

# **ΒΑΡΒΑΚΕΙΟ ΠΡΟΤΥΠΟ ΛΥΚΕΙΟ**

**ΟΜΙΛΟΣ ΒΙΟΛΟΓΙΑΣ**

**«Βασικές Αρχές Βιοπληροφορικής και  
Εφαρμογές στη Βιολογία»**

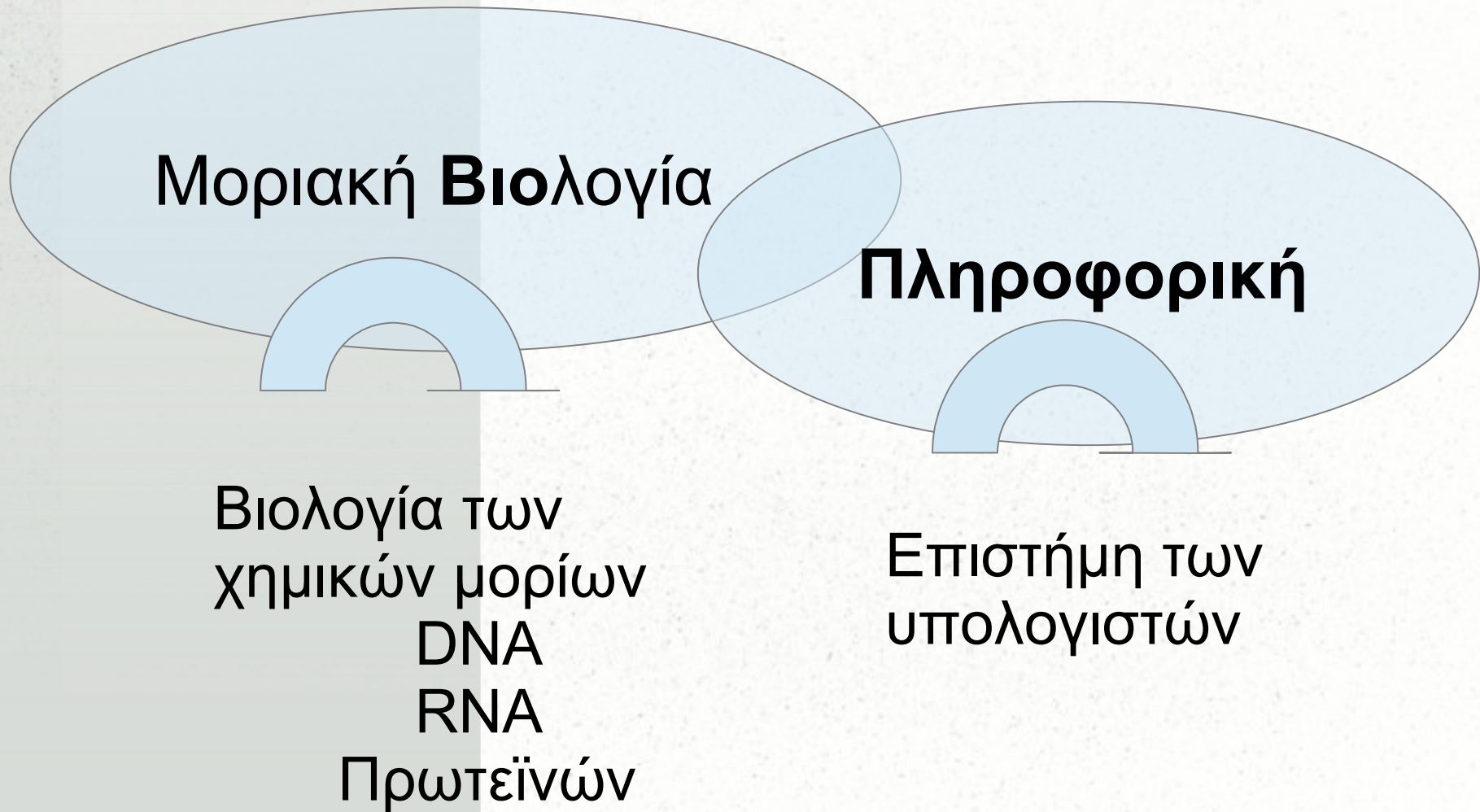
**Εισαγωγή στη Βιοπληροφορική**

---

**Μαρία Καπασά**

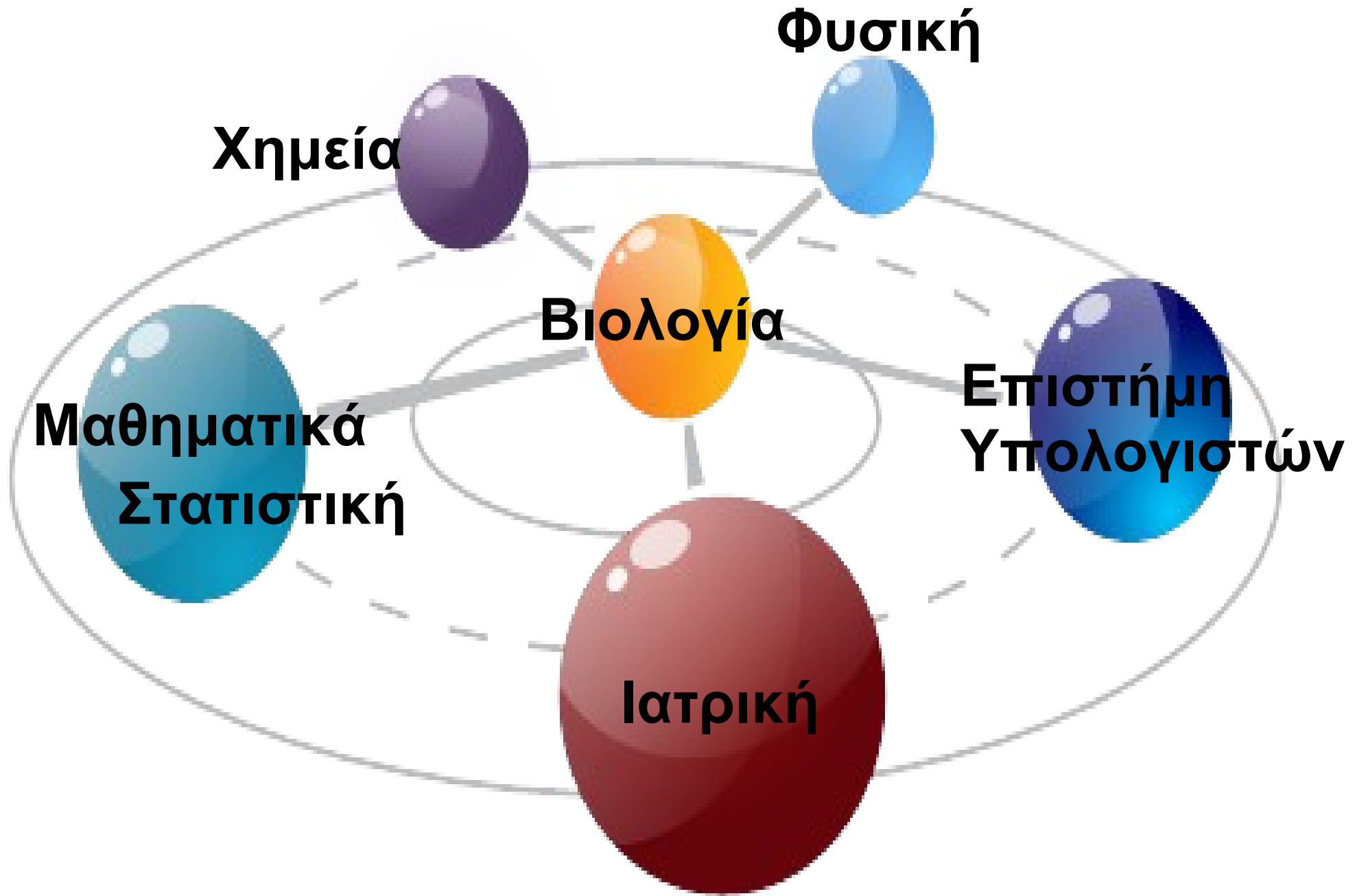
**Βιολόγος MSc, PhD**

# Τι είναι η Βιοπληροφορική



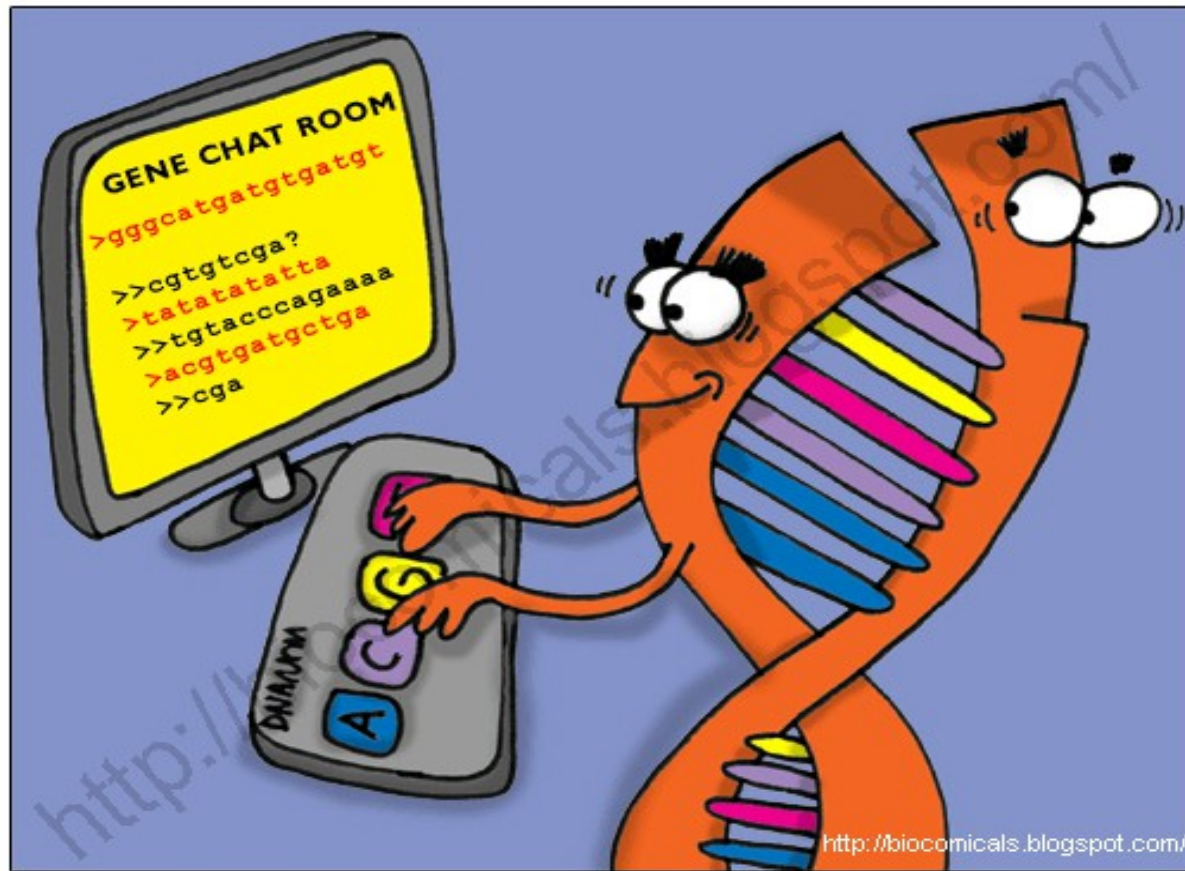
**Η Βιοπληροφορική απαντά σε βιολογικά ερωτήματα μέσα από τη χρήση μεθόδων & εργαλείων της επιστήμης των υπολογιστών**

# Βιοπληροφορική και άλλες Επιστήμες



# Τι κάνει η Βιοπληροφορική

Εφαρμογή μεθόδων και τεχνικών που αναπτύσσονται από την επιστήμη των υπολογιστών με τη συνδρομή των μαθηματικών, της φυσικής και της χημείας στη διαχείριση πληροφοριών που εμπεριέχονται στα βιομόρια με στόχο την κατανόηση και οργάνωση των πληροφοριών ώστε να απαντηθούν βιολογικά ερωτήματα.



# Συνήθη Ζητούμενα της Βιοπληροφορικής

---

- Κατανόηση της οργάνωσης και της λειτουργίας των ζωντανών οργανισμών
- Αναγνώριση γενετικών ανωμαλιών
- Μελέτη των μηχανισμών γενετικών και μη ασθενειών
- Ανακάλυψη και σχεδιασμός φαρμάκων
- Γενετική τροποποίηση κυττάρων και οργανισμών
- Γονιδιακή θεραπεία



