, PROSITE) (PAM, BLOSUM)
- ✓ Βιολογικές Βάσεις Δεδομένων NCBI, EBI, DDJB και εργαλεία προσπέλασης τους.
- ✓ Φυλογενετικά δένδρα
- ✓ Ακολουθιοποίηση γονιδιώματος, Sanger και Next Generation Sequencing (De Bruijn graphs και Burrows Wheeler μετασχηματισμός)
- ✓ Πρόβλεψη με Hidden Markov Models (Viterbi)
- ✓ Protein folding (ESMFold and AlphaFold) .
- ✓ Γονιδιακές εκφράσεις. Τεχνικές ομαδοποίησης και κατηγοριοποίησης βιολογικών δεδομένων (clustering and categorization techniques) με σκοπό την πρόβλεψη της συμπεριφοράς βιολογικών μορίων
- ✓ Δίκτυα Αλληλεπίδρασης και αλγόριθμοι γραφημάτων.

✓ Δεύτερο μέρος

-- Η Θεωρητική Βάση του Μοριακού Σχεδιασμού

-- Μοριακά Μοντέλα και Βιοχημική Πληροφορία

-- Η Βασιζόμενη στη Δομή Σχεδίαση Φαρμάκων

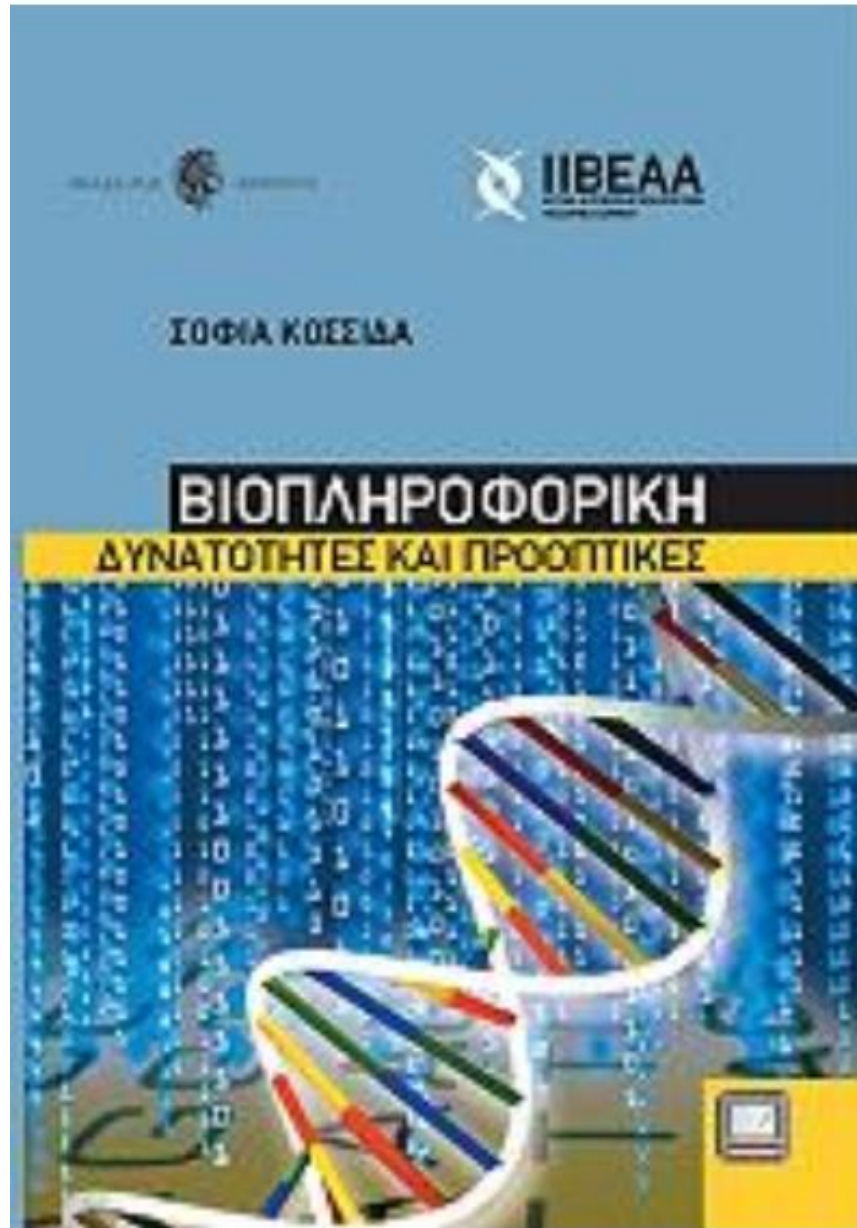
(ένα σύνολο από 2-3 διαλέξεις από τον κ. Μεγαλοοικονόμου σε αντικείμενα εφαρμογών τεχνικών Βιοπληροφορικής σε πραγματικά προβλήματα)

Εξέταση Μαθήματος

Η εξέταση του μαθήματος συνίσταται:

1. Στην παράδοση μιας εργασίας από ομάδες 1-2 ατόμων → 30% βαθμού, βασισμένη στο βιβλίο: Dan Gusfield, Algorithms on trees strings and sequences, Computer Science and Computational Biology, Cambridge University Press και Jonathan Pevzner Βιοπληροφορική και Λειτουργική Γονιδιωματική.
2. Παρουσίαση & Προφορική Εξέταση πάνω στις σημειώσεις του μαθήματος και σε μία επιπλέον εργασία → 70% βαθμού. Η προφορική εξέταση είναι (κυρίως) πάνω στις (ηλεκτρονικά διαθέσιμες) σημειώσεις του μαθήματος και σε επιλεγμένα κεφάλαια του προτεινόμενου συγγράμματος του Εύδοξου
3. Η δεύτερη εργασία θα είναι είτε θέμα δικό μου είτε θέμα του κ. Μεγαλοοικονόμου.

Εύδοξος σύγγραμμα 1



Εύδοξος σύγγραμμα 2 (κύριο σύγγραμμα)



Απλοποιημένο πιο λεπτομερές

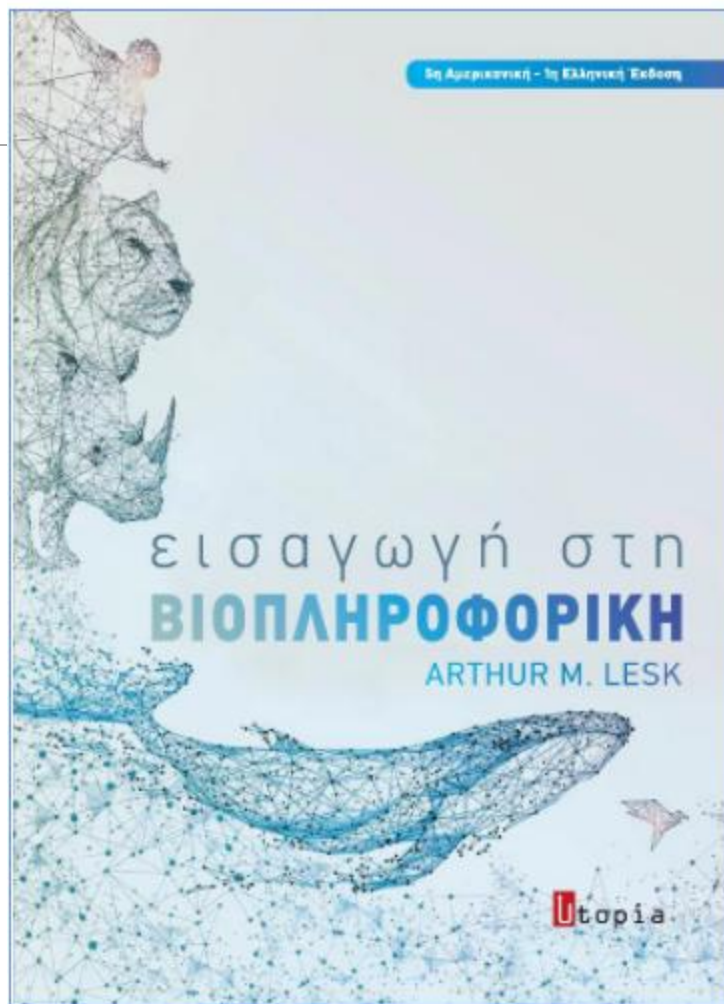
Bioinformatics algorithms : an active learning approach /
Phillip Compeau & Pavel Pevzner, Publisher: Active Learning
Publishers; 3rd edition (January 1, 2018)

<https://www.bioinformaticsalgorithms.org/>

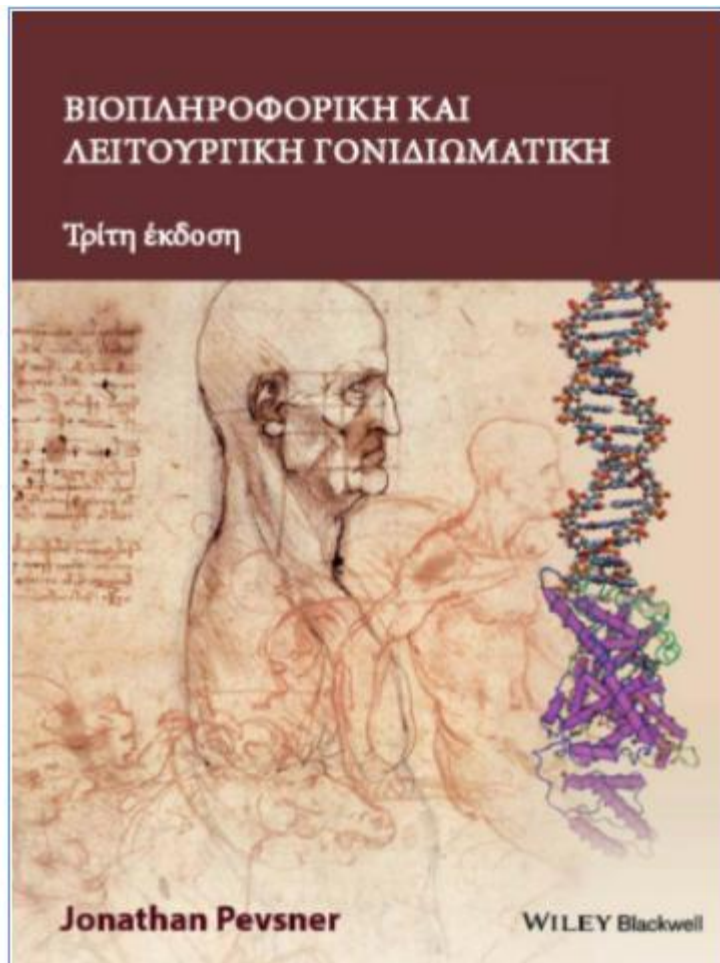
<https://rosalind.info/problems/locations/>

<https://boinc.bakerlab.org/>

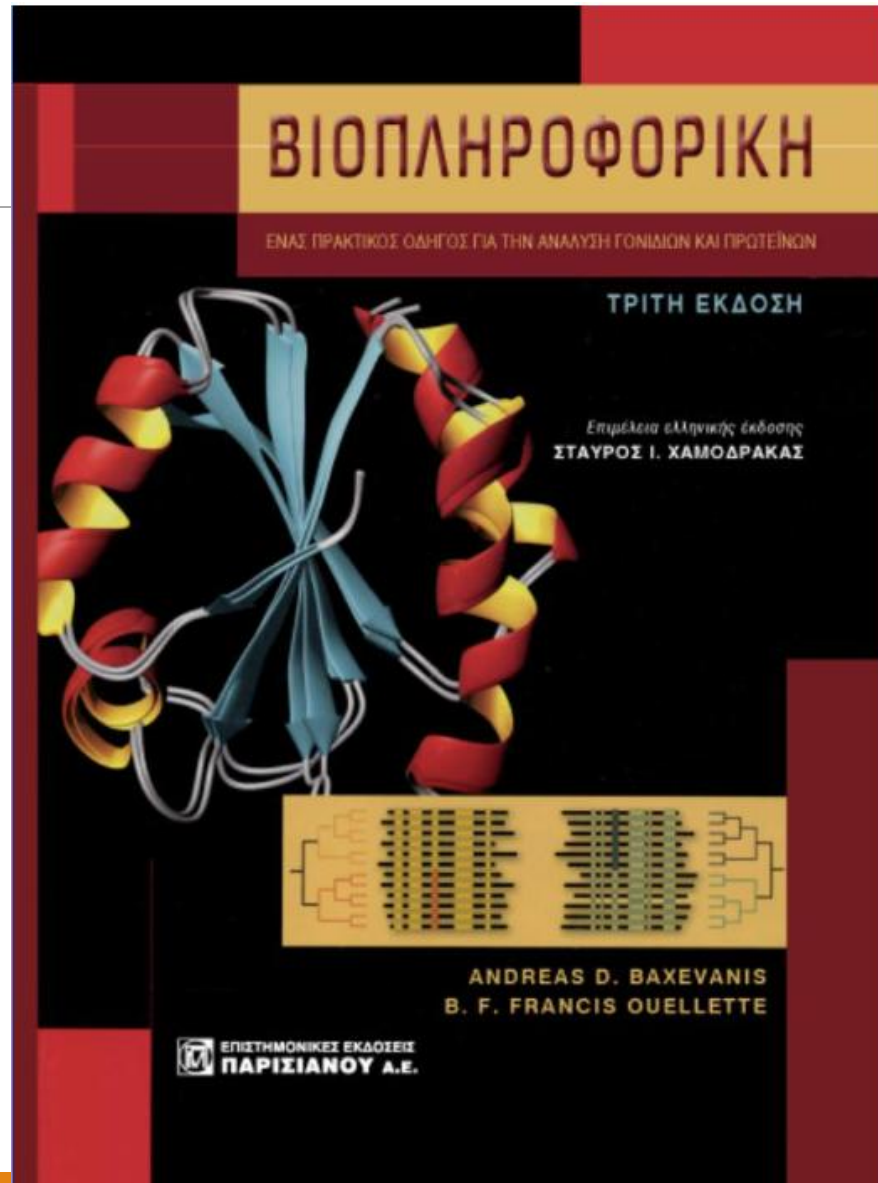
Εύδοξος σύγγραμμα 3



Εύδοξος σύγγραμμα 4

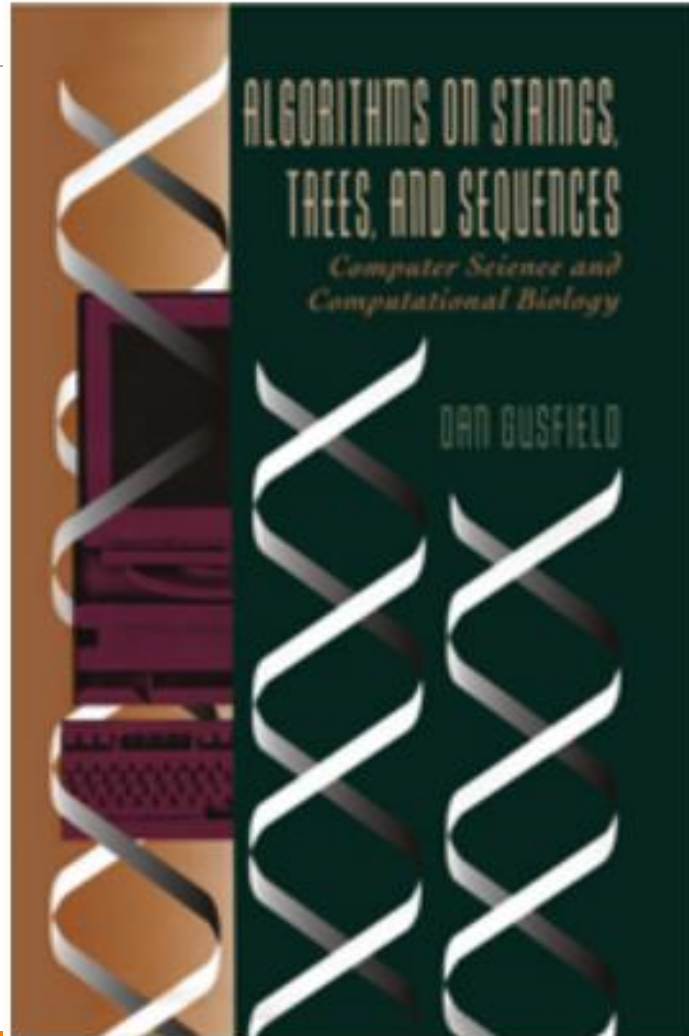


Εύδοξος σύγγραμμα 5



Βασικό βιβλίο για ασκήσεις

**Algorithms On Strings, Trees And Sequences: Computer Science
And Computational Biology - Dan Gusfield** Cambridge University
Press (1997)



Βιβλία

Richard Durbin, Sean R. Eddy
Anders Krogh, Graeme Mitchison

Ανάλυση βιολογικών αλληλουχιών

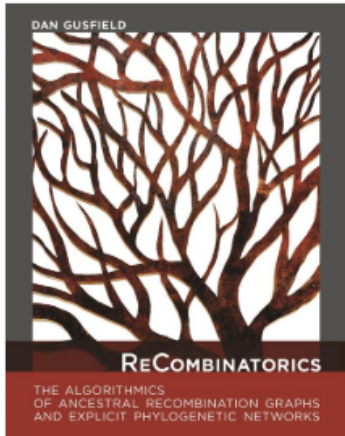
Πιθανοκρατικά μοντέλα πρωτεϊνών
και νουκλεϊκών οξέων



Επιστημονική επιμέλεια: Ιωάννης Ζ. Εμίρης

 πεδίο

Βιβλία



ReCombinatorics: The Algorithmics of Ancestral Recombination Graphs and Explicit Phylogenetic Networks



By Dan Gusfield

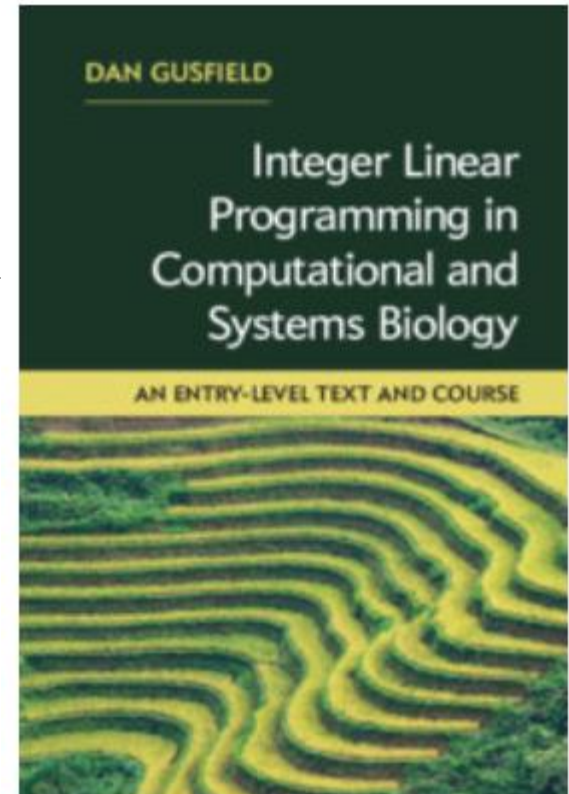
The MIT Press

DOI: <https://doi.org/10.7551/mitpress/9432.001.0001>

ISBN electronic: 9780262324472

Publication date: 2014

Βιβλία



Publisher: Cambridge University Press
Online publication date: May 2019
Print publication year: 2019
Online ISBN: 9781108377737
DOI: <https://doi.org/10.1017/9781108377737>

Subjects: Genomics, Bioinformatics and Systems Biology, Life Sciences, Computer Science, Computational Biology and Bioinformatics

Βιβλία

Μπάγκος Παντελεήμων, Βιοπληροφορική, Έκδοση: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών 2015(<https://repository.kallipos.gr/handle/11419/5016>).

Νικολάου, Χριστόφορος Χουβαρδάς, Παναγιώτης, Υπολογιστική Βιολογία, Έκδοση: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών , 2015 (<https://repository.kallipos.gr/handle/11419/1577>).

A. Baxevanis, B.F. Ouellette, Βιοπληροφορική, 2η/2012
Παρισιανού Μονοπρόσωπη ανώνυμη εκδοτική εισαγωγική εμπορική εταιρεία επιστημονικών βιβλίων

Βασικές Έννοιες (1)

Βιοπληροφορική είναι η διαχείριση της Βιολογίας σε όρους μορίων (με την έννοια της Φυσικής Χημείας) και η εφαρμογή “τεχνικών πληροφορικής” (εφαρμοσμένα μαθηματικά, επιστήμη των υπολογιστών και στατιστική) για την κατανόηση και οργάνωση πληροφορίας που σχετίζεται με τα μόρια σε μεγάλη κλίμακα.

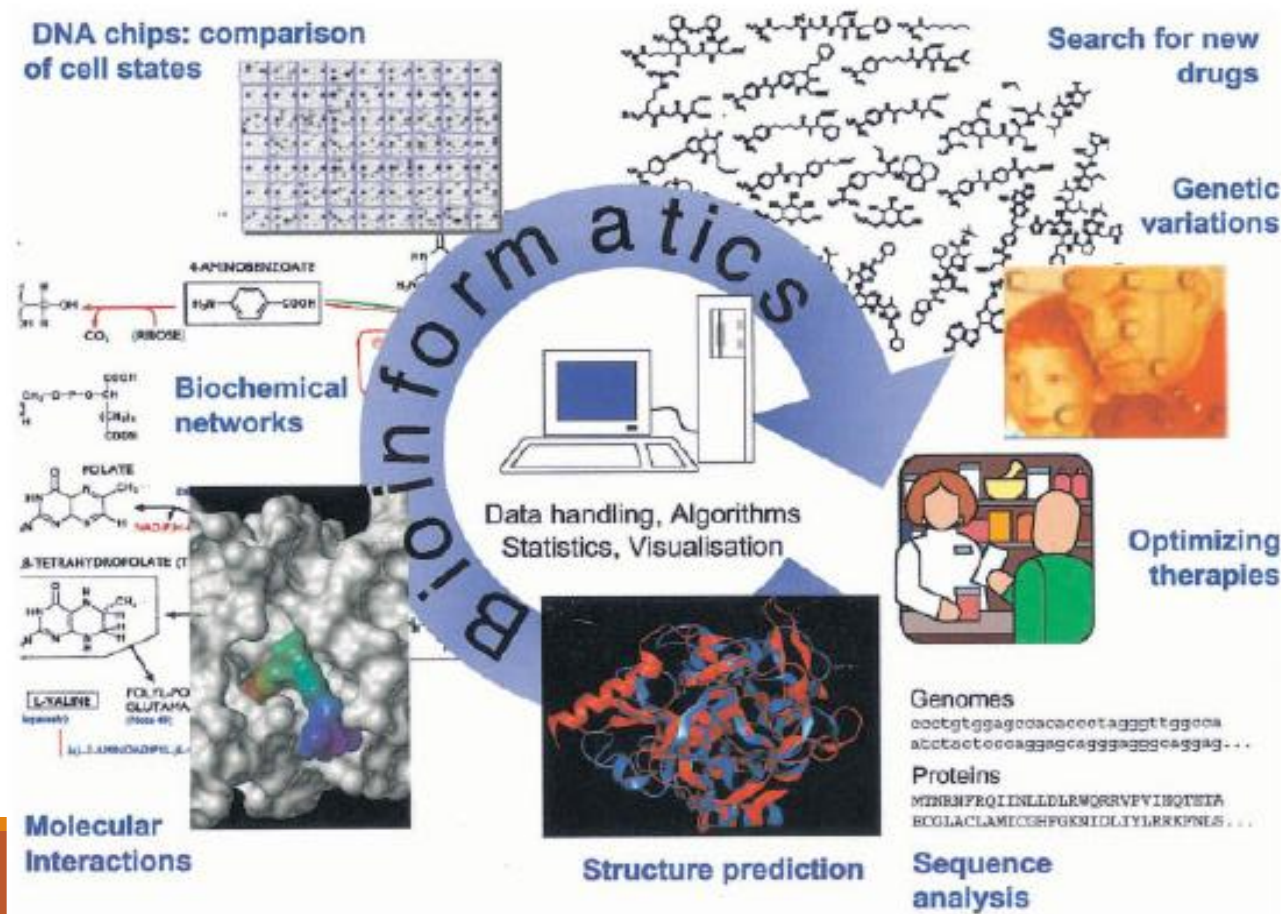


Figure from
“Bioinformatics, from
Genomes to Drugs”.
T. Lengauer

Πληροφορίες και «μηχανικός εξοπλισμός» των κυττάρων

- ✓ Το κύτταρο αποθηκεύει όλες τις πληροφορίες για να δημιουργήσει αντίγραφά του
 - ✓ Το ανθρώπινο γονιδίωμα περιέχει περίπου 3 δισεκατομμύρια ζεύγη βάσεων
 - ✓ Σχεδόν όλα τα κύτταρα στο ανθρώπινο σώμα περιέχουν το ίδιο σύνολο γονιδίων
 - ✓ Όμως, δεν χρησιμοποιούνται ή εκφράζονται όλα τα γονίδια από τα κύτταρα αυτά
- ✓ «Μηχανικός εξοπλισμός»:
 - Συλλογή και κατασκευή συστατικών στοιχείων
 - Εκτέλεση αντιγραφής
 - Έναρξη διαδικασίας διπλασιασμού(Το κύτταρο μοιάζει με εργοστάσιο κατασκευής αυτοκινήτων)

Ορολογία

- ✓ Γονιδίωμα: το γενετικό υλικό ενός οργανισμού

- ✓ Γονίδιο: διακριτή μονάδα κληρονομικών πληροφοριών που βρίσκεται στα χρωμοσώματα και αποτελείται από DNA.
- ✓ Γονότυπος: Η γενετική σύσταση ενός οργανισμού
- ✓ Φαινότυπος: τα φυσικά εκφρασμένα (κληρονομικά) χαρακτηριστικά ενός οργανισμού
- ✓ Νουκλεϊ(νι)κό οξύ: Βιολογικά μόρια (RNA και DNA) που δίνουν τη δυνατότητα στους οργανισμούς να αναπαραχθούν

Περισσότερη ορολογία

- ✓ Το **γονιδίωμα** είναι το πλήρες σύνολο του DNA ενός οργανισμού
 - ✓ Ένα βακτήριο περιέχει περίπου 600.000 **ζεύγη βάσεων DNA**
 - ✓ Το γονιδίωμα των ανθρώπων και των ποντικών περιέχει περίπου 3 δισεκατομμύρια
- ✓ Το ανθρώπινο γονιδίωμα διαθέτει 24 διαφορετικά χρωμοσώματα
 - ✓ Κάθε χρωμόσωμα περιέχει πολλά **γονίδια**
- ✓ **Γονίδια**
 - ✓ Βασικές φυσικές και λειτουργικές μονάδες της κληρονομικότητας
 - ✓ Συγκεκριμένες αλληλουχίες βάσεων DNA που κωδικοποιούν οδηγίες για την παρασκευή **πρωτεϊνών**
- ✓ **Πρωτεΐνες**
 - Αποτελούν τα συστατικά στοιχεία της κυτταρικής δομής
 - Μεγάλα, σύνθετα μόρια τα οποία αποτελούνται από μικρότερες υπομονάδες που ονομάζονται **αμινοξέα**

Όλες οι μορφές ζωής εξαρτώνται από 3 ζωτικά μόρια

✓ Μόρια DNA

- ✓ Αποθηκεύουν πληροφορίες για τον τρόπο λειτουργίας του κυττάρου

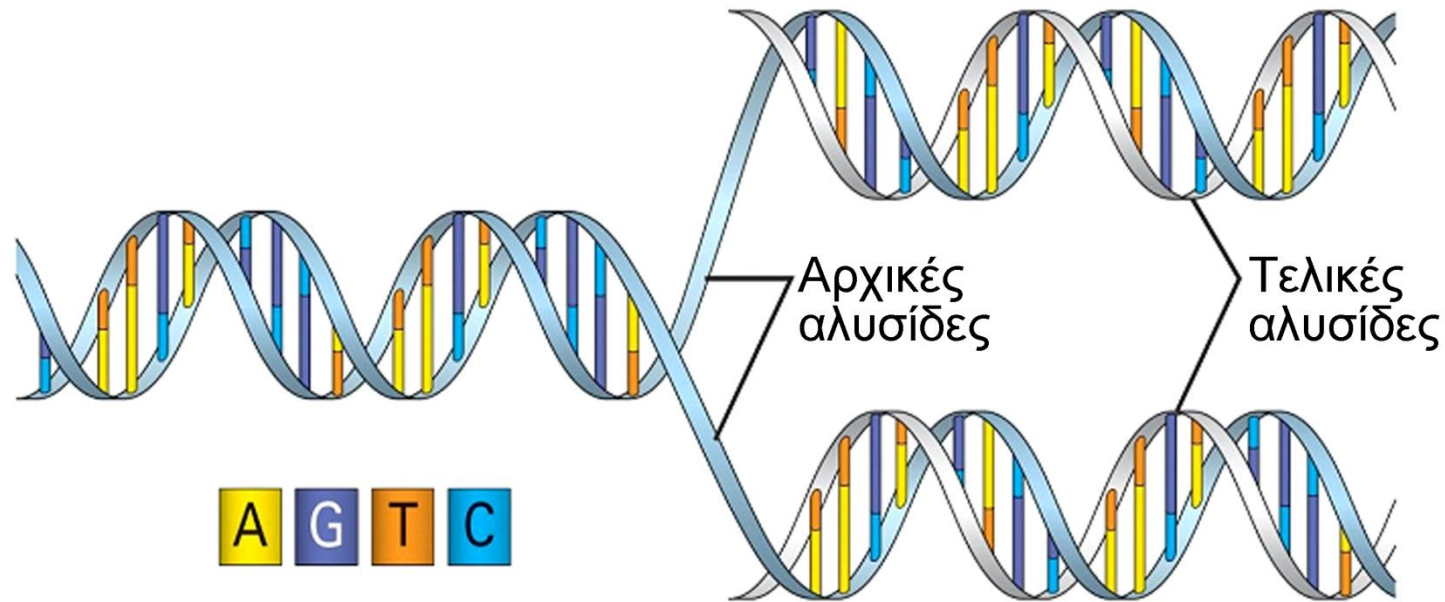
✓ Μόρια RNA

- ✓ Μεταφέρουν μικρά κομμάτια πληροφοριών σε διαφορετικά σημεία του κυττάρου
- ✓ Παρέχουν πρότυπα για τη σύνθεση πρωτεϊνών

✓ Πρωτεΐνες

- Σχηματίζουν ένζυμα που στέλνουν σήματα σε άλλα κύτταρα και ρυθμίζουν τις δραστηριότητες των γονιδίων
- Σχηματίζουν τα κύρια συστατικά μέρη του σώματος (π.χ., τα μαλλιά, το δέρμα, κλπ.)