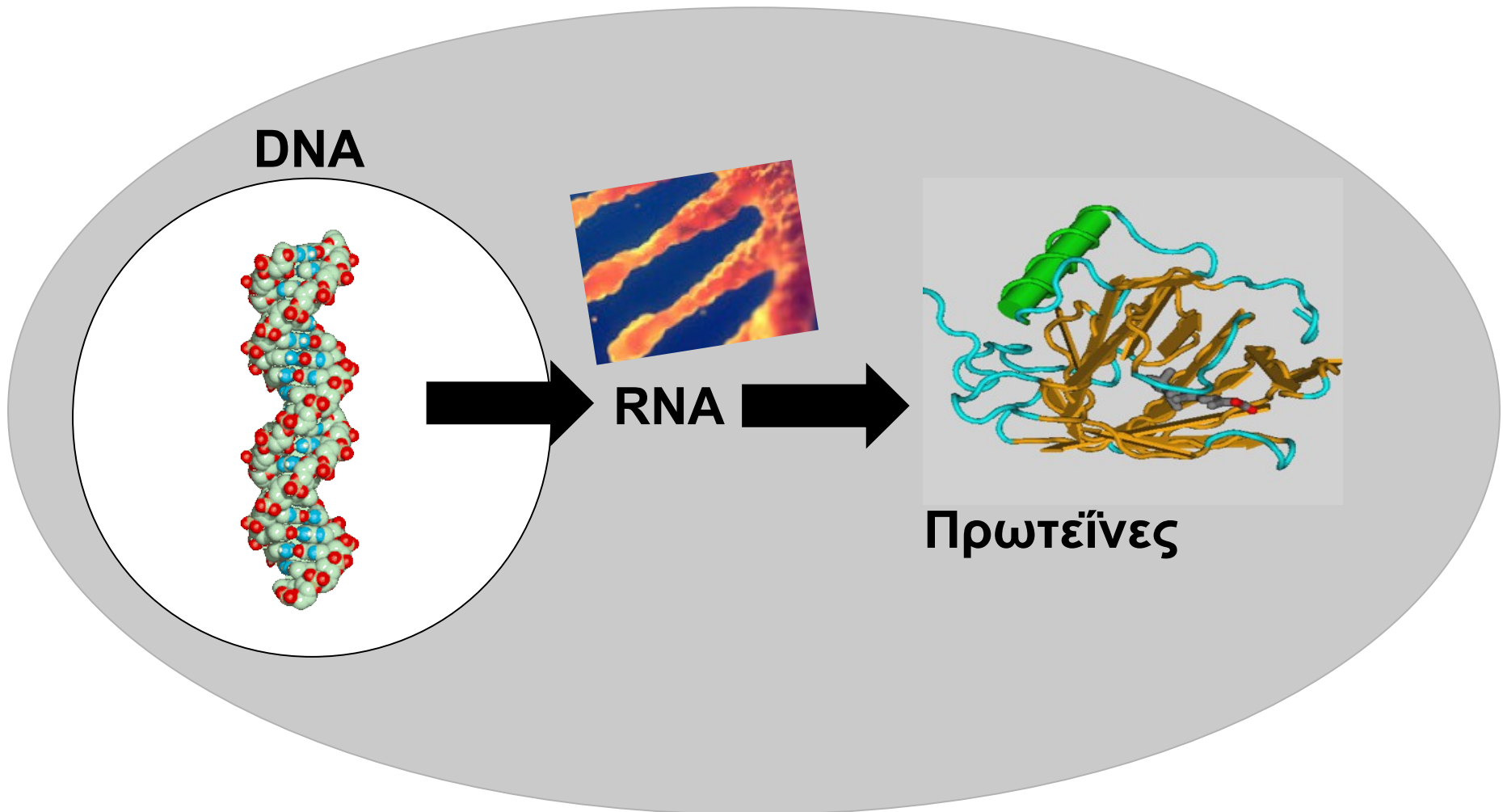
# Τι κάνουμε συνήθως στη Βιοπληροφορική

---

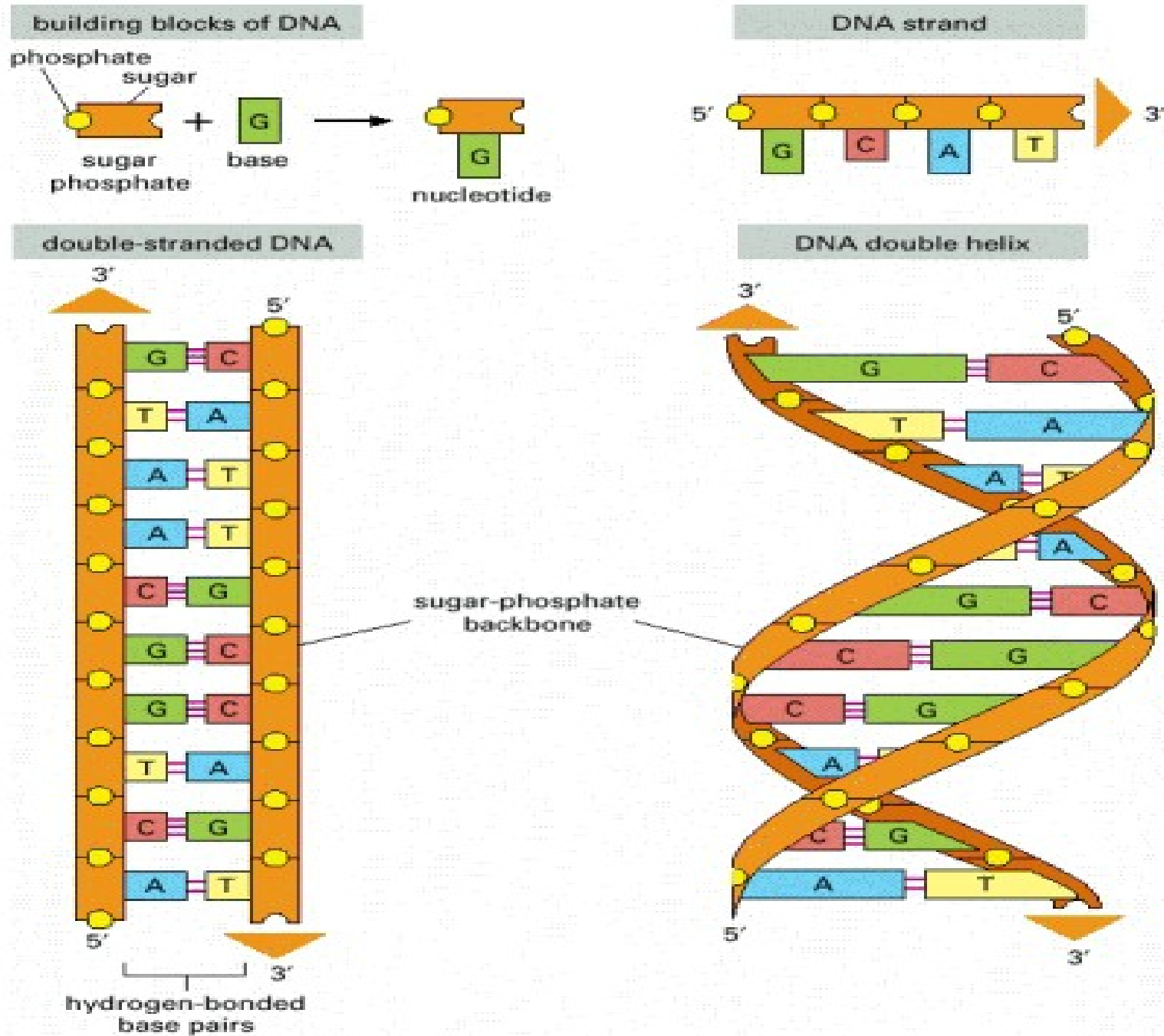
- Ανεύρεση και εξαγωγή αλληλουχιών DNA, RNA ή πρωτεϊνών από βάσεις δεδομένων.
- Σχολιασμός της δομής και λειτουργίας τους με πληροφορίες από πειράματα που έχουν καταχωρηθεί στις βάσεις.
- Διερεύνηση των συνθηκών κάτω από τις οποίες συντίθενται αυτά τα μόρια στα κύτταρα και στους οργανισμούς.
- Μελέτη των μεταξύ τους αλληλεπιδράσεων.
- Συσχέτιση με τον φαινότυπο (εμφάνιση και λειτουργία) των κυττάρων και των οργανισμών.