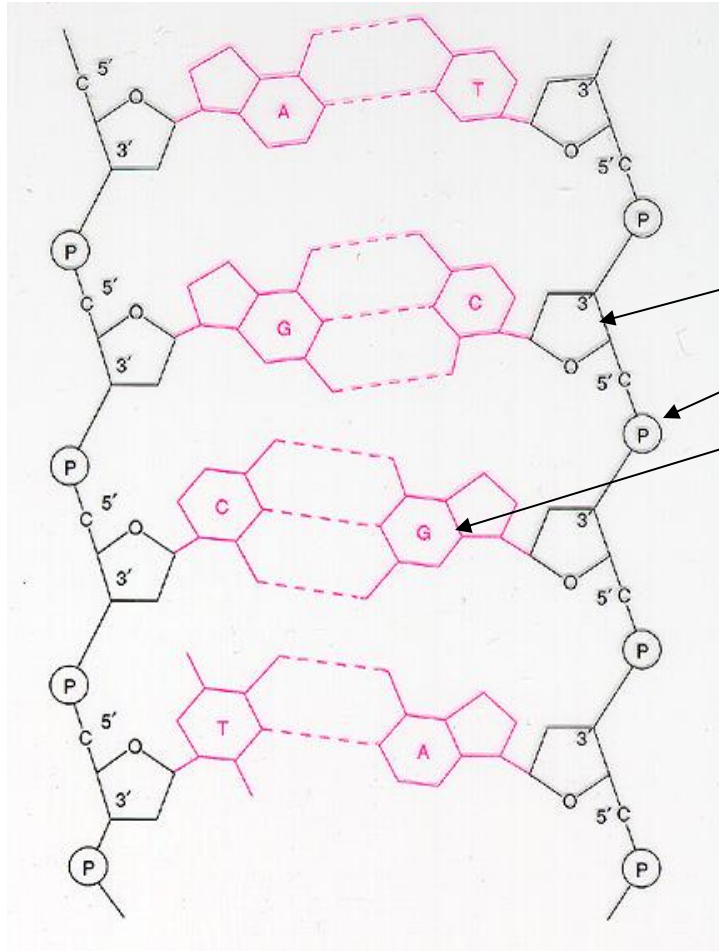
DNA: Ο κώδικας της ζωής



Η δομή και ο κώδικας τεσσάρων γραμμάτων του γονιδιώματος που κωδικοποιούν για όλους τους ζωντανούς οργανισμούς

Αδενίνη (A), γουανίνη (G), θυμίνη (T), και κυτοσίνη (C): δημιουργούν ζεύγη A-T και C-G σε συμπληρωματικές αλυσίδες

DNA (συνέχεια)



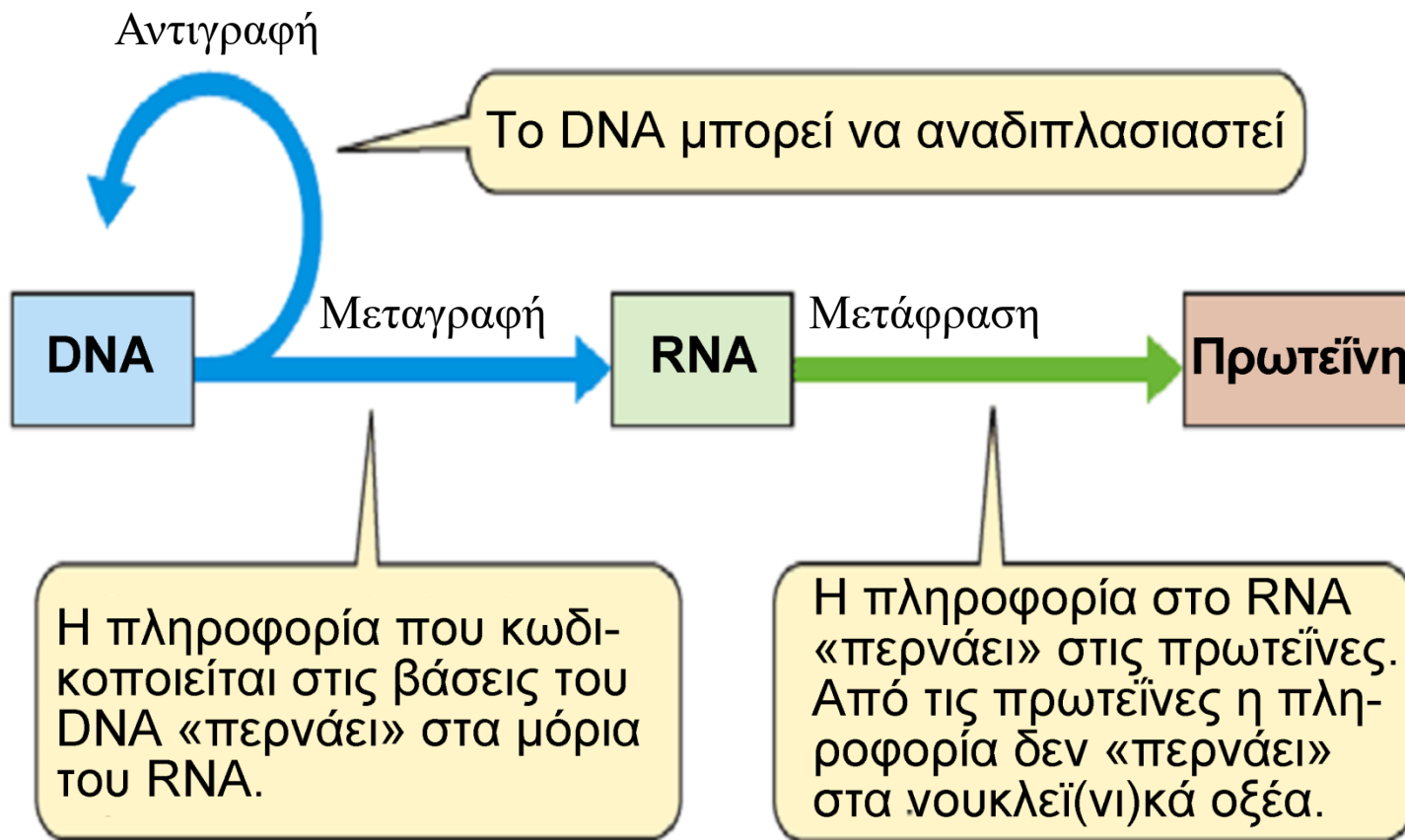
Το DNA έχει διπλή ελικοειδή δομή που αποτελείται από

- ένα μόριο σακχάρου
- μια φωσφορική ομάδα
- και μια βάση (A, C, G, T)

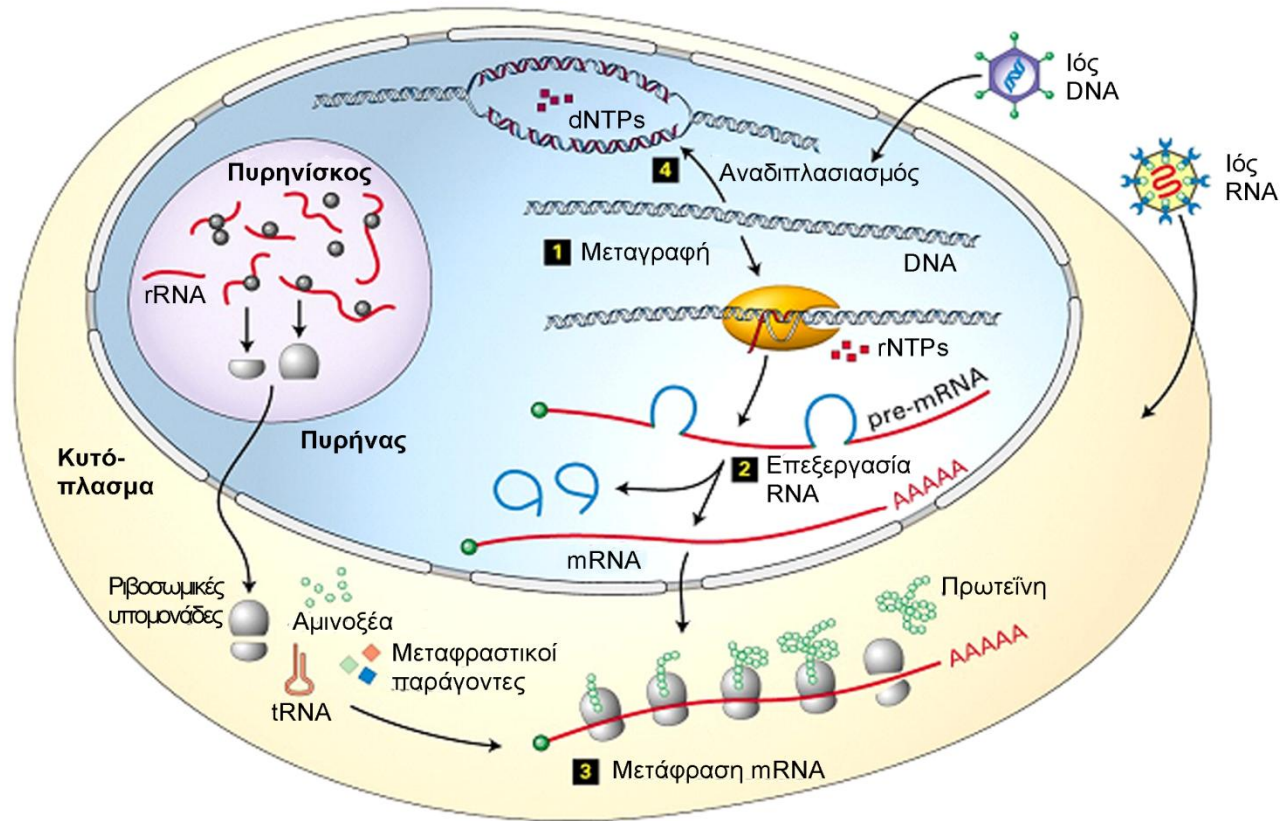
Χημική Βάση της Διαφοράς

- Κάθε νουκλεοτίδιο αποτελείται από:
 - Φωσφορική ομάδα ($-\text{PO}_4$)
 - Δεοξυριβόζη (σάκχαρο 5 ατόμων άνθρακα)
 - Αζωτούχα βάση (A, T, G, C)
- 5' άκρο: Φωσφορική ομάδα συνδεδεμένη στον 5ο άνθρακα
- 3' άκρο: Υδροξυλομάδα ($-\text{OH}$) στον 3ο άνθρακα

Το DNA, το RNA, και η ροή των πληροφοριών



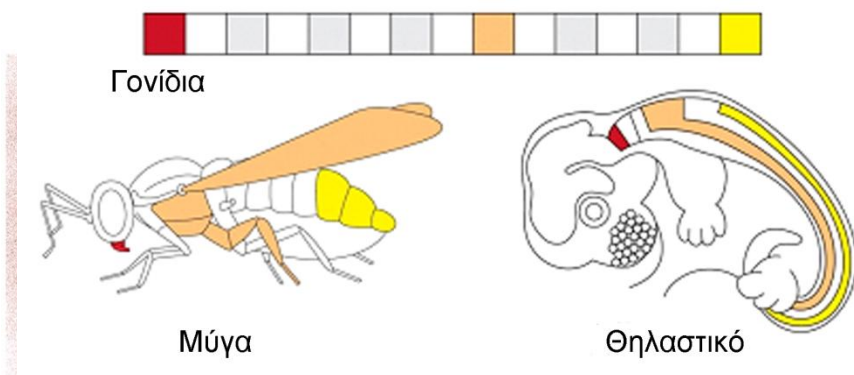
Από το DNA, στο RNA, στις πρωτεΐνες



Το γονίδιο εκφράζεται σε δύο βήματα

- 1) Μεταγραφή: σύνθεση RNA
- 2) Μετάφραση: σύνθεση πρωτεϊνών

DNA: η γενετική σύσταση



Τα γονίδια κληρονομούνται και εκφράζονται

- **γονότυπος** (γενετική σύσταση)
- **φαινότυπος** (φυσική έκφραση)



- Στα αριστερά, φαίνονται οι φαινότυποι του ματιού για τα γονίδια των πράσινων και μαύρων ματιών.

Πληροφορίες του κυττάρου: το εγχειρίδιο οδηγιών της ζωής

- ✓ Το DNA, το RNA, και οι πρωτεΐνες αποτελούν παραδείγματα συμβολοσειρών που έχουν γραφτεί είτε με το αλφάβητο τεσσάρων γραμμάτων του DNA και του RNA (A C G T/U)
- ✓ είτε με το αλφάβητο είκοσι γραμμάτων (αμινοξέων) των πρωτεϊνών. Κάθε αμινοξύ κωδικοποιείται από μια τριάδα νουκλεοτιδίων που ονομάζεται κωδικόνιο (Leu, Arg, Met, κλπ.)

		Δεύτερο γράμμα				
		U	C	A	G	
Πρώτο γράμμα	U	UUU Φαινυ- λαανίνη UUA Λευκίνη	UCU Σερίνη UCC UCA UCG	UAU Τυροσίνη UAC UAA Κωδικόνιο τερματισμού UAG Κωδικόνιο τερματισμού	UGU Κυστεΐνη UGC UGA Κωδικόνιο τερματισμού UGG Τρυπτοφάνη	U C A G
	C	CUU Λευκίνη CUC CUA CUG	CCU Προλίνη CCC CCA CCG	CAU Ιστιδίνη CAC CAA Γλουταμίνη CAG	CGU Αργινίνη CGC CGA CGG	U C A G
	A	AUU Ισολευκίνη AUC AUA AUG Μεθονίνη Κωδικόνιο εκκίνησης	ACU Θρεονίνη ACC ACA ACG	AAU Ασπαραγίνη AAC AAA Λυσίνη AAG	AGU Σερίνη AGC AGA Αργινίνη AGG	U C A G
	G	GUU Βαλίνη GUC GUA GUG	GCU Αλανίνη GCC GCA GCG	GAU Ασπα- ραγμικό οξύ GAC GAA Γλουταμινικό οξύ GAG	GGU Γλυκίνη GGC GGA GGG	U C A G

Βασικές Έννοιες (2)

- ✓ Το DNA αποτελείται από 1 διπλή έλικά βάσεων. Οι βάσεις ενώνονται σε συγκεκριμένη σειρά και αποθηκεύουν τη γενετική πληροφορία κάθε οργανισμού: A (αδενίνη), T (θυμίνη), C (κυτοσίνη), G (γουανίνη)
- ✓ Κάθε μόριο του DNA μπορεί να θεωρηθεί ως μία συμβολοσειρά με αλφάβητο {A,C,T,G}
- ✓ Διπλή έλικά, η γνώση της μίας έλικας συνεπάγεται και τη γνώση της άλλης (A-T, C-G)

Γονιδιώματα (Genomes)

- ✓ Ο όρος γονιδίωμα-genome, αναφέρεται σε ολόκληρη την ακολουθία DNA ενός ζωντανού οργανισμού,
- ✓ Το ανθρώπινο γονιδίωμα αποτελείται από 46 χρωμοσώματα,
- ✓ Κάθε κύτταρο περιλαμβάνει ολόκληρο το γονιδίωμα ενός οργανισμού (διαφοροποίηση ευκαρυωτικών από προκαρυωτικά κύτταρα)

Πρωτεΐνες - Proteins

- ✓ Οι πρωτεΐνες είναι μόρια που αποτελούνται από ένα ή περισσότερα πολυπεπτίδια,
- ✓ Ένα πολυπεπτίδιο, είναι ένα πολυμερές που δομείται από αμινοξέα,
- ✓ Τα κύτταρα κατασκευάζουν τις πρωτεΐνες τους από 20 διαφορετικά αμινοξέα,
- ✓ Μια ακολουθία πρωτεΐνης μπορεί να θεωρηθεί ως μια συμβολοσειρά, από ένα αλφάβητο 20 χαρακτήρων, $\Sigma = \{\text{Ala, Arg, Asp, Asn, Cys, Glu, Gln, Gly, Hsi, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val}\}$.

Γονίδια - Genes

- ✓ Τα γονίδια αποτελούν τις βασικές μονάδες της κληρονομικότητας,
- ✓ Ένα γονίδιο κωδικοποιεί-encodes μία πρωτεΐνη αφού αποθηκεύει τις απαραίτητες πληροφορίες για την κατασκευή της,
- ✓ Το ανθρώπινο γονιδίωμα αποτελείται από περίπου ~40.000 γονίδια

Το Βασικό Δόγμα της Μοριακής Βιολογίας

✓ DNA (RNA-polymerase) -> pre-RNA (T->U (uracil))

✓ Pre-RNA (Spliceosome) -> RNA

✓ RNA (Ribosome) -> Protein

(the first step for eukariotes)

-- Η διαδικασία εμπλέκει το τμήμα του DNA που κωδικοποιεί ένα γονίδιο

-- Ο όρος έκφραση γονιδίου αναφέρεται στην παραγωγή του RNA από το γονίδιο.

✓ Κεντρικό δόγμα

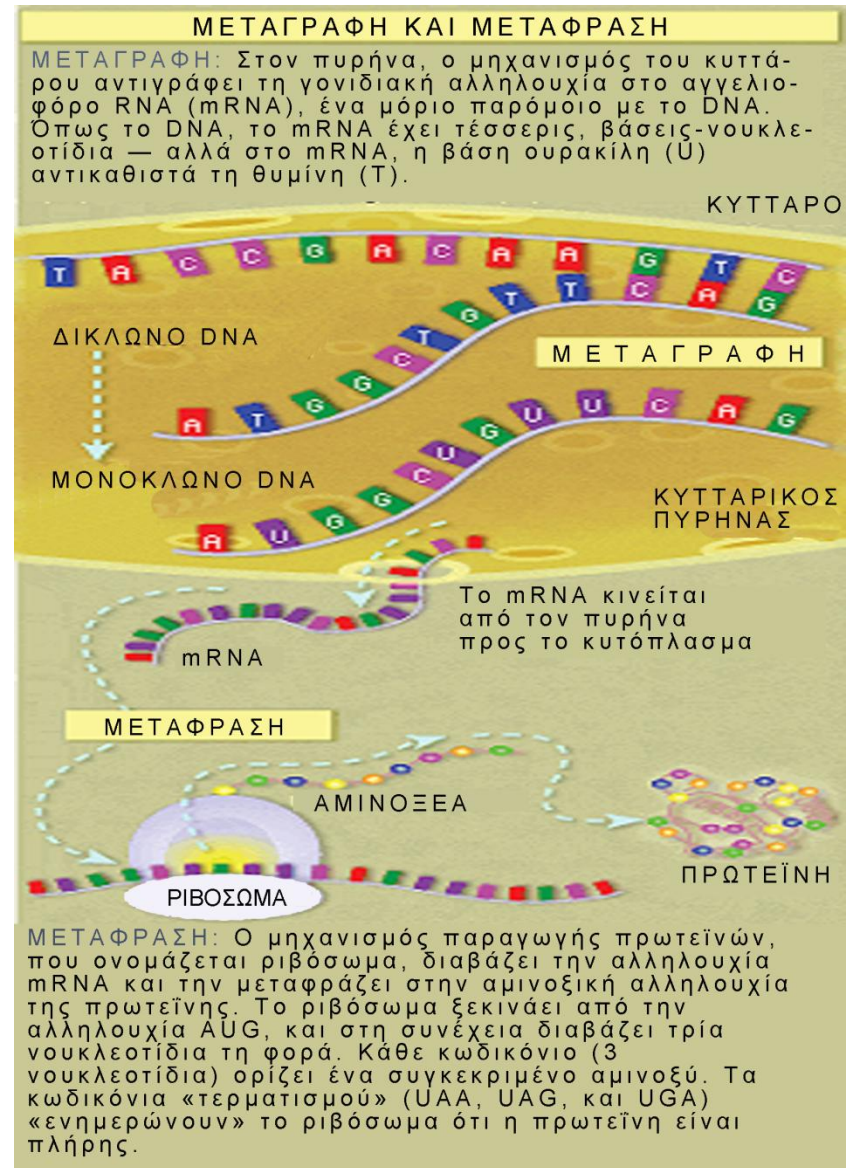
(DNA → RNA → πρωτεΐνη) Το δόγμα ότι το DNA διευθύνει τη μεταγραφή του σε RNA, το οποίο μεταφράζεται μετά σε πρωτεΐνη.

✓ Μεταγραφή

(DNA → RNA) Η διαδικασία που μεταφέρει γενετικές πληροφορίες από το DNA στο RNA.

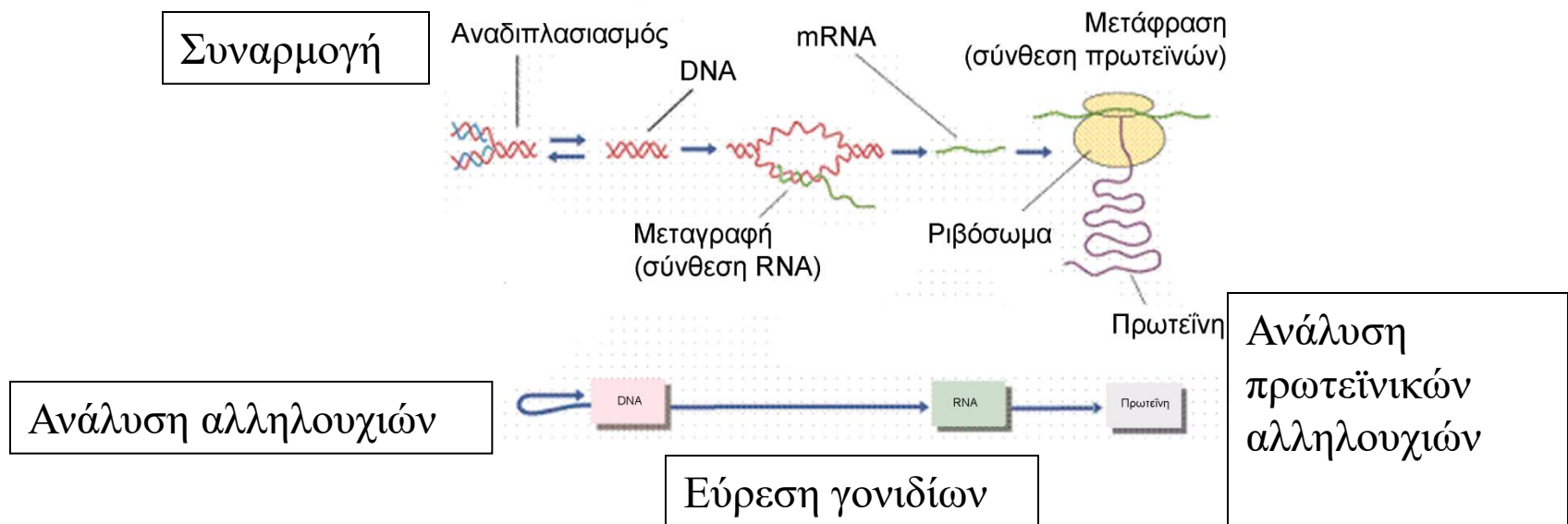
✓ Μετάφραση

(RNA → πρωτεΐνη) Η διαδικασία μετασχηματισμού του RNA σε πρωτεΐνη, όπως καθορίζεται από τον γενετικό κώδικα.



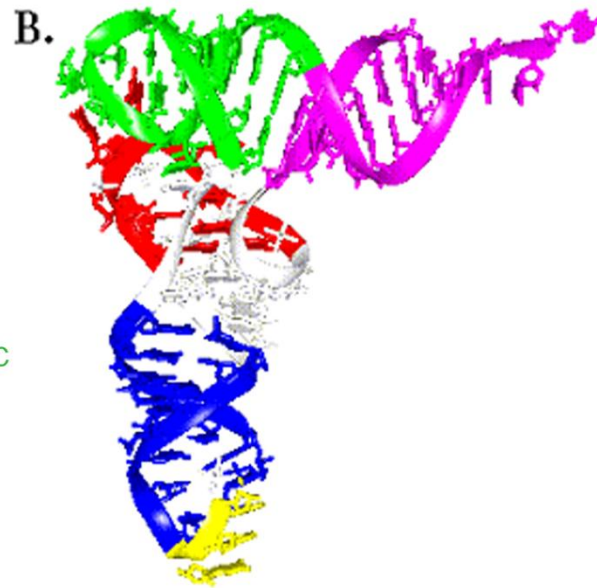
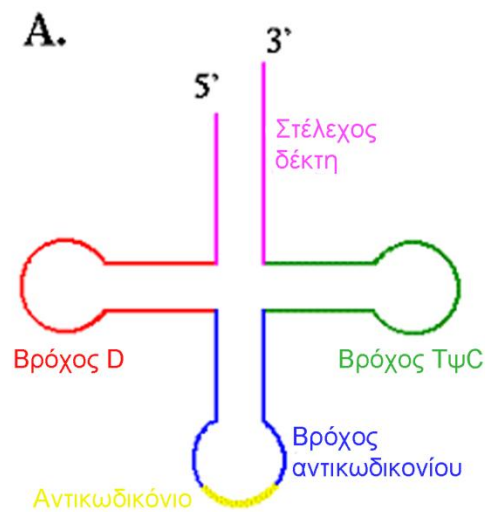
Το κεντρικό δόγμα της βιολογίας

Οι πληροφορίες για την κατασκευή πρωτεϊνών είναι αποθηκευμένες στο DNA. Υπάρχει μια διαδικασία (μεταγραφή και μετάφραση) με την οποία το DNA μετατρέπεται σε πρωτεΐνη. Με την κατανόηση της διαδικασίας αυτής και του τρόπου με τον οποίο ρυθμίζεται, μπορούμε να κάνουμε προβλέψεις και να φτιάξουμε μοντέλα των κυττάρων.



RNA

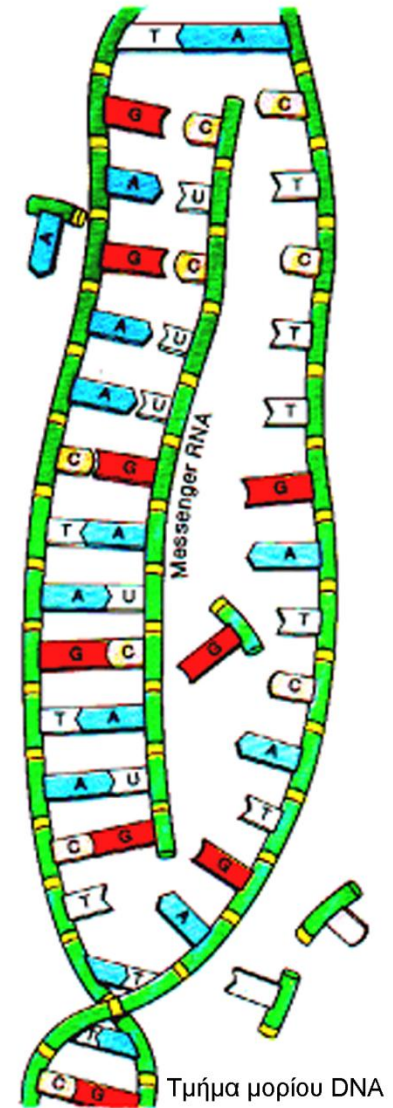
- ✓ Το RNA είναι χημικά παρόμοιο με το DNA. Συνήθως αποτελείται από μία μόνο αλυσίδα. Η T (θυμίνη) αντικαθίσταται από τη U (ουρακίλη).
- ✓ Μερικές μορφές του RNA μπορούν να σχηματίσουν δευτεροταγείς δομές «δημιουργώντας ζεύγη» με τον εαυτό τους. Αυτό μπορεί να αλλάξει τις ιδιότητές τους δραματικά. Το DNA και το RNA μπορούν να «ζευγαρώσουν» μεταξύ τους.



Γραμμική και τριδιάστατη
προβολή του tRNA :

<http://www.cgl.ucsf.edu/home/glasfeld/tutorial/trna/trna.gif>

- ΥΠΟΜΝΗΜΑ
Τ = Θυμίνη
C = Κυτοσίνη
A = Αδερίνη
G = Γουανίνη



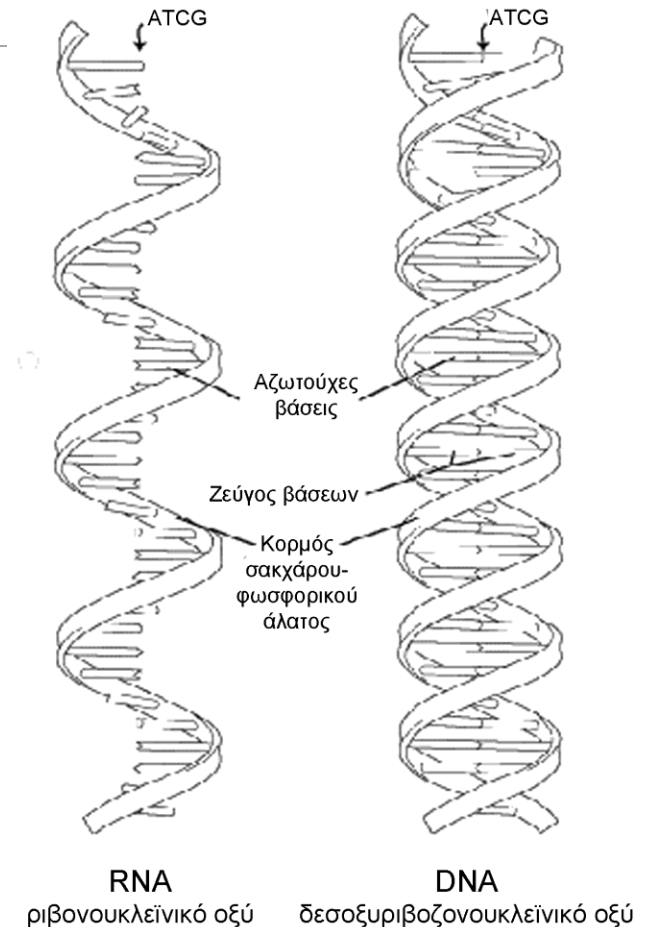
DNA → RNA: Μεταγραφή

Το DNA μεταγράφεται από μια πρωτεΐνη γνωστή ως πολυμεράση *RNA*

Η διαδικασία δημιουργεί μια αλυσίδα βάσεων που θα γίνουν mRNA

Το RNA και το DNA είναι παρόμοια, με τη διαφορά ότι το RNA έχει μία αλυσίδα και άρα είναι λιγότερο σταθερό από το DNA

- Επίσης, στο RNA, χρησιμοποιείται η βάση ουρακίλη (U) αντί της αντίστοιχης θυμίνης (T) στο DNA



Μεταγραφή (συνέχεια)

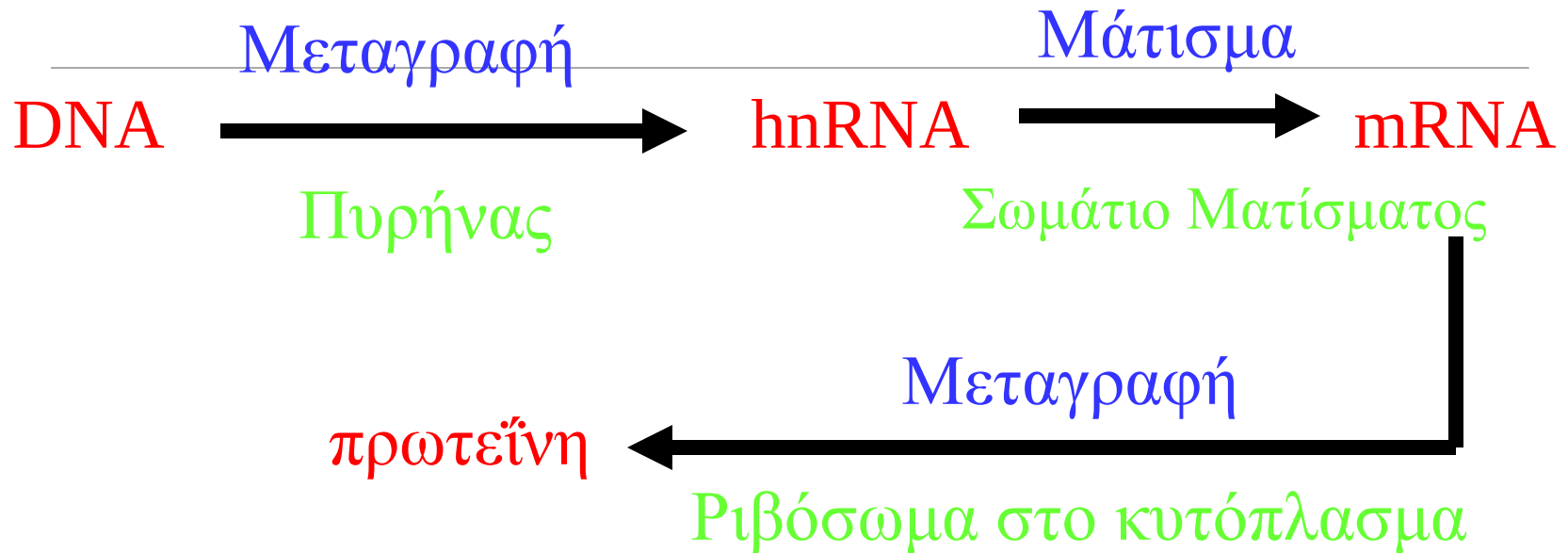
Η μεταγραφή ρυθμίζεται σε εξαιρετικό βαθμό. Το DNA βρίσκεται συνήθως σε μια πακεταρισμένη μορφή όπου δεν μπορεί να μεταγραφεί.