# Ας γνωρίσουμε όμως πρώτα τα Βιομόρια

## Συνήθεις ύποπτοι

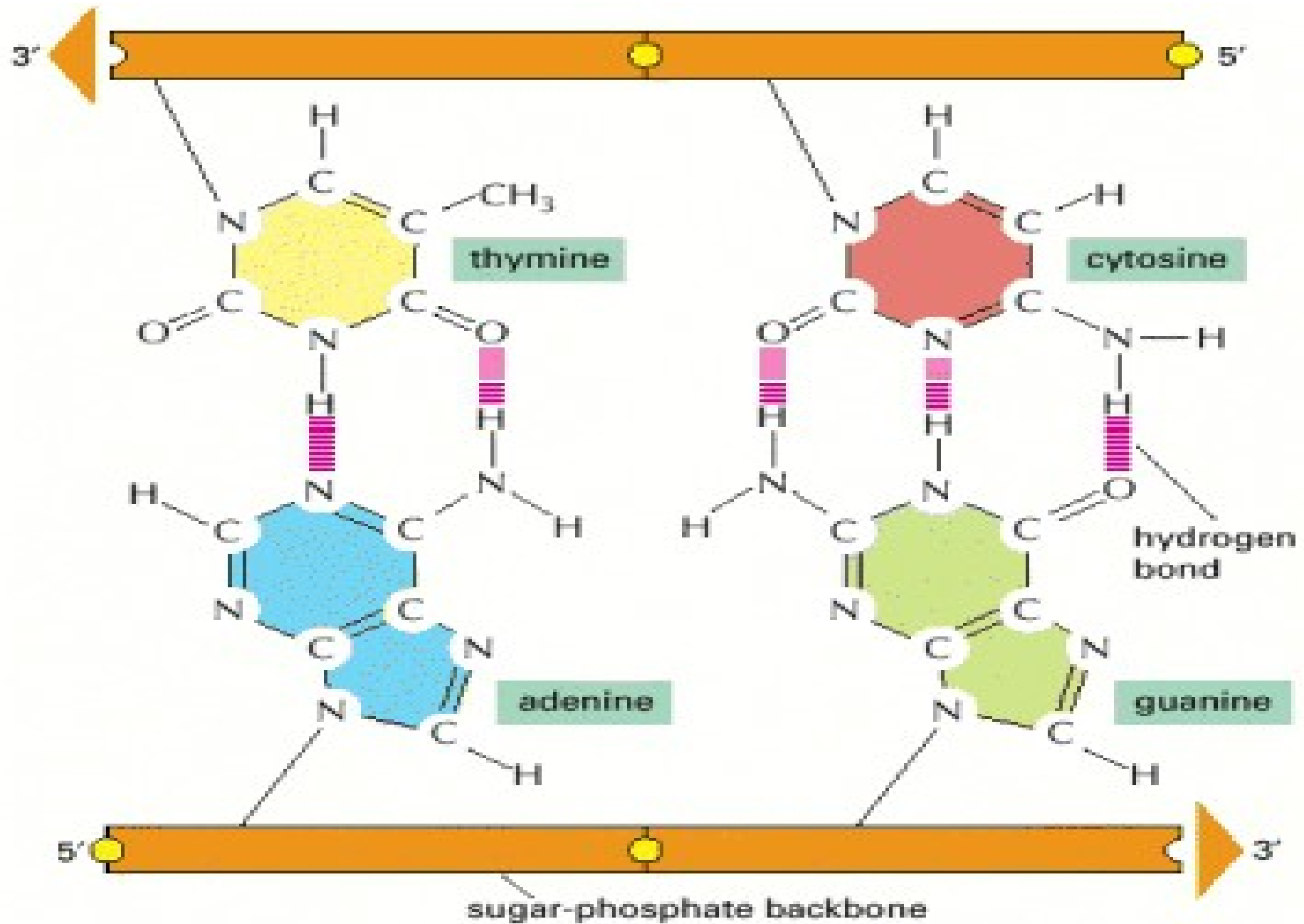


# DNA

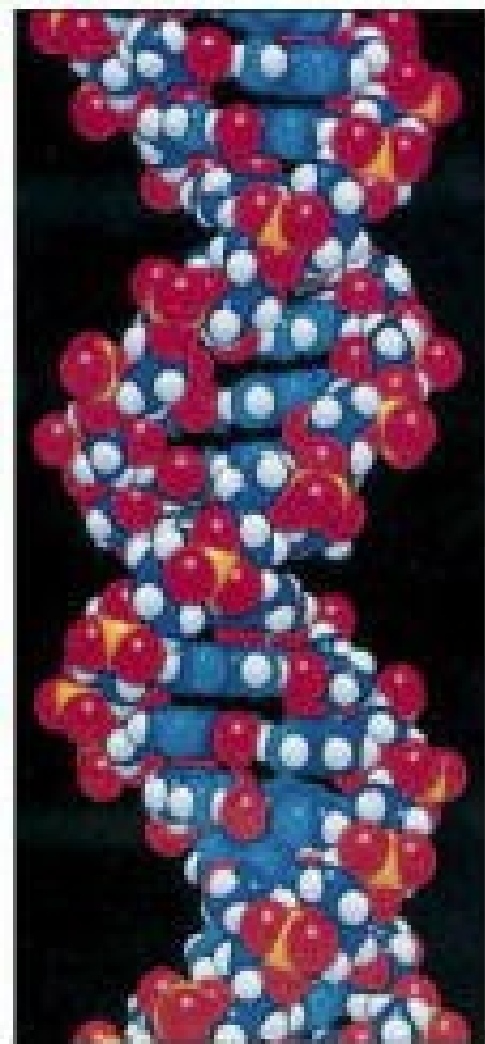




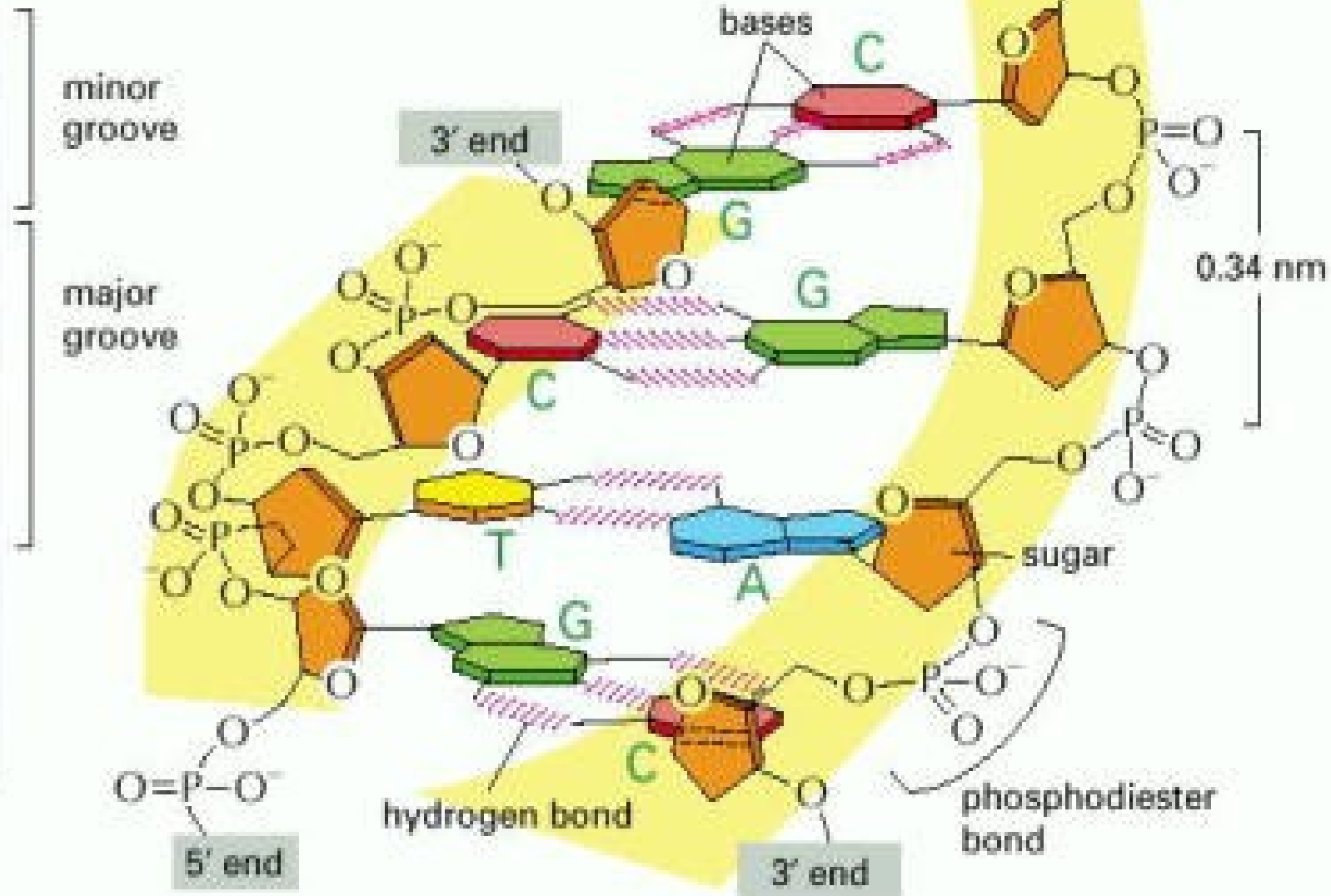
# DNA



# DNA



2 nm

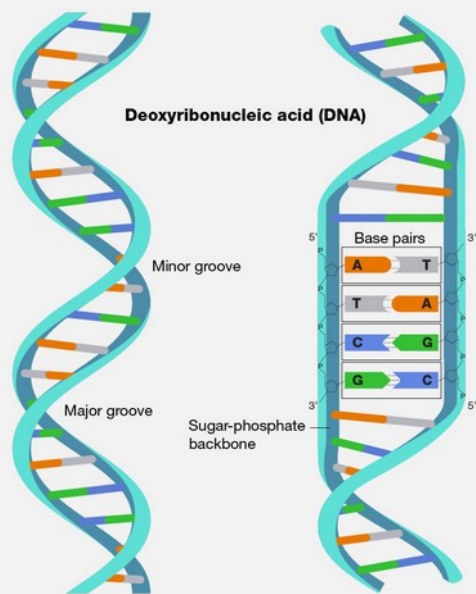


(B)

(A)

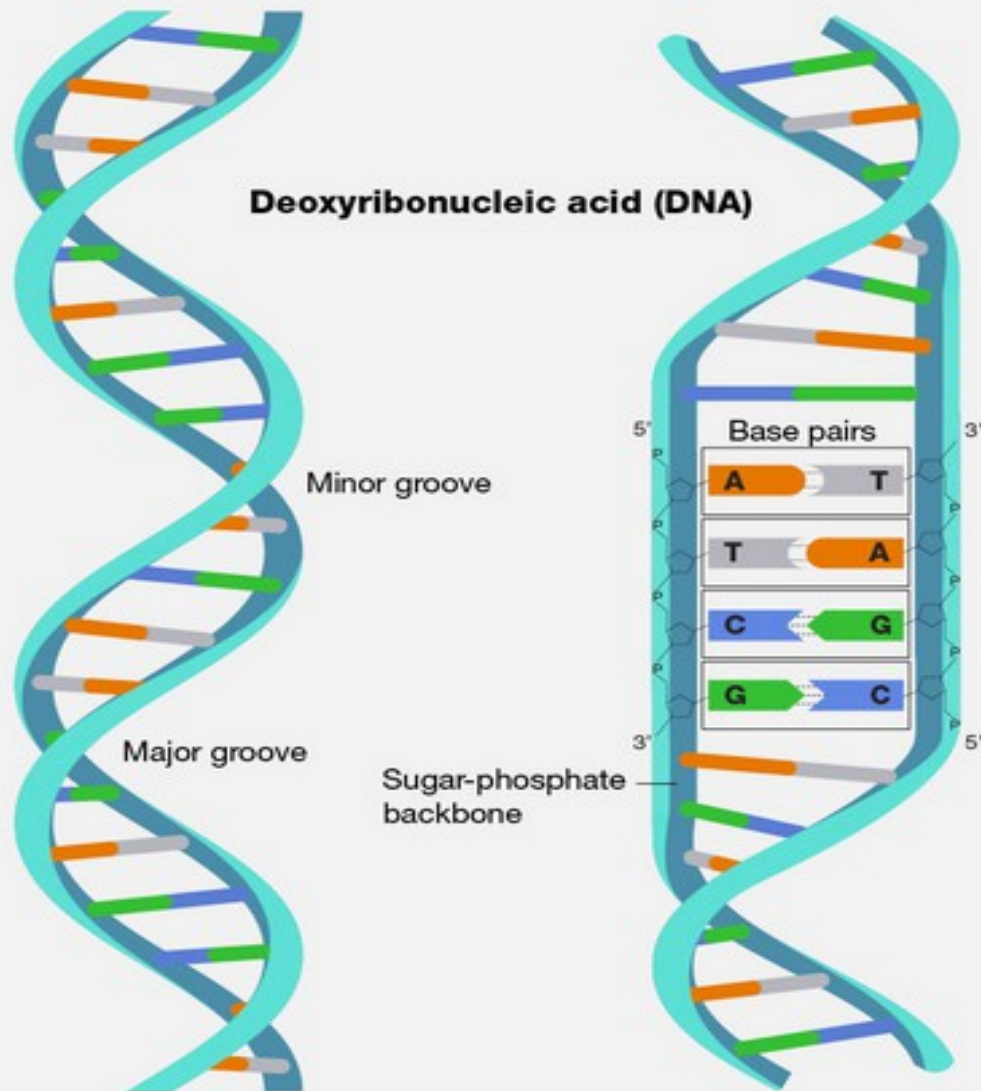
# Ας γνωρίσουμε όμως πρώτα τα Βιομόρια

## Τι είναι η αλληλουχία DNA & RNA



- Απομόνωση του DNA από τα κύτταρα ενός οργανισμού.
- Προσδιορισμός της συγκεκριμένης σειράς όπου οι βάσεις των νουκλεοτιδίων διατάσσονται στις αλυσίδες της διπλοέλικας του DNA.
- Συμπληρωματικότητα των βάσεων: Απέναντι από μια αδενίνη (A) βρίσκεται μια θυμίνη (T) και απέναντι από μια κυτοσίνη (C) βρίσκεται πάντα μια γουανίνη (G).
- Χρησιμοποιώντας την αρχή της συμπληρωματικότητας των βάσεων (A-T, C-G) προσδιορίζουμε τη σειρά με την οποία διατάσσονται τα νουκλεοτίδια στην μία έλικα του DNA από τη γνώση της άλλης.
- Με τη βοήθεια του ίδιου κανόνα προσδιορίζουμε την αλληλουχία των νουκλεοτιδίων στο RNA που προκύπτει από το DNA.