Η έναρξη της μεταγραφής απαιτεί έναν υποκινητή, μια μικρή συγκεκριμένη αλληλουχία DNA στην οποία μπορεί να προσδεθεί η πολυμεράση (~40 ζεύγη βάσεων που βρίσκονται πριν από το γονίδιο).

Η εύρεση των περιοχών των υποκινητών είναι ένα πρόβλημα που έχει λυθεί εν μέρει, και το οποίο σχετίζεται με την εύρεση μοτίβων.

Επίσης, ενδέχεται να υπάρχουν καταστολείς (repressors) και αναστολείς (inhibitors) που ενεργούν με διάφορους τρόπους για να σταματήσουν τη μεταγραφή. Αυτό δυσκολεύει πολύ την κατανόηση της ρύθμισης της μεταγραφής των γονιδίων.

Επανεξέταση του κεντρικού δόγματος



Κανόνας ζευγαρώματος βάσεων: η A και η T ή U συνδέονται με 2 δεσμούς υδρογόνου, ενώ οι G και C με 3 δεσμούς υδρογόνου.

Σημείωση: Ορισμένο mRNA παραμένει RNA (δηλαδή tRNA, rRNA).

Ορολογία του ματίσματος

Εξώνιο: Ένα τμήμα του γονιδίου που εμφανίζεται και στα αρχικά και στα ώριμα μετάγραφα του mRNA.

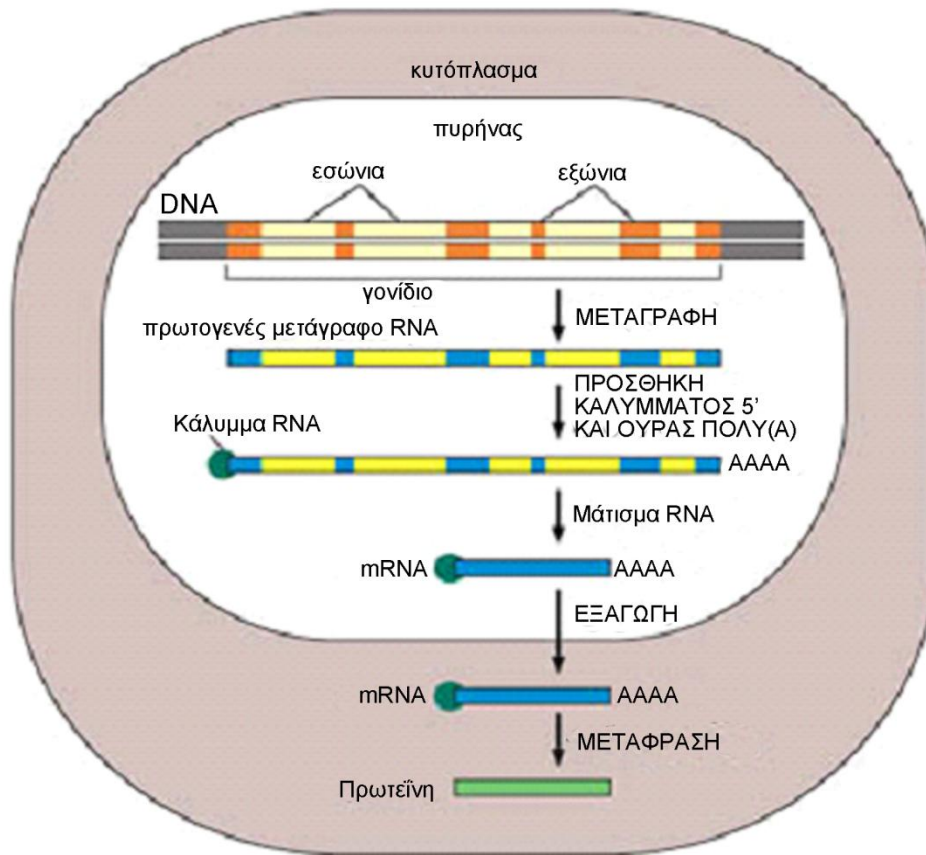
Εσώνιο: Ένα τμήμα του γονιδίου που μεταγράφεται αλλά αποκόπτεται πριν από τη μετάφραση.

Δομή βρόχου: Η δομή που σχηματίζει ένα εσώνιο στο mRNA κατά τη διάρκεια της αποκοπής/ματίσματος.

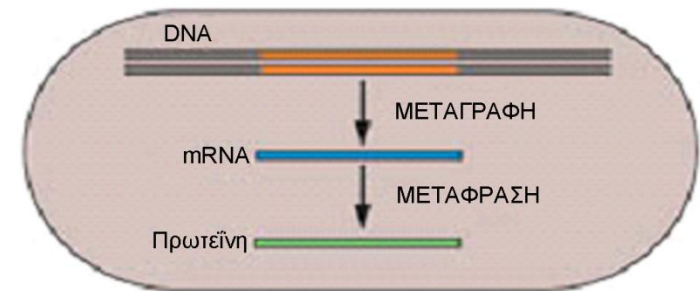
Σωμάτιο ματίσματος: Ένα οργανίδιο που εκτελεί τις αντιδράσεις ματίσματος μέσω των οποίων το pre-mRNA μετατρέπεται σε ώριμο mRNA.

Μάτισμα

(A) ΕΥΚΑΡΥΩΤΕΣ



(B) ΠΡΟΚΑΡΥΩΤΕΣ

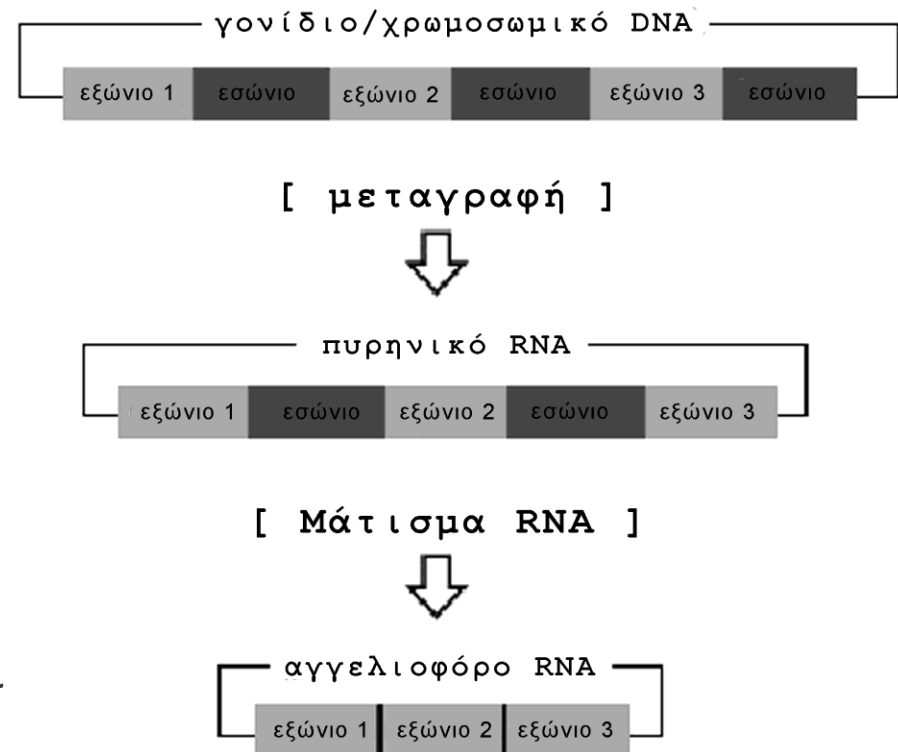


Μάτισμα (Ευκαρυώτες)

Το ανεπεξέργαστο RNA αποτελείται από εσώνια και εξώνια. Τα εσώνια αφαιρούνται προτού το υπόλοιπο RNA εκφραστεί και μετατραπεί σε πρωτεΐνη.

Μερικές φορές, εναλλακτικά ματίσματα μπορούν να δημιουργήσουν διαφορετικές λειτουργικές πρωτεΐνες.

Ένα τυπικό ευκαρυωτικό γονίδιο έχει 4-20 εσώνια. Ο εντοπισμός τους με αναλυτικές μεθόδους δεν είναι εύκολος.



Neil C. Jones, Pavel A. Pevzner, Εισαγωγή στους Αλγορίθμους Βιοπληροφορικής, Εκδόσεις Κλειδάριθμος (<http://bix.ucsd.edu/bioalgorithms/slides.php>)

Βασικό δόγμα Μοριακής Βιολογίας

- ✓ Το [μύριο του DNA](https://public.ornl.gov/site/gallery/default.cfm) (https://public.ornl.gov/site/gallery/default.cfm - U.S. Department of Energy Genome Programs)
- ✓ [Το βασικό δόγμα της Μοριακής Βιολογίας](http://www.accesexcellence.org/AB/GG)
(<http://www.accesexcellence.org/AB/GG>)
- ✓ [Αντιγραφή](#)
- ✓ [Μεταγραφή](#)
- ✓ [Σύνθεση](#)
- ✓ [Ο γενετικός κώδικας](#)
(<http://www.accesexcellence.org/AB/GG>)
- ✓ [Συνοψίζοντας](#)
(<http://www.doegenomes.org> - U.S. Department of Energy Genome Programs)
- ✓ [Μικρές αλλαγές συνεπάγονται διαφοροποιήσεις](#)
(<http://www.doegenomes.org> - U.S. Department of Energy Genome Programs)
- ✓ Genomics, and proteomics

Βασικές Έννοιες

- Το DNA και το RNA έχουν δύο κατευθύνσεις: $5' \rightarrow 3'$ και $3' \rightarrow 5'$
- Οι κατευθύνσεις βασίζονται στη δομή της δεοξυριβόζης (DNA) και ριβόζης (RNA)
- Η αντιπαράλληλη διάταξη είναι απαραίτητη για τη λειτουργία των νουκλεϊκών οξέων

Δομή του DNA και RNA

- DNA: Δίκλωνη έλικά με κατευθύνσεις $5' \rightarrow 3'$ και $3' \rightarrow 5'$
- RNA: Μονόκλωνο μόριο που συντίθεται προς την κατεύθυνση $5' \rightarrow 3'$
- Η φωσφοδιεστερική σύνδεση συνδέει το $5'$ άκρο ενός νουκλεοτιδίου με το $3'$ του επόμενου

Παράδειγμα Αλληλουχίας DNA

5' - ATCGTACG - 3'

3' - TAGCATGC - 5' (αντισυμπληρωματικός κλώνος)

- Ο ένας κλώνος είναι αντίστροφος του άλλου λόγω της αντιπαράλληλης φύσης του DNA

Σημασία της Κατεύθυνσης στην Αντιγραφή DNA

- Η DNA πολυμεράση προσθέτει νουκλεοτίδια μόνο στο 3' άκρο
- Η αντιγραφή προχωρά από 5' προς 3'
- Ο ηγούμενος κλώνος συντίθεται συνεχώς
- Ο καθυστερημένος κλώνος συντίθεται ασυνεχώς μέσω τμημάτων Okazaki

Σημασία της Κατεύθυνσης στη Μεταγραφή RNA

- Η RNA πολυμεράση διαβάζει το DNA από 3' → 5'
- Το παραγόμενο mRNA συντίθεται προς 5' → 3'
- Παράδειγμα:
 - DNA: 3' - TACGGCAT - 5'
 - mRNA: 5' - AUGCCGUA - 3' (αντιγραφή με αντικατάσταση T → U)

Σημασία της Κατεύθυνσης στη Μετάφραση

- Η μετάφραση γίνεται από 5' → 3'
- Τα ριβοσώματα διαβάζουν το mRNA σε τριάδες (κωδικόνια)
- Παράδειγμα:
 - mRNA: 5' - AUG CCG UAA - 3'
 - Κωδικόνια: AUG (Μεθειονίνη), CCG (Προλίνη), UAA (Σήμα λήξης)

Υπολογισμός Αντίστροφου Συμπληρώματος DNA

- Για τη σωστή ανάλυση, πρέπει να βρεθεί το αντίστροφο συμπλήρωμα:

- DNA: 5' - ATCGTACG - 3'

- Αντίστροφο Συμπλήρωμα: 3' - TAGCATGC - 5'

- Διαβάζοντας από 5' προς 3': 5' - GCATGCTA - 3'

Εφαρμογές σε Αλγορίθμους Βιοπληροφορικής

- BLAST, Smith-Waterman, Needleman-Wunsch: Σύγκριση αλληλουχιών
- Gene Prediction: Ανάλυση γονιδίων
- Genome Assembly: Σύνδεση αλληλουχιών με βάση τη σωστή κατεύθυνση
- Phylogenetics: Ανάλυση εξελικτικών σχέσεων με σωστό προσανατολισμό

Συμπέρασμα

- Η κατεύθυνση $3' \rightarrow 5'$ και $5' \rightarrow 3'$ είναι ζωτικής σημασίας για τη μοριακή βιολογία
- Οι αλγόριθμοι πρέπει να λαμβάνουν υπόψη τη σωστή κατεύθυνση για ακρίβεια
- Η λανθασμένη κατεύθυνση μπορεί να οδηγήσει σε λάθος βιολογικά συμπεράσματα
- Η σωστή ανάλυση αλληλουχιών είναι κρίσιμη στη βιοπληροφορική και τη γενετική

Εννοιες

- ✓ Genomics (μελέτη γονιδίων)
- ✓ Proteomics (μελέτη πρωτεϊνών)

Προβλήματα βιοπληροφορικής που βασίζονται στην αλληλουχία

Γονιδιωματική

- Συναρμογή τμημάτων της αλληλουχίας DNA
 - Δεν είναι εφικτό να διαβάσουμε ολόκληρη την αλληλουχία.
 - «Κόβουμε» σε μικρά τμήματα χρησιμοποιώντας ένζυμα περιορισμού.
 - Μετά, πρέπει να εκτελέσουμε τη συναρμογή τμημάτων. Επικαλυπτόμενες ομοιότητες με τμήματα που ταιριάζουν
 - NP-πλήρες πρόβλημα
- Εύρεση γονιδίων
 - Προσδιορίζουμε τα ανοιχτά πλαίσια ανάγνωσης (open reading frames, ORF)
 - Τα εξώνια αφαιρούνται μέσω ματίσματος.
 - «Σκουπίδια» στο ενδιάμεσο των γονιδίων

Προβλήματα βιοπληροφορικής που βασίζονται στην αλληλουχία

Πρωτεϊνωματική (proteomics)

- Προσδιορισμός των λειτουργικών περιοχών στην αλληλουχία της πρωτεΐνης
 - Προσδιορισμός λειτουργικών τμημάτων στις πρωτεΐνες
- Αναδίπλωση πρωτεϊνών
 - Αλληλουχία 1D → Δομή 3D
 - Τι είναι αυτό που οδηγεί αυτή τη διαδικασία;

DNA... και μετά;

DNA → μεταγραφή → RNA → μετάφραση → Πρωτεΐνη

Ριβονουκλεϊ(νι)κό οξύ (RNA)

- Είναι ο αγγελιοφόρος
 - Ένα προσωρινό αντίγραφο
- Γιατί όχι DNA → Πρωτεΐνη;
 - Το DNA βρίσκεται στον πυρήνα και οι πρωτεΐνες παρασκευάζονται έξω από τον πυρήνα
 - Προσθέτει ένα βήμα «διόρθωσης». (Μεταγραφή = DNA→RNA)

Άρα, στην πραγματικότητα... DNA → pre-mRNA → mRNA → Πρωτεΐνη

- Προκαρυώτες
 - Το γονίδιο είναι συνεχές. Η μετάφραση είναι εύκολη.
- Ευκαρυώτες
 - Εσώνια και εξώνια
 - Αρκετά εξώνια σε διαφορετικές θέσεις πρέπει να συνενωθούν μέσω ματίσματος για να δημιουργήσουν μια πρωτεΐνη. (Μάτισμα)
 - Pre-mRNA (μη ματισμένο RNA)
 - Το σωμάτιο ματίσματος αποκόπτει και αφαιρεί τα εσώνια, και φτιάχνει έτσι επεξεργασμένο mRNA.