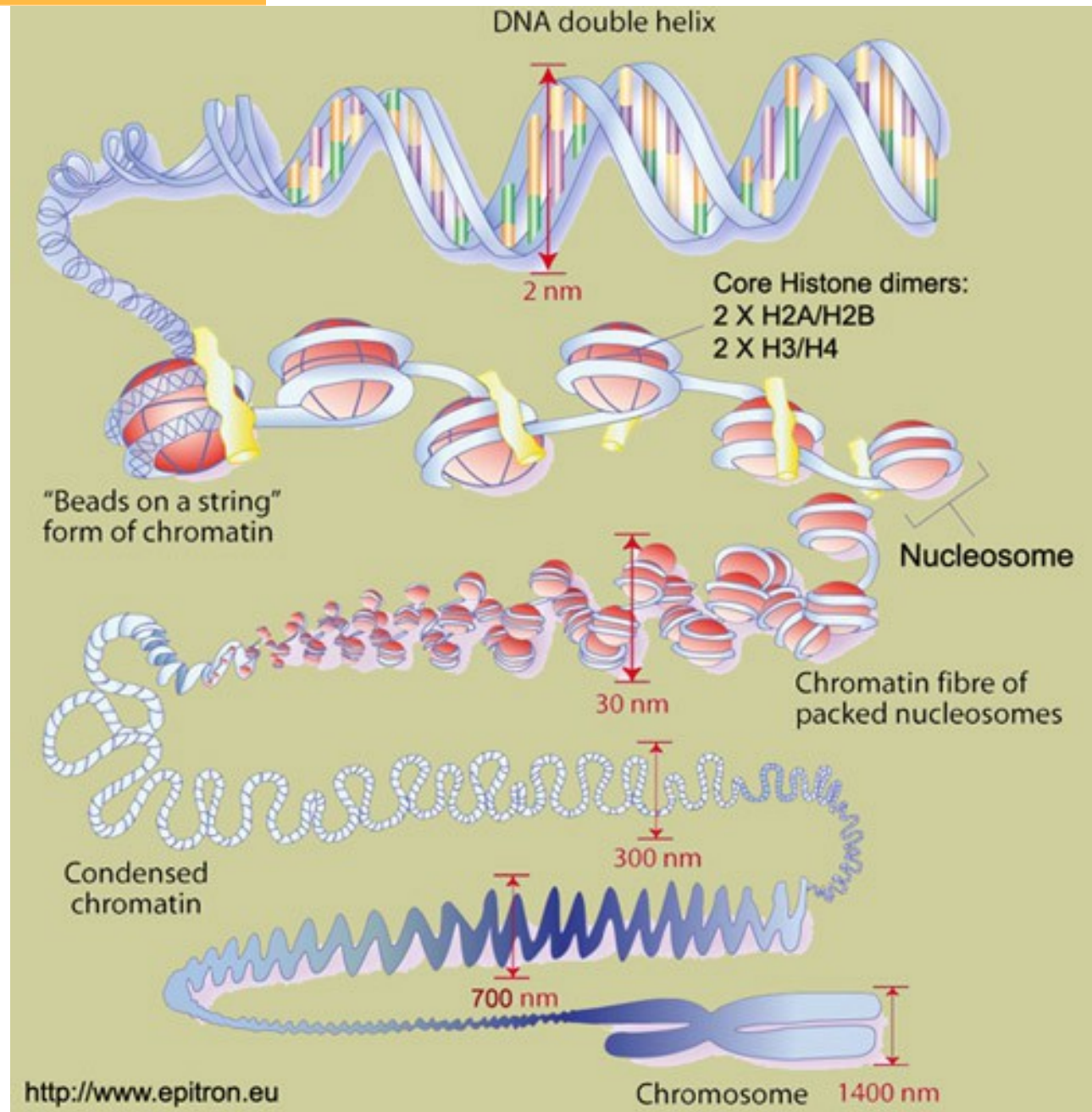
# Αλληλουχία DNA



**ATCG  
TAGC**

# DNA

Στο πυρήνα του  
διπλοειδούς  
ανθρώπινου  
κυττάρου,  
το DNA κατανέμεται  
σε  
46 γραμμικά μόρια  
συνολικού μήκους  
2m!





# Ας δούμε όμως πρώτα πως γίνεται η απομόνωση του DNA

---







# Απομόνωση DNA από φυτικά κύτταρα

Υλικά:

Πάγος

Μπανάνες

Για το διάλυμα εκχύλισης του DNA (100 mL )

Υγρό πιάτων (όχι συμπυκνωμένο) 10 mL

Μαγειρικό αλάτι  $\frac{1}{4}$  κουταλιού

Απιονισμένο νερό 90 mL

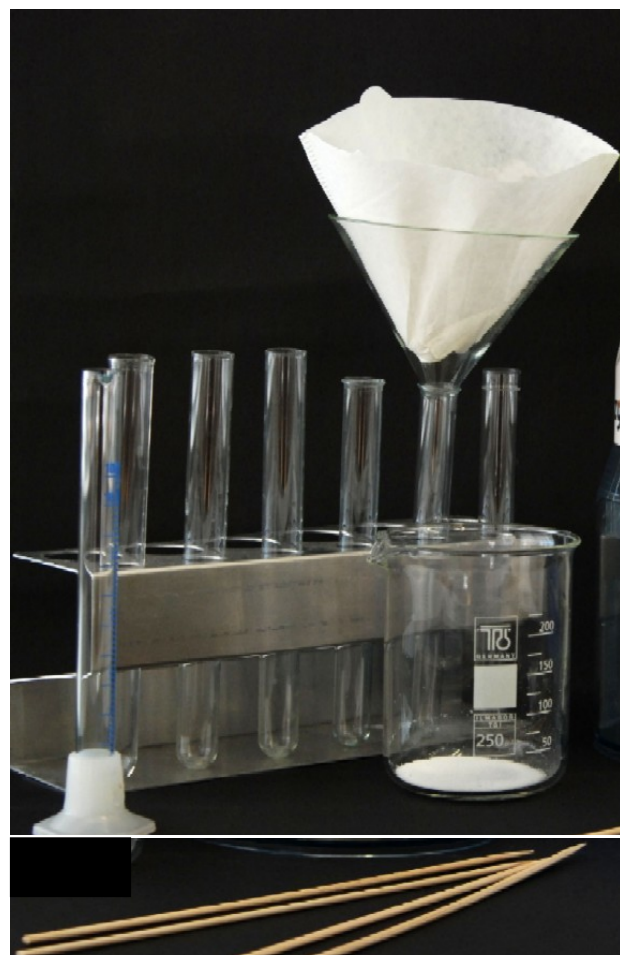
Για την απομόνωση του DNA

Αιθανόλη 6 ml παγωμένη (-20 °C)

Σταγόνες διαλύματος πεψίνης 1%w/v

Όργανα κ Συσκευές:

- Ηθμός ή γάζα
- Σακουλάκι ή Γουδί
- Χωνί
- Δοκιμαστικός σωλήνας
- Ποτήρι ζέσεως
- Ξύλινος ή γυάλινος αναδευτήρας
- Πιπέτα Pasteur ή σύριγγα



# Εκτέλεση Πειράματος

## Βήμα 1

Πολτοποιούμε το υλικό μας μέσα στο σακουλάκι

## Βήμα 2

Προσθέτουμε περίπου 10mL διαλύματος εκχύλισης (ίση ποσότητα με το ομογενοποίημα) και πιέζουμε ελαφρά ώστε να αναμιχθούν.

## Βήμα 3

Τοποθετούμε ένα χωνί σε ένα δοκιμαστικό σωλήνα και βάζουμε μία γάζα/ηθμός μέσα στο χωνί.

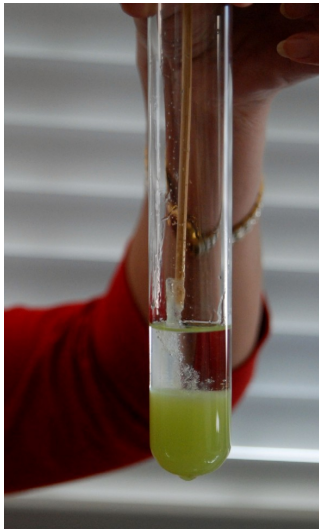
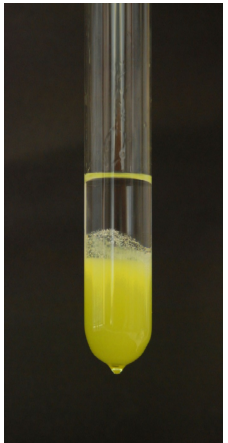
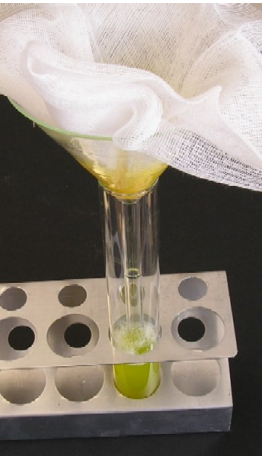
Ρίχνουμε το μίγμα του διαλύματος εκχύλισης με το πολτοποιημένο υλικό μας, στον ηθμό και φιλτράρουμε το μίγμα μέσα στο σωλήνα.

## Βήμα 4

Με τη βοήθεια της πιπέτας προσθέτουμε 2-3 σταγόνες διαλύματος του ενζύμου (πεψίνης), ανακατεύουμε ελαφρά και περιμένουμε λίγα λεπτά. Έπειτα προσθέτουμε σιγά σιγά την παγωμένη αλκοόλη στο διήθημα μέχρι να σχηματιστούν δύο διακριτές φάσεις. Κρατάμε το δοκιμαστικό σωλήνα ακίνητο και παρατηρούμε τις δύο φάσεις. Τα λιπίδια και οι πρωτεΐνες κινούνται στον πυθμένα της υδάτινης (κάτω) στοιβάδας, ενώ τα νουκλεϊκά οξέα ανέρχονται στη στοιβάδα της αλκοόλης (πάνω)

## Βήμα 5

Με τη βοήθεια της ράβδου “ψαρεύουμε” το DNA και την απομακρύνουμε τον δοκιμαστικό σωλήνα



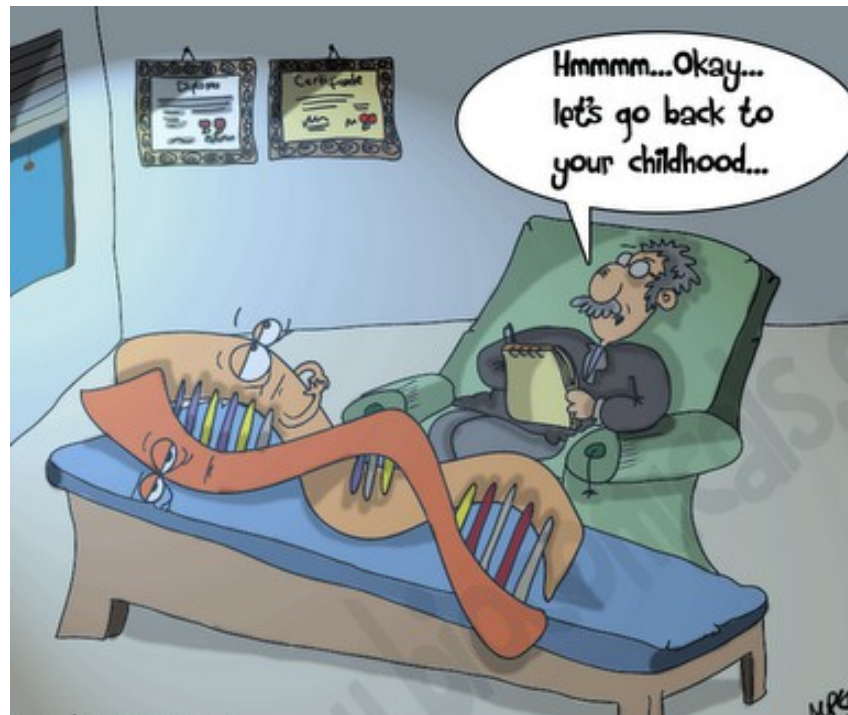
## Ερωτήσεις- Προβληματισμοί

- Το παραπάνω πείραμα θα μπορούσε να γίνει χρησιμοποιώντας κάποιο άλλο υλικό π.χ ζωικού ιστού ή ακόμη και ανθρώπινου;
- Γιατί πολτοποιούμε την μπανάνα ;
- Ποιος ο ρόλος του απορρυπαντικού και της πρωτεΐνάσης στο διάλυμα εκχύλισης;
- Γιατί ξαναπροσθέτουμε πρωτεΐνάση στο διήθημα;
- Γιατί το DNA συλλέγεται στην φάση της αλκοόλης;
- Το RNA συλλέγεται στην φάση της αλκοόλης;
- Αν παρατηρήσουμε το DNA στο μικροσκόπιο θα δούμε την διπλοέλικα;

# Αλληλούχηση DNA

## Προσδιορισμός της νουκλεοτιδικής αλληλουχίας (sequencing)

Ανάλυση κατά Sanger (1997): PCR με ραδιοσημασμένα ddNTPs, ηλεκτροφόρηση σε πήκτωμα ακρυλαμίδης & αυτοραδιογραφία

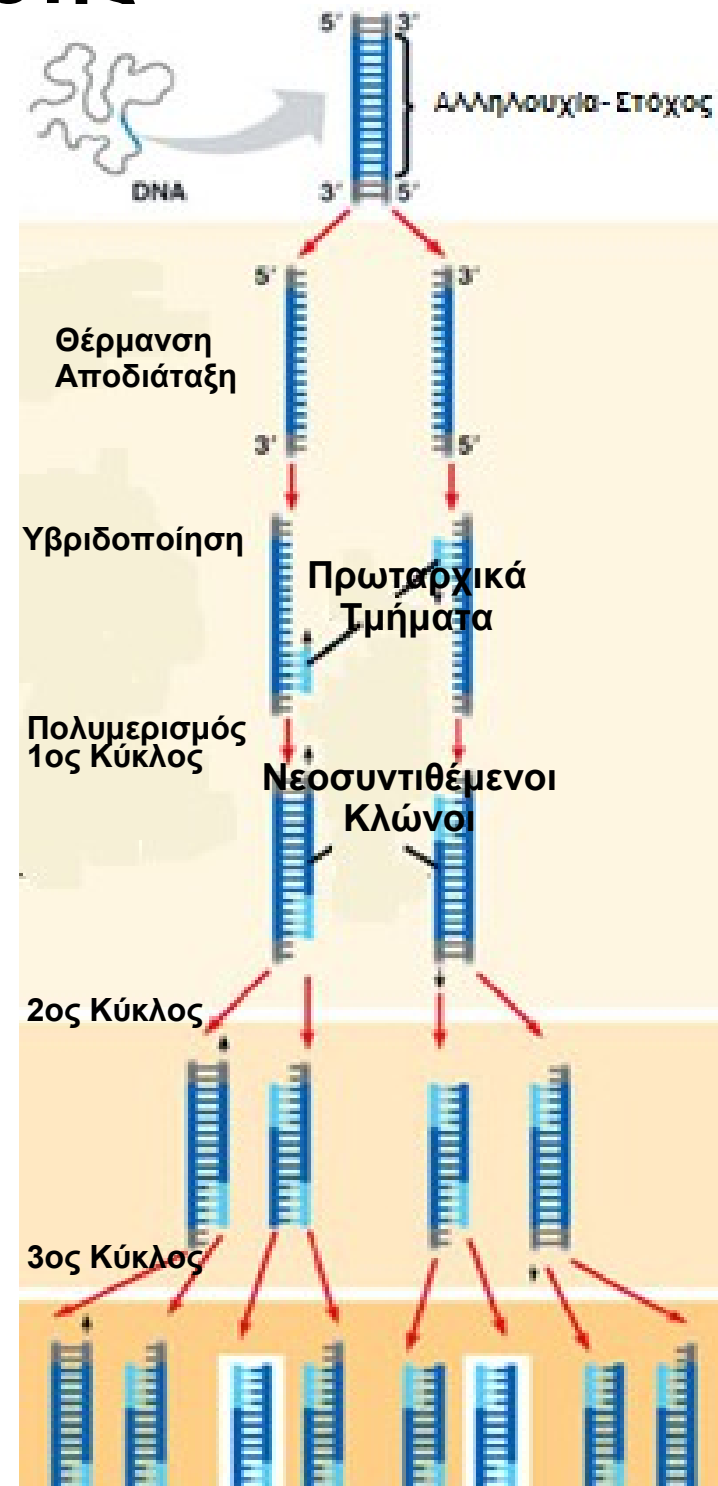
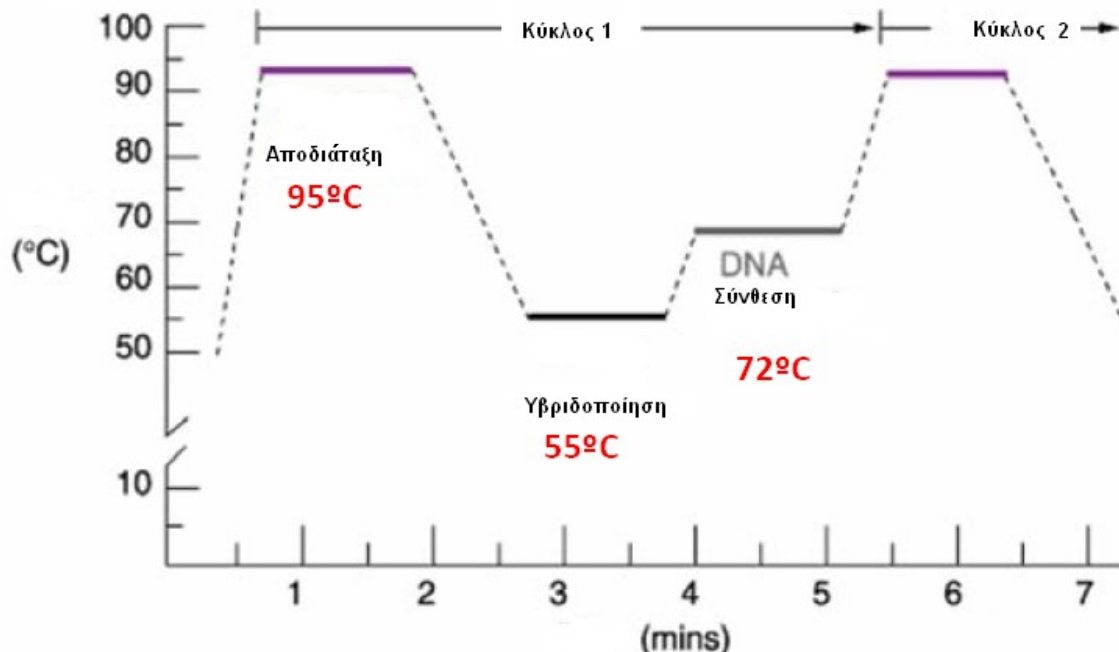


ggaattactgggctgtaaagcgcacgcaggcggttgcccaagtcagatgtgaaagccccgggcttaacctgggaa  
ctgcattgaaactgggctgactagagtatgaaagagaaagcgggaatttcagtgtagcagtgaaa tgcgtagata  
ttggaaggaa caccgatggc gaaggcagct ttctgggtcg ata