Πρωτεΐνες

Εκτελούν τη χημεία του κυττάρου

- 20 αμινοξέα
-

Πιο πολύπλοκο πολυμερές από το DNA

- Μια αλληλουχία από 100 έχει 20^{100} συνδυασμούς
- Η ανάλυση αλληλουχιών είναι δύσκολη εξαιτίας της πολυπλοκότητας
- Μόνο λίγες από τις πιθανές αλληλουχίες χρησιμοποιούνται όντως στη ζωή.
(Ισχυρό επιχείρημα υπέρ της Εξέλιξης)

Το RNA μεταφράζεται σε πρωτεΐνη, η οποία μετά αναδιπλώνεται

- Από την αλληλουχία στη δομή 3D (Το πρόβλημα της αναδίπλωσης πρωτεϊνών)
- Η μετάφραση γίνεται στα ριβοσώματα
- 3 γράμματα του DNA $\rightarrow$ 1 αμινοξύ
 - 64 δυνατοί συνδυασμοί αντιστοιχίζονται σε 20 αμινοξέα
 - Εκφυλισμός του γενετικού κώδικα
 - Πολλά κωδικόνια αντιστοιχίζονται στα ίδια αμινοξέα

Γιατί χρειάζεται η βιοπληροφορική;

Η βιοπληροφορική είναι ο συνδυασμός της βιολογίας και της πληροφορικής.

Οι τεχνολογίες προσδιορισμού αλληλουχίας του DNA έχουν δημιουργήσει τεράστιες ποσότητες πληροφορίας, οι οποίες μπορούν να αναλυθούν αποτελεσματικά μόνο με υπολογιστές.

Μέχρι τώρα, έχει προσδιοριστεί η αλληλουχία για >100 000 είδη

- Ο άνθρωπος, ο αρουραίος, ο χιμπατζής, η κότα, και πολλά άλλα.

Καθώς αυξάνεται υπερβολικά ο όγκος και η πολυπλοκότητα των πληροφοριών, χρειάζονται περισσότερα υπολογιστικά εργαλεία για την ταξινόμηση των δεδομένων.

- Η βιοπληροφορική έρχεται να σώσει την κατάσταση!!!

Τι είναι η βιοπληροφορική;

Η βιοπληροφορική ορίζεται γενικά ως η ανάλυση, πρόβλεψη, και μοντελοποίηση των βιολογικών δεδομένων με τη βοήθεια των υπολογιστών



Βιο-πληροφορίες

Από τη στιγμή που ανακαλύψαμε τον τρόπο που το DNA λειτουργεί ως «εγχειρίδιο οδηγιών» της ζωής, η βιολογία έχει γίνει επιστήμη των πληροφοριών

Τώρα που έχουμε προσδιορίσει την αλληλουχία πολλών διαφορετικών οργανισμών, είμαστε σε θέση να βρούμε το νόημα του DNA μέσω της *συγκριτικής γονιδιωματικής*, κατά τρόπο παρόμοιο με τη συγκριτική γλωσσολογία.

Σιγά-σιγά, μαθαίνουμε το «συντακτικό» του DNA

Πληροφορίες αλληλουχιών

Πολλές γραπτές γλώσσες αποτελούνται από διαδοχικά σύμβολα

Ακριβώς όπως το ανθρώπινο κείμενο, οι γονιδιωματικές αλληλουχίες αναπαριστούν μια γλώσσα που γράφεται με τα σύμβολα A, T, C, G

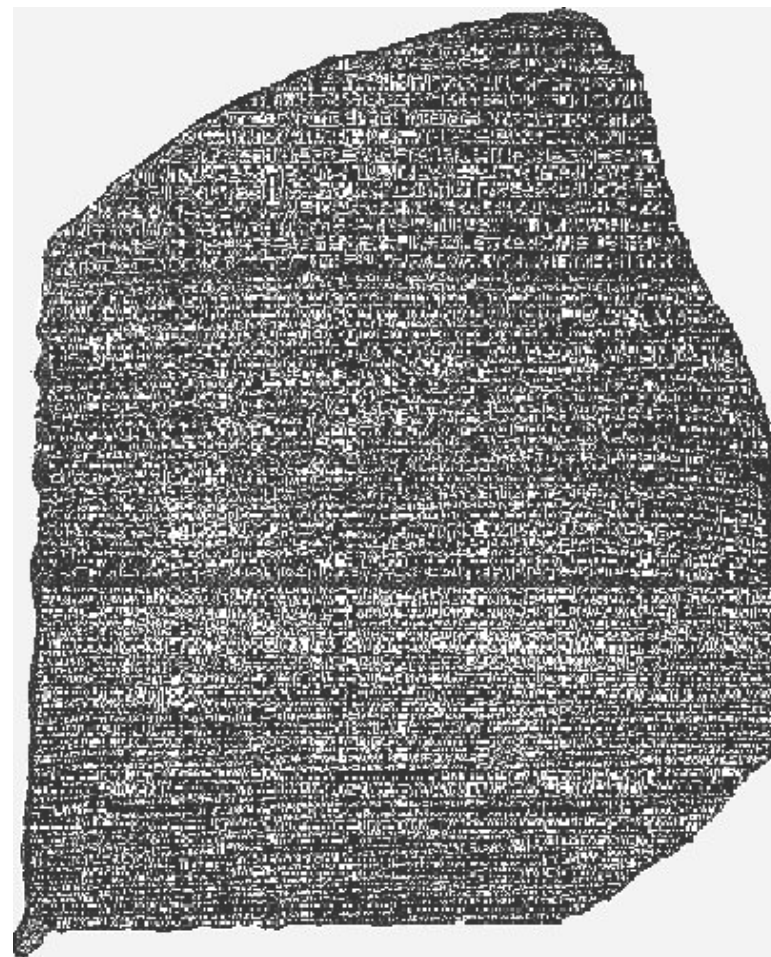
Πολλές τεχνικές αποκωδικοποίησης του DNA δεν διαφέρουν πολύ από τις αντίστοιχες τεχνικές αποκωδικοποίησης μιας αρχαίας γλώσσας

Η στήλη της Rosetta

Με τη στήλη της Rosetta, οι γλωσσολόγοι μπόρεσαν να λύσουν τον κώδικα των αιγυπτιακών ιερογλυφικών

Η εγχάρακτη επιγραφή στην ελληνική γλώσσα παρείχε ενδείξεις σχετικά με τη σημασία των ιερογλυφικών.

Αυτό είναι ένα παράδειγμα της συγκριτικής γλωσσολογίας

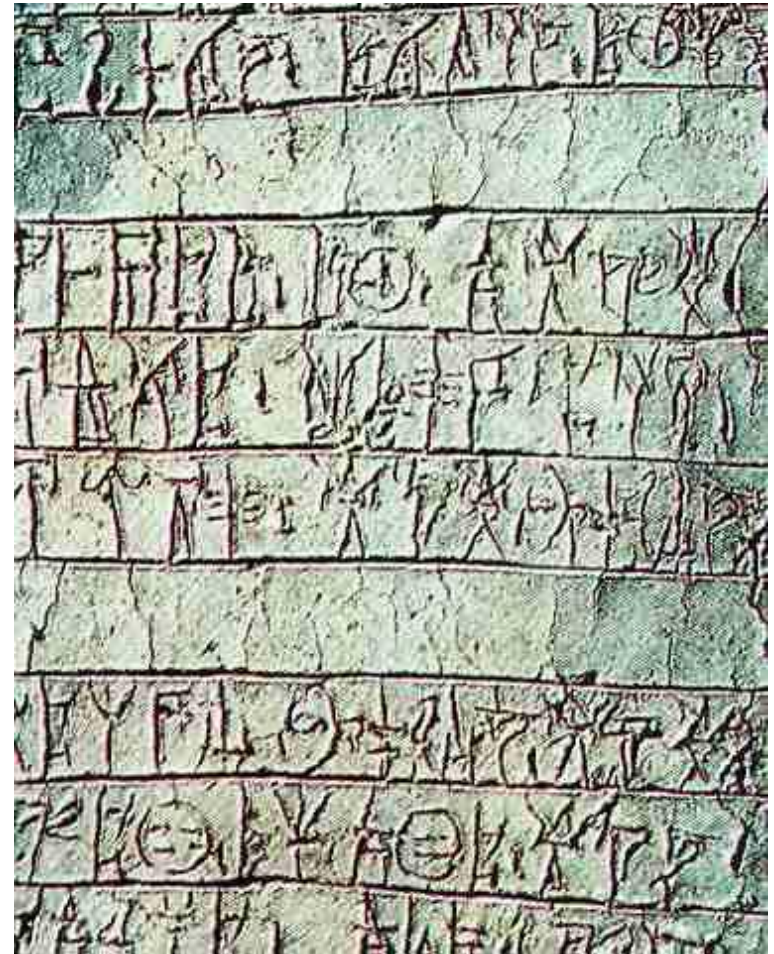


Γραμμική Β

Στις αρχές του εικοστού αιώνα, οι αρχαιολόγοι ανακάλυψαν πήλινες πινακίδες στο νησί της Κρήτης

Η άγνωστη γλώσσα ονομάστηκε «Γραμμική Β»

Η γραφή θεωρήθηκε ότι ανήκε σε μια αρχαία μινωική γλώσσα, και αποτέλεσε μυστήριο για 50 χρόνια



Γραμμική Β

Την ίδια εποχή που αποκρυπτογραφήθηκε η δομή του DNA, ο Michael Venter αποκωδικοποίησε τη Γραμμική Β χρησιμοποιώντας ειδικές γνώσεις μαθηματικής κρυπτανάλυσης

Πρόσεξε ότι μερικές λέξεις στη Γραμμική Β χρησιμοποιούνταν αποκλειστικά για το νησί, και υπέθεσε ότι αποτελούν ονόματα πόλεων

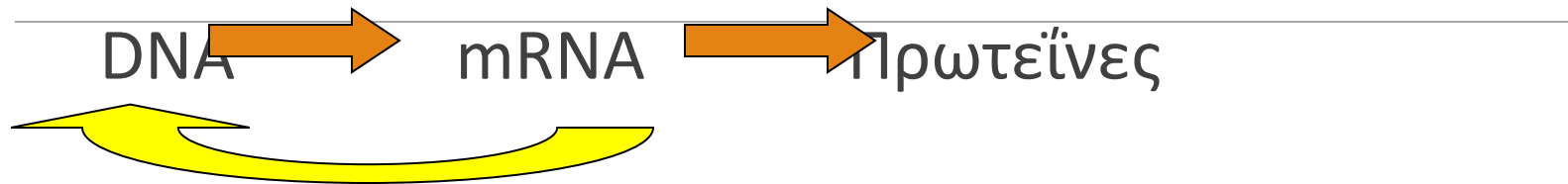
Με αυτή τη γνώση, μπόρεσε να αποκωδικοποιήσει τη γραφή, η οποία αποδείχθηκε ότι ήταν η ελληνική γλώσσα με διαφορετικό αλφάβητο

Αποκωδικοποίηση των αμινοξέων

Νωρίτερα, ένα πείραμα στις αρχές της δεκαετίας του 1900 έδειξε ότι όλες οι πρωτεΐνες αποτελούνται από αλληλουχίες 20 αμινοξέων

Εξαιτίας αυτού, κάποιοι υπέθεσαν ότι το «σχέδιο» της ζωής βρισκόταν στα πολυπεπίδια

Κεντρικό δόγμα



Το DNA στο χρωμόσωμα μεταγράφεται σε mRNA, το οποίο εξάγεται από τον πυρήνα στο κυτόπλασμα. Εκεί, μεταφράζεται σε πρωτεΐνη.

Αργότερα, ανακαλύφθηκε ότι μπορούμε επίσης να μεταβούμε από το mRNA στο DNA (ρετροϊοί).

Ακόμη, το mRNA μπορεί να υποβληθεί σε εναλλακτικό μάτισμα, και να προκύψουν έτσι διαφορετικά πρωτεϊνικά προϊόντα.

Από τη δομή στη λειτουργία

Η οργανική χημεία μας δείχνει ότι η δομή των μορίων προσδιορίζει τις δυνατές αντιδράσεις τους.

Μία μέθοδος μελέτης των πρωτεϊνών είναι να συμπεράνουμε τη λειτουργία τους με βάση τη δομή τους, ειδικά για τις ενεργές θέσεις.

Δύο σύντομες εφαρμογές της βιοπληροφορικής

BLAST (Basic Local Alignment Search Tool, Βασικό εργαλείο αναζήτησης τοπικών στοιχίσεων)

PROSITE (**PRO**tein **SITEs** and **patterns database**, Βάση δεδομένων για μοτίβα και θέσεις πρωτεϊνών)

BLAST

Ένα υπολογιστικό εργαλείο που μας επιτρέπει να συγκρίνουμε δεδομένες αλληλουχίες με τις καταχωρίσεις στις τρέχουσες βιολογικές βάσεις δεδομένων.

Ένα εξαιρετικό εργαλείο για την πρόβλεψη των λειτουργιών μιας αλληλουχίας που βασίζεται σε ομοιότητες στοίχισης με γνωστά γονίδια.

BLAST

NCBI Blast - Microsoft Internet Explorer provided by Compaq

File Edit View Favorites Tools Help

Address <http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi> Go Links

Google Linear B Search Web Search Site PageRank 71 blocked AutoFill Options Linear B

NCBI *formatting* **BLAST**

Nucleotide Protein Translations Retrieve results for an RID

Your request has been successfully submitted and put into the Blast Queue.

Query = (183 letters)

The request ID is 1082998002-15402-91580503850.BLASTQ3

Format! or **Reset all**

The results are estimated to be ready in 13 seconds but may be done sooner.

Please press "FORMAT!" when you wish to check your results. You may change the formatting options for your result via the form below and press "FORMAT!" again. You may also request results of a different search by entering any other valid request ID to see other recent jobs.