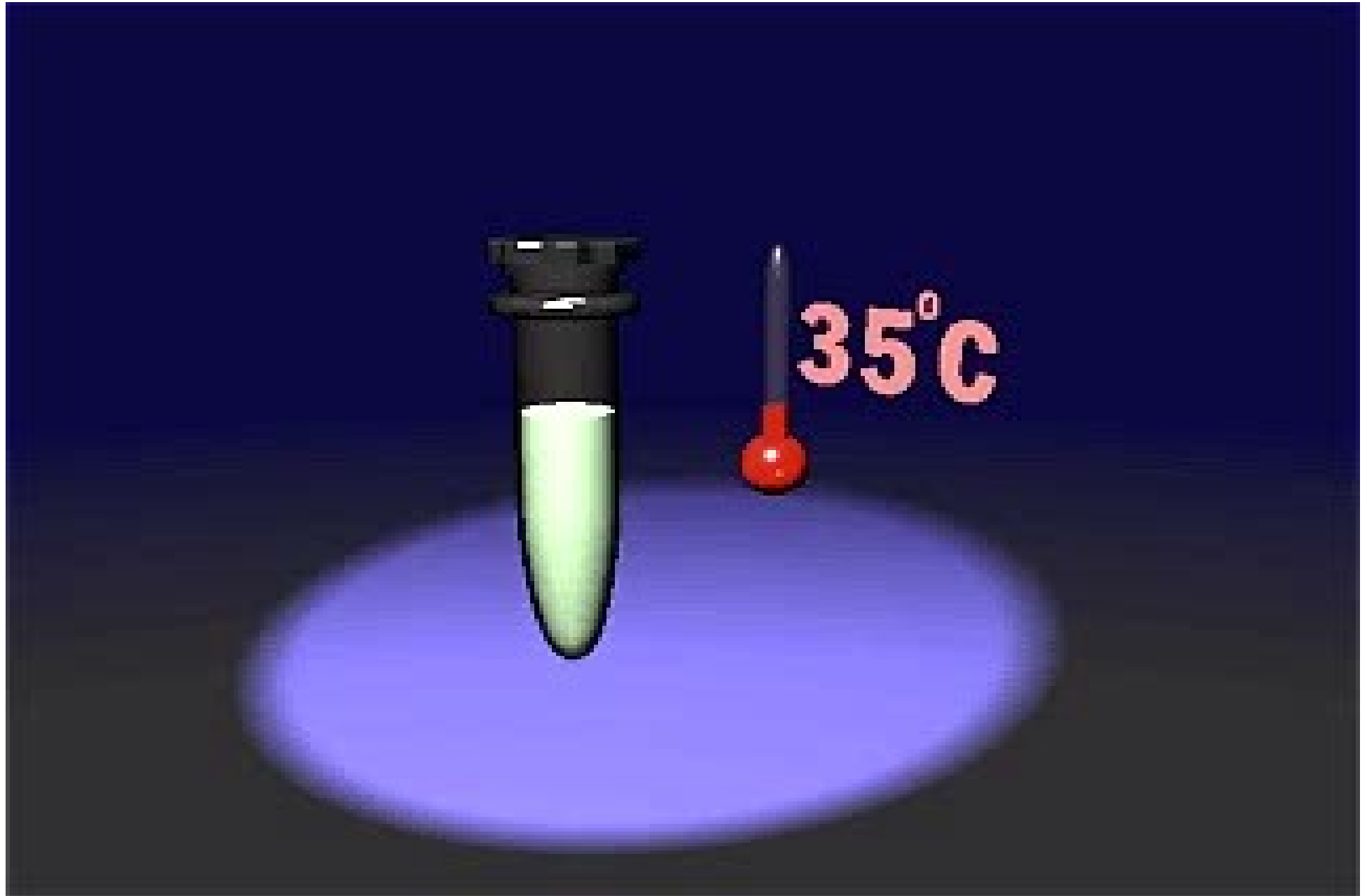
# Αλυσιδωτή Αντίδραση Πολυμεράσης PCR-Polymerase Chain Reaction

- Κλωνοποίηση επίλεκτων αλληλουχιών *in vitro* για ανίχνευση & χαρακτηρισμό
- Πρωταρχικά τμήματα (primers) 20-30 νουκλεοτιδίων συμπληρωματικά του DNA-στόχου
- Θερμοανθεκτική (Tac) πολυμεράση
- Μία τυπική PCR κάνει 30 κύκλους των 60-90s
- Μεγάλη ευαισθησία
- Πολλές εφαρμογές



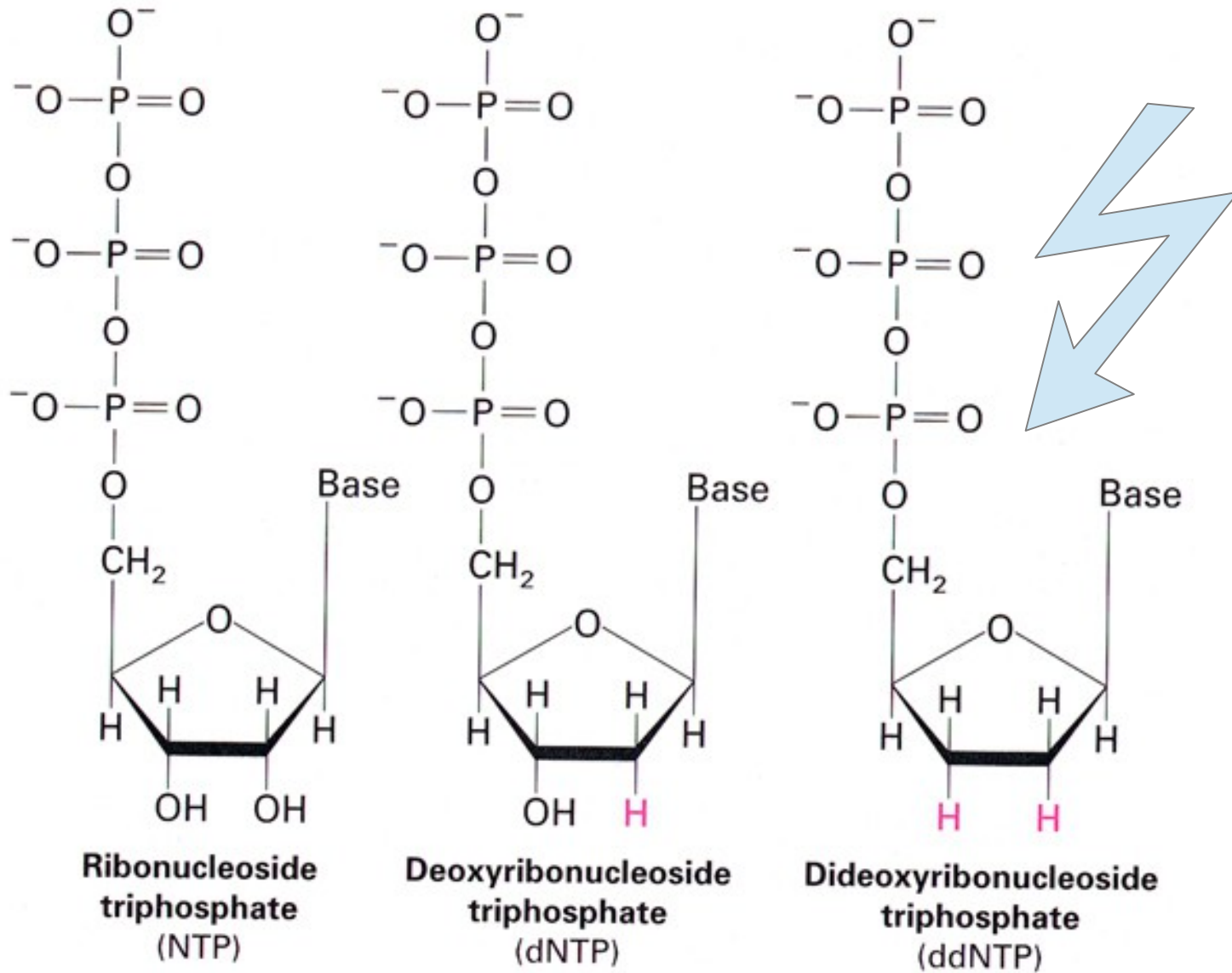
# Αλυσιδωτή Αντίδραση Πολυμεράσης PCR

---



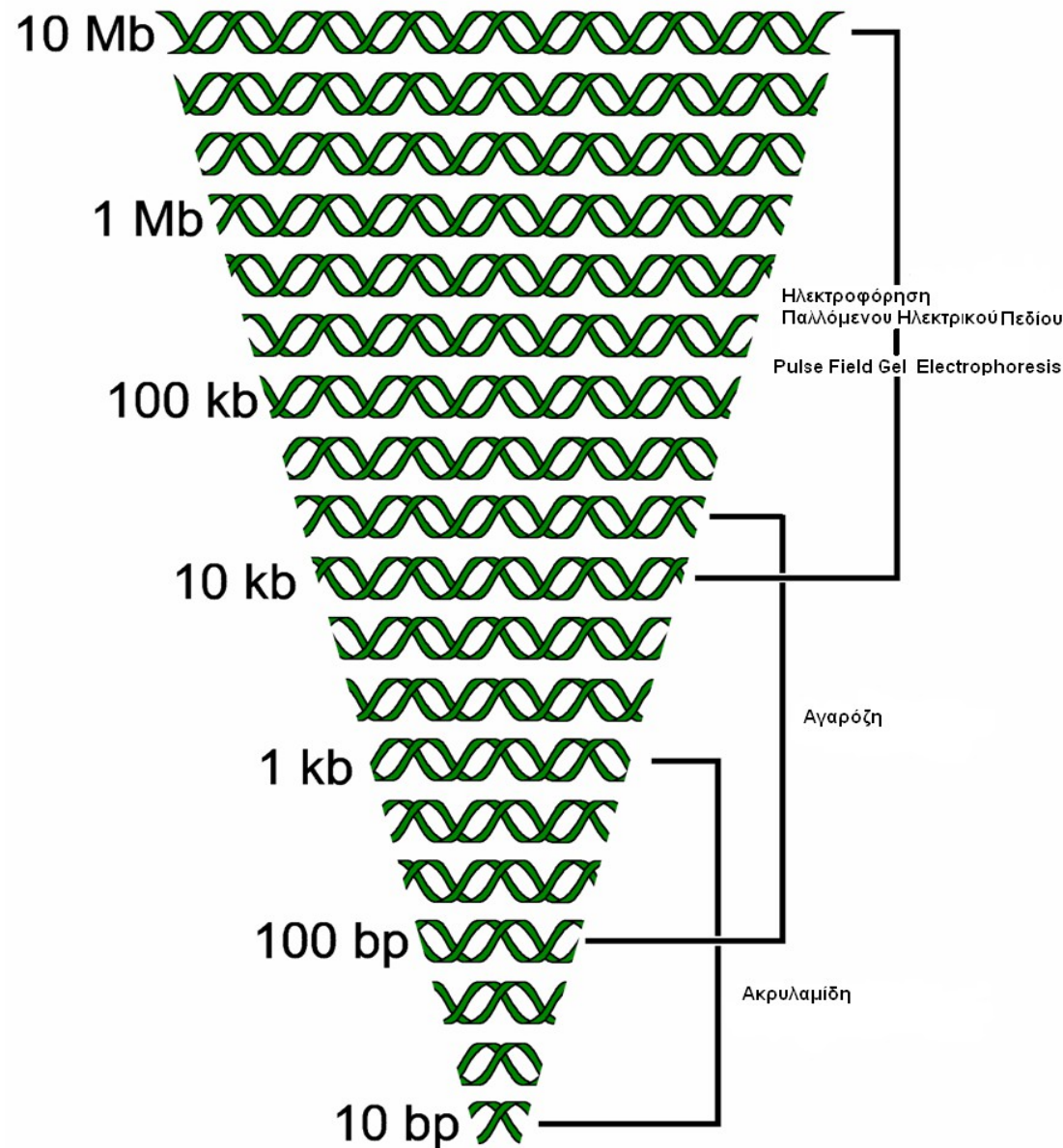


# NTP, dNTPs & ddNTPs

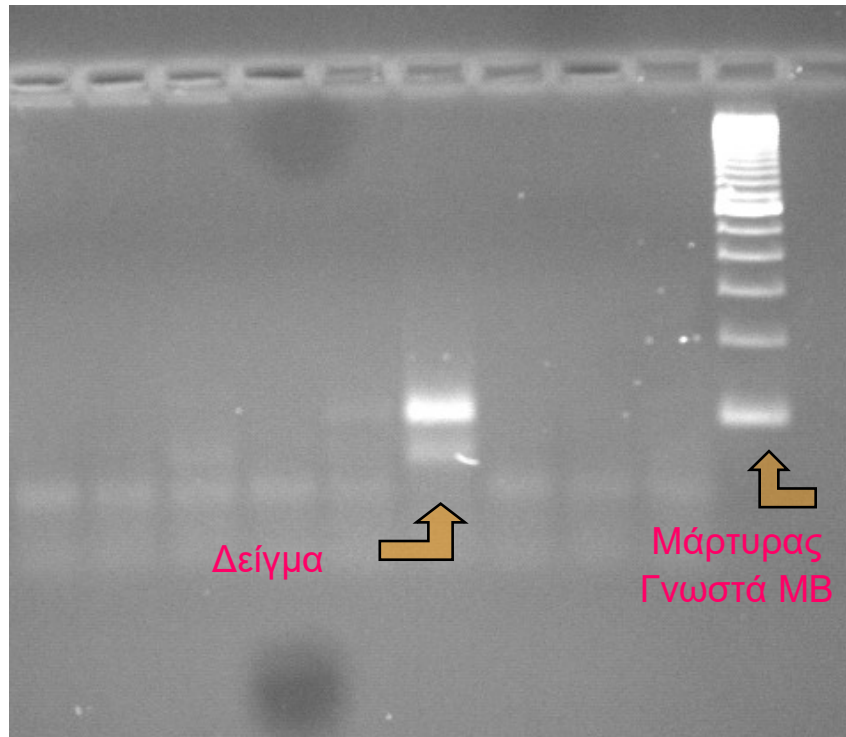


# Ηλεκτροφόρηση σε Πήκτωμα

Μέθοδος διαχωρισμού βιομορίων με βάση το μέγεθός τους