Format

Show ☒ Graphical Overview ☒ Linkout ☒ Sequence Retrieval ☒ NCBI-gi Alignment in HTML [format](#)

Use new formatter ☐ Masking Character Default(X for protein, n for nucleotide) Masking Color Black

Number of: Descriptions 100 Alignments 50

[Alignment view](#) Pairwise

start 181 RhinmanJngP23 Google Search: Ro... Google Search: Wa... NCBI Blast - Micros... 9:48 AM

Ο αρχικός ρόλος της βιοπληροφορικής

Σύγκριση αλληλουχιών

Αναζητήσεις σε βάσεις δεδομένων που περιέχουν αλληλουχίες

Σύγκριση βιολογικών αλληλουχιών

Needleman- Wunsch, 1970

- Αλγόριθμος δυναμικού προγραμματισμού για τη στοίχιση αλληλουχιών

	A	D	C	N	S	R	Q	C	L	C	R	P	M
A	8	7	6	6	5	4	4	3	3	2	1	0	0
S	7	7	6	6	6	4	4	3	3	2	1	0	0
C	6	6	7	6	5	4	4	4	3	3	1	0	0
S	6	6	6	5	6	4	4	3	3	2	1	0	0
N	5	5	5	6	5	4	4	3	3	2	1	0	0
R	4	4	4	4	4	5	4	3	3	2	2	0	0
C	3	3	4	3	3	3	3	4	3	3	1	0	0
K	3	3	3	3	3	3	3	3	3	2	1	0	0
C	2	2	3	2	2	2	2	3	2	3	1	0	0
R	2	1	1	1	1	2	1	1	1	1	2	0	0
D	1	2	1	1	1	1	1	1	1	1	1	0	0
P	0	0	0	0	0	0	0	0	0	0	0	1	0

Πρώιμη στοίχιση (ταίριασμα) αλληλουχιών

Η εύρεση των θέσεων κοπής για γνωστά ένζυμα περιορισμού μέσα σε μια αλληλουχία DNA (εξαιρετικά στοιχειώδης εφαρμογή)

Στοίχιση μιας πρωτεϊνικής αλληλουχίας με ένα μοτίβο βαθμολόγησης

Η συνένωση μικρών τμημάτων DNA σε μία συνεχόμενη αλληλουχία

- Η τεχνική χρησιμοποιήθηκε μαζί με την PCR ([Αλυσιδωτή αντίδραση πολυμεράσης](#)) και τον αυτοματοποιημένο προσδιορισμό αλληλουχίας HT (high throughput) για να δημιουργηθούν τα πάρα πολλά δεδομένα αλληλουχιών που έχουμε σήμερα

Βιολογικές βάσεις δεδομένων

Άπειρα βιολογικά δεδομένα και δεδομένα αλληλουχιών είναι διαθέσιμα δωρεάν μέσω των ηλεκτρονικών βάσεων δεδομένων

Υπολογιστικοί αλγόριθμοι χρησιμοποιούνται για την αποδοτική αποθήκευση τεράστιων ποσοτήτων βιολογικών δεδομένων

Παραδείγματα

NCBI GeneBank <http://ncbi.nih.gov>

Τεράστια συλλογή βάσεων δεδομένων, από τις οποίες ξεχωρίζει η βάση δεδομένων με αλληλουχίες νουκλεοτιδίων

Protein Data Bank <http://www.pdb.org>
Βάση δεδομένων με τριτοταγείς δομές πρωτεϊνών

SWISSPROT <http://www.expasy.org/sprot/>
Βάση δεδομένων με σχολιασμένες αλληλουχίες πρωτεϊνών

PROSITE <http://kr.expasy.org/prosite>
Βάση δεδομένων με μοτίβα ενεργών θέσεων πρωτεϊνών

Βάση δεδομένων PROSITE

Βάση δεδομένων με ενεργές θέσεις πρωτεϊνών.

Ένα εξαιρετικό εργαλείο που προβλέπει την ύπαρξη ενεργών θέσεων σε μια άγνωστη πρωτεΐνη με βάση μια πρωτοταγή αλληλουχία.

PROSITE

The screenshot shows a Microsoft Internet Explorer browser window with the title "List of PROSITE documentation entries - Microsoft Internet Explorer provided by Compaq". The address bar displays "http://www.expasy.org/cgi-bin/prosite-list.pl". The search bar contains the text "PROSITE". The main content area features the PROSITE logo and the text "PROSITE Database of protein families and domains". Below this, there is a section titled "Browse PROSITE documentation entries Release 18.26, of 26-Apr-2004". A list of links is provided, including "Post-translational modifications", "Compositional biased regions", "Domains", "DNA or RNA associated proteins", "Enzymes", "Electron transport proteins", "Other transport proteins", "Structural proteins", "Receptors", "Cytokines and growth factors", "Hormones and active peptides", "Toxins", "Inhibitors", "Protein secretion and chaperones", and "Others". A list of documentation entries is shown, with the first entry being "PD000020 2 Kringle domain signature and profile". The entry is marked with an asterisk, indicating it has been updated since the last release. The text explains that the asterisk indicates an update and that the accession number is PD000020, which describes two patterns.

start 181 RhinmanJngP23 Google Search... Google Search... List of PROSIT... untitled - Paint 9:52 AM

Ανάλυση αλληλουχιών

Μερικοί αλγόριθμοι αναλύουν βιολογικές αλληλουχίες για να βρουν μοτίβα

- Θέσεις ματίσματος του RNA
- ORFs (ανοιχτά πλαίσια ανάγνωσης)
- Τάσεις των αμινοξέων σε μια πρωτεΐνη
- Συντηρημένες περιοχές σε
 - αλληλουχίες AA [πιθανή ενεργή θέση]
 - DNA/RNA [πιθανή θέση πρόσδεσης πρωτεΐνης]

Άλλοι κάνουν προβλέψεις με βάση την αλληλουχία

- Αναδίπλωση δευτεροταγούς δομής πρωτεϊνών/RNA

Η αλληλουχία προσδιορίστηκε, τι ακολουθεί μετά;
Η ανίχνευση φυλογενετικών σχέσεων

- Βρίσκουμε οικογενειακές σχέσεις μεταξύ των ειδών, παρακολουθώντας τις ομοιότητες μεταξύ τους.

Σχολιασμός γονιδίων (συνεργατική γονιδιωματική)

- Σύγκριση παρόμοιων ειδών

Προσδιορισμός ρυθμιστικών δικτύων

- Οι μεταβλητές που καθορίζουν τον τρόπο αντίδρασης του σώματος σε ορισμένα ερεθίσματα.

Πρωτεϊνωματική

- Από την αλληλουχία DNA σε μια αναδιπλωμένη πρωτεΐνη

Μοντελοποίηση

Η μοντελοποίηση βιολογικών διεργασιών μας δείχνει αν κατανοούμε μια δεδομένη διεργασία

Λόγω του μεγάλου πλήθους μεταβλητών που υπάρχουν στα βιολογικά προβλήματα, χρειαζόμαστε πανίσχυρους υπολογιστές για να αναλύσουμε ορισμένα βιολογικά ερωτήματα

Μοντελοποίηση πρωτεϊνών

Οι αλγόριθμοι απεικόνισης ενεργών θέσεων της κβαντικής χημείας μας επιτρέπουν να δούμε τους πιθανούς μηχανισμούς αντίδρασης και δημιουργίας δεσμών

Η ομόλογη μοντελοποίηση πρωτεϊνών (homologous protein modeling) είναι μια συγκριτική πρωτεϊνωματική μέθοδος με την οποία προσδιορίζεται η τριτοταγής δομή μιας άγνωστης πρωτεΐνης

Οι αλγόριθμοι πρόβλεψης τριτοταγούς αναδίπλωσης απέχουν πολύ από το ιδανικό, αλλά μπορούμε να προβλέψουμε τη δευτεροταγή δομή με ακρίβεια ~80%.

Μοντελοποίηση ρυθμιστικών δικτύων

Τα πειράματα με μικροσυστοιχίες DNA μας επιτρέπουν να συγκρίνουμε τις διαφορές της έκφρασης για δύο διαφορετικές καταστάσεις

Οι αλγόριθμοι για την ομαδοποίηση των επιπέδων γονιδιακής έκφρασης συμβάλλουν στον εντοπισμό πιθανών ρυθμιστικών δικτύων

Άλλοι αλγόριθμοι εκτελούν στατιστική ανάλυση για να βελτιώσουν το λόγο σήματος προς θόρυβο

Μοντελοποίηση της βιολογίας των συστημάτων

Προβλέψεις των αλληλεπιδράσεων ολόκληρων κυττάρων.

- Διεργασίες οργανιδίων, μοντελοποίηση επιπέδων έκφρασης

Είναι σήμερα εφικτή για συγκεκριμένες διεργασίες (π.χ., μεταβολισμός στο βακτήριο *E. coli*, απλά κύτταρα)
Ανάλυση ισορροπίας ροής (flux balance analysis)

Το μέλλον...

Η βιοπληροφορική βρίσκεται ακόμα σε νηπιακή ηλικία
Έχουμε να μάθουμε ακόμα πολλά σχετικά με το πώς οι
πρωτεΐνες μπορούν να χειρίζονται μια αλληλουχία
ζευγών βάσεων με τόσο συγκεκριμένο τρόπο, το οποίο
έχει σαν αποτέλεσμα έναν πλήρως λειτουργικό
οργανισμό.

Επομένως, πώς μπορούμε να χρησιμοποιήσουμε
αυτές τις πληροφορίες προς όφελος της
ανθρωπότητας, χωρίς να τις καταχραστούμε;

Πηγές

Daniel Sam, “Greedy Algorithm”, παρουσίαση.

Glenn Tesler, “Genome Rearrangements in Mammalian Evolution: Lessons from Human and Mouse Genomes”, παρουσίαση.

Ernst Mayr, “What evolution is”.

Neil C. Jones, Pavel A. Pevzner, “Εισαγωγή στους αλγορίθμους βιοπληροφορικής”.

Alberts, Bruce, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter. Molecular Biology of the Cell. New York: Garland Science. 2002.

Mount, Ellis, Barbara A. List. Milestones in Science & Technology. Phoenix: The Oryx Press. 1994.

Voet, Donald, Judith Voet, Charlotte Pratt. Fundamentals of Biochemistry. New Jersey: John Wiley & Sons, Inc. 2002.

Campbell, Neil. Biology, 3η έκδοση. The Benjamin/Cummings Publishing Company, Inc., 1993.

Snustad, Peter και Simmons, Michael. Principles of Genetics. John Wiley & Sons, Inc, 2003.

Πηγές

Daniel Sam, “Greedy Algorithm”, παρουσίαση.

Glenn Tesler, “Genome Rearrangements in Mammalian Evolution: Lessons from Human and Mouse Genomes”, παρουσίαση.

Ernst Mayr, “What evolution is”.

Neil C. Jones, Pavel A. Pevzner, “Εισαγωγή στους αλγορίθμους βιοπληροφορικής”.

Alberts, Bruce, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter. Molecular Biology of the Cell. New York: Garland Science. 2002.

Mount, Ellis, Barbara A. List. Milestones in Science & Technology. Phoenix: The Oryx Press. 1994.

Voet, Donald, Judith Voet, Charlotte Pratt. Fundamentals of Biochemistry. New Jersey: John Wiley & Sons, Inc. 2002.

Campbell, Neil. Biology, 3η έκδοση. The Benjamin/Cummings Publishing Company, Inc., 1993.

Snustad, Peter και Simmons, Michael. Principles of Genetics. John Wiley & Sons, Inc, 2003.

Πηγές

- ✓ Neil C. Jones, Pavel A. Pevzner, “Εισαγωγή στους αλγορίθμους βιοπληροφορικής”.
- ✓ Alberts, Bruce, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter. Molecular Biology of the Cell. New York: Garland Science. 2002.
- ✓ Mount, Ellis, Barbara A. List. Milestones in Science & Technology. Phoenix: The Oryx Press. 1994.
- ✓ Voet, Donald, Judith Voet, Charlotte Pratt. Fundamentals of Biochemistry. New Jersey: John Wiley & Sons, Inc. 2002.
- ✓ Campbell, Neil. Biology, 3η έκδοση. The Benjamin/Cummings Publishing Company, Inc., 1993.
- ✓ Snustad, Peter και Simmons, Michael. Principles of Genetics. John Wiley & Sons, Inc, 2003.
- ✓ Daniel Sam, “Greedy Algorithm”, παρουσίαση.
- ✓ Glenn Tesler, “Genome Rearrangements in Mammalian Evolution: Lessons from Human and Mouse Genomes”, παρουσίαση.
- ✓ Ernst Mayr, “What evolution is”.

Σύγχρονοι Στόχοι Μοριακής Βιολογίας

- ✓ Ακολουθιοποίηση και σύγκριση των γονιδιωμάτων διαφορετικών οργανισμών (εξελικτική πορεία, επακριβής συσχέτιση).
- ✓ Αναγνώριση γονιδίων και καθορισμός των λειτουργιών που ρυθμίζουν (αναγνώριση σημείων πρόσδεσης πρωτεϊνών, και από εκεί αναγνώριση γονιδίων).
- ✓ Κατανόηση της γονιδιακής έκφρασης (κάθε γονίδιο δραστηριοποιείται μετά την παραγωγή της αντίστοιχης έκφρασης, μελέτη της διαδικασίας δραστηριοποίησης).^[?]
- ✓ Κατανόηση Γενετικών Ασθενειών (μετάλλαξη γονιδίων).

Τα δεδομένα από τα οποία προκύπτουν τα συμπεράσματα προέρχονται είτε από πειράματα, είτε από βιολογικές βάσεις δεδομένων

Επίλυση Υπολογιστικών Προβλημάτων με εργαλεία Βιοπληροφορικής

- ✓ Διασύνδεση της γονιδιακής πληροφορίας (σταδιακή ανάγνωση και επανασύνδεση)
- ✓ Σύγκριση ακολουθιών (αλγόριθμοι ανάκτησης πληροφορίας μέσω σχηματικών ομοιοτήτων).
- ✓ Κατηγοριοποίηση πρωτεϊνών.
- ✓ Εξαγωγή πληροφοριών από γονιδιακές εκφράσεις.
- ✓ Αναπαράσταση κυττάρων ως μεταγραφικών δικτύων.

Τομείς Έρευνας Βιοπληροφορικής

- ✓ Υλοποίηση και Σχεδιασμός υπολογιστικών εργαλείων για αυτόματη ανάκτηση γνώσης από Βάσεις Βιολογικών Δεδομένων.
- ✓ Ανάλυση Ακολουθιών Βιολογικών Δεδομένων
- ✓ Κατηγοριοποίηση Βιολογικών Δεδομένων
- ✓ Μοριακή Μοντελοποίηση
- ✓ Ανάλυση Πρωτεϊνών
- ✓ Σχεδιασμός Φαρμάκων με χρήση Η/Υ

Υπολογιστικά Εργαλεία

- ✓ Η διαχείριση δεδομένων της Μοριακής Βιολογίας παρουσιάζει αυξημένες απαιτήσεις και το μοντέλο της παραδοσιακής σχεσιακής Β.Δ. δε φαίνεται ικανοποιητικό. Στόχος είναι ο σχεδιασμός και υλοποίηση ενός μοντέλου που επιτρέπει την αυτόματη ανάκτηση γνώσης (automated knowledge discovery) από μεγάλο όγκο δεδομένων.
- ✓ Αναγνώριση κοινών δομικών χαρακτηριστικών, όχι μόνο σε επίπεδο ακολουθίας, αλλά και σε δισδιάστατο (2D) ή τρισδιάστατο (3D) επίπεδο.
- ✓ Ανίχνευση της ομοιότητας μεταξύ 2D ή 3D σχημάτων

Ανάλυση Ακολουθιών

- ✓ Ακριβές Ταίριασμα
- ✓ Προσεγγιστικό Ταίριασμα
- ✓ Στοίχιση (ολική ή τοπική)
- ✓ Ψάξιμο για μέγιστες κοινές υποακολουθίες
- ✓ Βάσεις δεδομένων βιολογικών μακρομορίων (συμπύεση)
- ✓ Αλγόριθμοι ψαξίματος, δεικτοδότηση και Cross-Referencing

Κατηγοριοποίηση/Ομαδοποίηση Βιολογικών Δεδομένων

- ✓ Η κατηγοριοποίηση γίνεται βάσει κοινών μοτίβων, δομικών ή λειτουργικών
- ✓ Μας ενδιαφέρουν και οι εφαρμογές ολοκλήρωσης διαφορετικού τύπου δεδομένων (data integration: sequences, 3D co-ordinates, functional knowledge)

Μοριακή Μοντελοποίηση

- ✓ Επιλογή του κατάλληλου μοντέλου που περιγράφει ικανοποιητικά τις ενδομοριακές συσχετίσεις του βιολογικού συστήματος που μελετάται.
- ✓ Υπολογισμός της ενεργειακής κατάστασης του συστήματος και ελαχιστοποίησή της και
- ✓ Ανάλυση των παραπάνω υπολογισμών και έλεγχος της τελικής διαμόρφωσης ώστε να ικανοποιούνται όλες οι συνθήκες και περιορισμοί που ο σχεδιαστής έχει θέσει.

Ανάλυση Πρωτεϊνών

- ✓ Καθορισμός της τρισδιάστατης δομής μιας πρωτεΐνης από την αμινοξέϊκή της ακολουθία
- ✓ Μελέτη της προσάραξη πρωτεϊνών (Protein Docking Problem) & πρωτεϊνών-DNA (DNA-Protein Docking Problem)

Σχεδιασμός Φαρμάκων με Η/Υ

- ✓ Οι σύγχρονοι ηλεκτρονικοί υπολογιστές αποθηκεύουν πολύτιμες πληροφορίες σχετικά με: 1) την τρισδιάστατη αρχιτεκτονική των μορίων, 2) τις φυσικοχημικές τους ιδιότητες, 3) τη σύγκριση ενός μορίου με άλλα μόρια, 4) τα σύμπλοκα μικρομορίων- μακρομορίων, 5) τις προβλέψεις για νέα μόρια.
- ✓ Ως πρώτος στόχος των επιστημόνων που ασχολούνται με τη σχεδίαση φαρμάκων με τη βοήθεια ηλεκτρονικού υπολογιστή τίθεται η αποτελεσματική απεικόνιση των δομών κανονικών και παθολογικών μορίων τα οποία στη συνέχεια συγκρίνονται με παθογενή ένζυμα και ενεργούς υποδοχείς αντίστοιχα οπότε και καθορίζεται ο στόχος σχεδιασμού.
- ✓ Έτσι αν γνωρίζουμε τη δομή μιας πρωτεΐνης και τον τρόπο που ο υποδοχέας ή η ενεργός περιοχή της δρα, μπορούμε να «χτίσουμε» και να προσομοιώσουμε την προσάραξη τους στην οθόνη του ηλεκτρονικού υπολογιστή εξοικονομώντας τον χρόνο και το κόστος που θα απαιτούσαν αντίστοιχες πειραματικές δοκιμές.

Σχεδιασμός Φαρμάκων με Η/Υ

- ✓ Αλγόριθμοι γραφικών
- ✓ Γεωμετρικοί υπολογισμοί
- ✓ Αριθμητικές μέθοδοι
- ✓ Γραφοθεωρητικές μέθοδοι

Βασικά Βήματα στη Σχεδίαση Φαρμάκων (in silico)

- ✓ Σχεδίαση της ένωσης οδηγού
- ✓ Βελτιστοποίηση ένωσης οδηγού
- ✓ In vitro και in vivo δοκιμές
- ✓ Τοξικολογικές δοκιμές
- ✓ Δοκιμές στον άνθρωπο
- ✓ Έλεγχος απόδοσης

Ορθολογικός Σχεδιασμός

Ακολουθίες Γονιδιωμάτων



Ακολουθίες πρωτεϊνών



Δομή πρωτεϊνών



Επιφάνεια πρωτεϊνών



Δυνάμεις πεδίου



Προσάραξη προσδέματος

Εύρεση γονιδίων

Πρόβλεψη δομής

Υπολογισμός γεωμετρίας

Μοριακές προσομοιώσεις

Προσάραξη μορίων

Προβλήματα

- ✓ η απουσία ενός γενικού και ενιαίου εργαλείου σχεδίασης μοριακών δομών που να περιλαμβάνει το σύνολο των βιολογικών μορίων,
- ✓ η αυξημένη υπολογιστική πολυπλοκότητα που εκφράζεται σε χρόνο και απαιτούμενους πόρους και η οποία αυξάνει εκθετικά με την αύξηση του μεγέθους του υπό εξέταση μορίου,
- ✓ η επιλογή του κατάλληλου μοντέλου αναπαράστασης (ανάλογα πάντα με το βιολογικό μόριο) και ο καθορισμός των κρίσιμων παραμέτρων (π.χ.: γεωμετρικές συντεταγμένες) που πρέπει να εξεταστούν ειδικότερα σε επίπεδο διανυσματικής γεωμετρίας,
- ✓ η αντιμετώπιση των σφαλμάτων στα δεδομένα εισόδου και η ανακατασκευή ενός τρισδιάστατου μοντέλου από ελλιπή ή λανθασμένα δεδομένα,
- ✓ η ταυτόχρονη αναπαράσταση ενός συνόλου φυσικοχημικών ιδιοτήτων (εντροπία, ενέργεια κ.α.) με τρόπο που η πληροφορία να είναι κατανοητή και ερμηνεύσιμη από τον ερευνητή,

Ερευνητικές κατευθύνσεις Βιοπληροφορικής

Data-driven advice for applying machine learning to bioinformatics problems

Randal S. Olson^{†*}, William La Cava^{†*}, Zairah Mustahsan, Akshay Varik, and Jason H. Moore[†]

*Institute for Biomedical Informatics, University of Pennsylvania
Philadelphia, PA 19104, USA*

[†]*E-mails: rso@randalolson.com, lacava@upenn.edu and jhmoore@upenn.edu*

As the bioinformatics field grows, it must keep pace not only with new data but with new algorithms. Here we contribute a thorough analysis of 13 state-of-the-art, commonly used machine learning algorithms on a set of 165 publicly available classification problems in order to provide data-driven algorithm recommendations to current researchers. We present a number of statistical and visual comparisons of algorithm performance and quantify the effect of model selection and algorithm tuning for each algorithm and dataset. The analysis culminates in the recommendation of five algorithms with hyperparameters that maximize classifier performance across the tested problems, as well as general guidelines for applying machine learning to supervised classification problems.

Keywords: machine learning; data science; best practices; benchmarking; bioinformatics

<https://arxiv.org/pdf/1708.05070.pdf>

(Βασισμένα στο πίνακα περιεχομένων του Srinivas Aluru, Sartaj Sahni
Handbook of Computational Molecular Biology , (Chapman & Hall/CRC
Computer and Information Science Series) 1st Edition, 2005,

- ✓ **Sequence alignment** (pairwise sequence alignment, spliced alignment and similarity based gene recognition, multiple sequence alignment, parametric sequence alignment)
- ✓ **String data structures** (lookup tables suffix trees and suffix arrays, suffix tree applications, enhanced suffix tree and applications)
- ✓ **Genome assembly and EST clustering** (Computational methods for genome assembly, assembling the human genome, comparative methods for sequence assembly, information theoretic approach to genome reconstruction, expressed sequence tags clustering and applications, algorithms for large scale clustering and assembly of biological sequence data)



(Βασισμένα στο πίνακα περιεχομένων του Srinivas Aluru, Sartaj Sahni
Handbook of Computational Molecular Biology , (Chapman & Hall/CRC
Computer and Information Science Series) 1st Edition, 2005,

- ✓ **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)
- ✓ **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)



(Βασισμένα στο πίνακα περιεχομένων του Srinivas Aluru, Sartaj Sahni
Handbook of Computational Molecular Biology , (Chapman &
Hall/CRC Computer and Information Science Series) 1st Edition, 2005,

✓ **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)

✓ **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)



Άρθρα – Βιβλία (1)

- ✓ J. Cohen, Bioinformatics-An introduction for Computer Scientists, ACM Computing Surveys, 36(2), 2004, 122-158
- ✓ N. Luscombe, D. Greenbaum, M. Gerstein, What is Bioinformatics? A proposed definition and overview of the field, Method Inform Med, 2001
- ✓ Thomas Lengauer, Raimund Mannhold, Hugo Kubinyi, and Hendrik Timmerman, Bioinformatics: from genomes to drugs, 2001 Wiley-VCH
- ✓ Thomas Lengauer, Bioinformatics: from genomes to therapies (two volumes) 2007 WILEY-VCH
- ✓ Neil Jones, Pavel Pevzner, An Introduction to Bioinformatics Algorithms, The MIT Press, 2004
- ✓ Srinivas Aluru, Sartaj Sahni Handbook of Computational Molecular Biology, (Chapman & Hall/CRC Computer and Information Science Series) 1st Edition, 2005,



Βιβλία (2)

Βιοπληροφορική, Δυνατότητες και Προοπτικές, Σοφία Κοσσίδα, 2008

Εισαγωγή στους Αλγοριθμούς Βιοπληροφορικής, Neil C. Jones, Pavel Pevzner, Εκδόσεις Κλειδάριθμος, 2008

[Bioinformatics Algorithms: An Active Learning Approach by Phillip Compeau, Pavel Pevzner \(2018\), third edition](#), Active Learning Publishers

Εισαγωγή στη Βιοπληροφορική, Arthur M. Lek, UTOPIA ΕΚΔΟΣΕΙΣ Μ. ΕΠΕ, 5η αμερικανική-1η ελληνική/2021

Βιοπληροφορική και Λειτουργική Γονιδιωματική, Jonathan Pevsner, ΑΚΑΔΗΜΑΪΚΕΣ ΕΚΔΟΣΕΙΣ Ι. ΜΠΑΣΔΡΑ & ΣΙΑ Ο.Ε., 1η/2019

Algorithms on Strings, Trees and Sequences, D. Gusfield, Cambridge University Press, 10th edition 2007

Μπάγκος Παντελεήμων, Βιοπληροφορική, Έκδοση: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών 2015(<https://repository.kallipos.gr/handle/11419/5016>).

Νικολάου, Χριστόφορος Χουβαρδάς, Παναγιώτης, Υπολογιστική Βιολογία, Έκδοση: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, 2015 (<https://repository.kallipos.gr/handle/11419/1577>).

Βιβλία (3)

Richard Durbin, Sean Eddy, Anders Krogh, Graeme Mitchison, Ανάλυση Βιολογικών Αλληλουχιών, Πιθανοκρατικά μοντέλα πρωτεϊνών και νουκλεϊκών όξεων, 2016.

Gonzalo Navarro, Compact Data Structures: A Practical Approach 2016
<https://www.cambridge.org/core/books/compact-data-structures/68A5983E6F1176181291E235D0B7EB44>

Βιολογία Κυττάρου - Μοριακή Προσέγγιση, Β. Μαρμάρας και Μ. Λαμπροπούλου - Μαρμάρα, Εκδόσεις ΤΥΡΟΡΑΜΑ, 2000

Bioinformatics- From Genomes to Drugs, Edited by Thomas Lengauer, Volumes I & II, WILEY-VCH, 2001

Fundamental Concepts of Bioinformatics, D. Krane and M.L.Raymer, Pearson Education, 2003

Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Edited by A.D.Baxevanis and B.F.F.Ouellette, Wiley-Liss, Inc., 2001

Main Journals

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

<http://bioinfolab.miamioh.edu/bioinfolab/journals.php>

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

Main Conferences

- ✓ RECOMB, (Research in Computational Molecular Biology)
- ✓ IEEE Computer Society Bioinformatics Conference
- ✓ PSB Pacific Symposium on Biocomputing
- ✓ ISMB Intelligent System for Molecular Biology

Ranking of Conferences: Bioinformatics

A+	• ACM RECOMB - Research in Computational Molecular Biology • ISMB - Intelligent Systems in Molecular Biology • PSB - Pacific Symposium on Biocomputing
A	• ECCB - European Conference on Computational Biology • ACM BCB - ACM Conference on Bioinformatics, Computational Biology and Biomedicine
B	• WABI - Algorithms in Bioinformatics • BIBE - IEEE International Symposium of Bioinformatic and Bioengineering

<https://personal.ntu.edu.sg/assourav/research/publications/ranking/conferences/bioinformatics.html>

Ιατρική Πληροφορική

Η ιατρική πληροφορική ορίζεται ως η τομή της επιστήμης των πληροφοριών, της πληροφορικής και της ιατρικής. Ασχολείται με τους πόρους, τις συσκευές και τις μεθόδους που απαιτούνται για να βελτιστοποιήσουν την απόκτηση, την αποθήκευση, την ανάκτηση και τη χρήση των πληροφοριών στην υγεία και στην βιοϊατρική.

Ο συνδυασμός ιατρικής πληροφορικής και βιοπληροφορικής μας οδηγεί στην **βιοϊατρική πληροφορική**.

Προχωρημένα θέματα βιοιατρικής

- ✓ Ιατρικές βάσεις δεδομένων και πρωτοβουλίες γενετικής
- ✓ Πληροφορίες πρωτεωμικής και ανάλυση
- ✓ Πρωτοβουλίες για επιλογή καταλληλότερης θεραπείας
- ✓ Πληροφορίες έκφρασης γονιδίων στην ιατρική διάγνωση και πρόγνωση
- ✓ Μοντελοποίηση και προσομοίωση βιολογικών δομών και διαδικασιών
- ✓ Ιατρική επισημείωση των βιολογικών βάσεων δεδομένων
- ✓ Λειτουργική και μοριακή επεξεργασία εικόνας
- ✓ Ενσωμάτωση των μοριακών δεδομένων ασθενών στους ηλεκτρονικούς ιατρικούς φακέλους
- ✓ Συστήματα λήψης κλινικών αποφάσεων
- ✓ Διεπαφές μοριακών πληροφοριών για τους ιατρούς
- ✓ Σηματολογική διαλειτουργικότητα και οντολογίες στη βιοϊατρική
- ✓ Τεχνολογίες για τη βιοϊατρική ολοκλήρωση πληροφοριών
- ✓ Πολυεπίπεδη μοντελοποίηση και κάθετη ολοκλήρωση πληροφοριών
- ✓ Εξόρυξη γνώσης από την βιοϊατρική πληροφορική

- ✓ Διαλειτουργικότητα δεδομένων και πρότυπα
- ✓ Βιοϊατρική πληροφορική στην Δημόσια Υγεία
- ✓ Σύνδεση βιοτραπεζών με βάσεις δεδομένων μεγάλης κλίμακας για να επιτρέψουν την εξόρυξη γνώσης
- ✓ Μητρώα που συνδέουν τα μοριακά, οικογενειακά, και κλινικά δεδομένα.
- ✓ Συστήματα πληροφοριών βιοάμυνας και δίκτυα
- ✓ Κάλυψη των αναγκών πληροφόρησης για την έρευνα στις μολυσματικές και τροπικές ασθένειες
- ✓ Αποτύπωση κινδύνου ασθενούς και διαχείριση τρόπου ζωής
- ✓ Επιβεβαίωση ταυτότητας και προσωπική γονιδιωματική
- ✓ Πληροφορική για την υποστήριξη φαρμακογενετικών και στρωματοποιημένων κλινικών δοκιμών
- ✓ Πληροφορική για να επιτρέψει την ανάπτυξη ιατρικών συσκευών και βιοαισθητήρων
- ✓ Εφαρμοσμένη φαρμακευτική ακρίβεια
- ✓ Εκτιμήσεις ασφάλειας και ακρίβειας
- ✓ Μετα-επιτήρηση μάρκετινγκ των φαρμάκων και φαρμακοεπαγρύπνηση
- ✓ Κλινικά και ηθικά ζητήματα σχετικά με την βιοϊατρική επεξεργασία δεδομένων

- ✓ Συστήματα λήψης κλινικών αποφάσεων
- ✓ Διεπαφές μοριακών πληροφοριών για τους ιατρούς
- ✓ Σημασιολογική διαλειτουργικότητα και οντολογίες στη βιοϊατρική
- ✓ Τεχνολογίες για τη βιοϊατρική ολοκλήρωση πληροφοριών
- ✓ Πολυεπίπεδη μοντελοποίηση και κάθετη ολοκλήρωση πληροφοριών
- ✓ Εξόρυξη γνώσης από την βιοϊατρική πληροφορική