# Πήκτωμα Αγαρόζης με EtBr

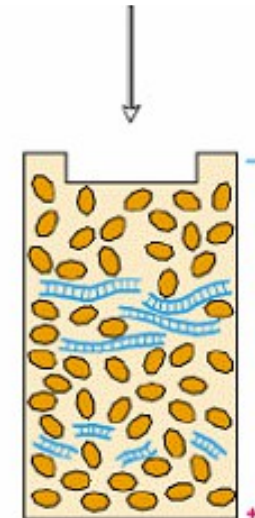
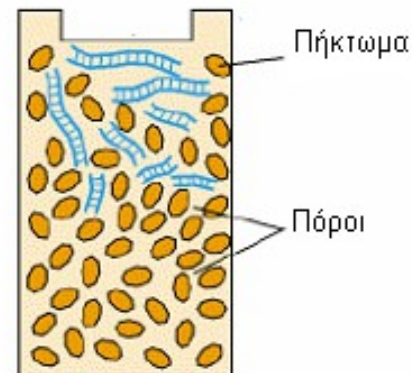


← “πηγαδάκια”  
1000 bp

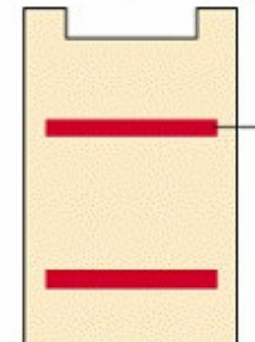
700 bp

600 bp

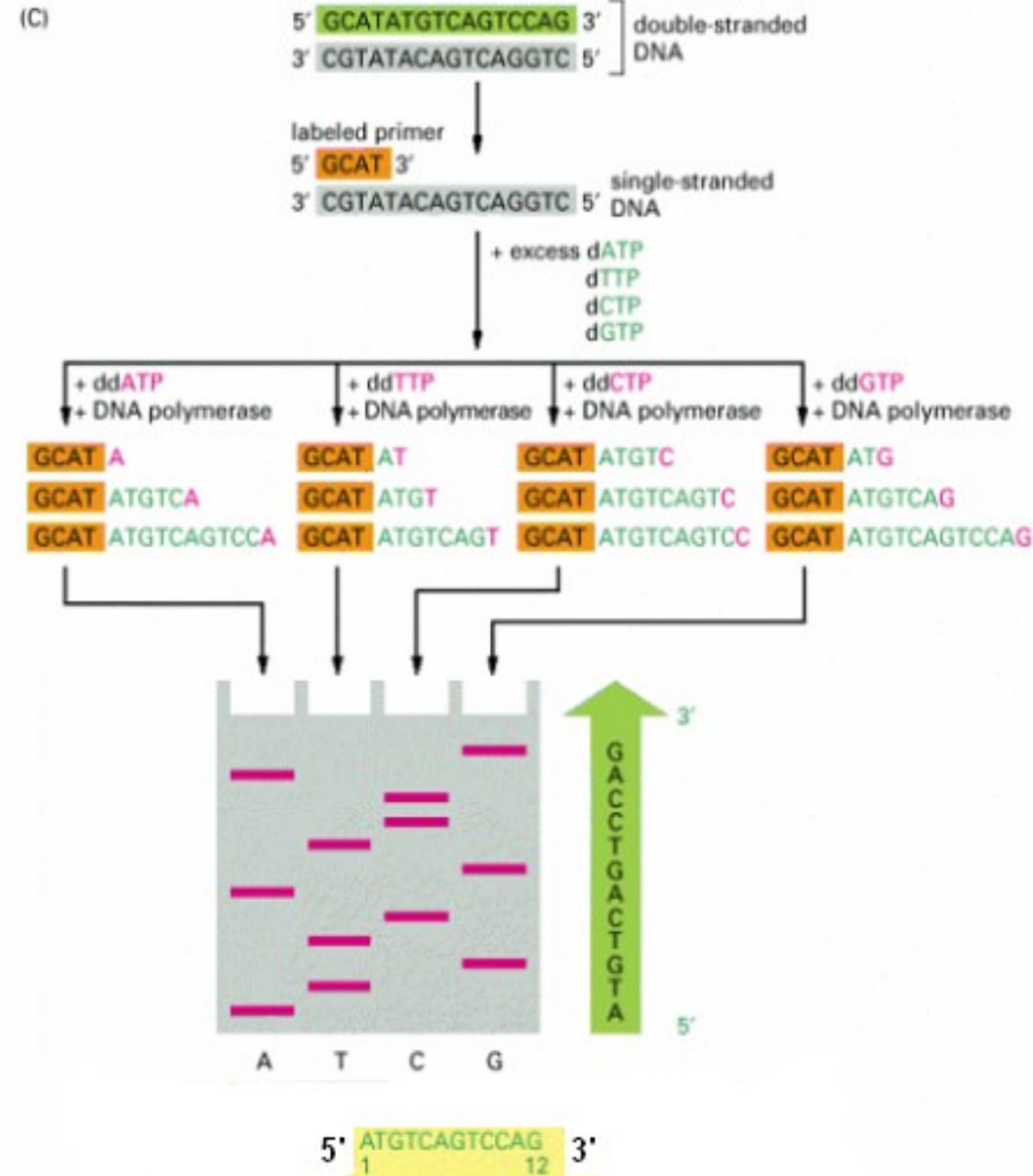
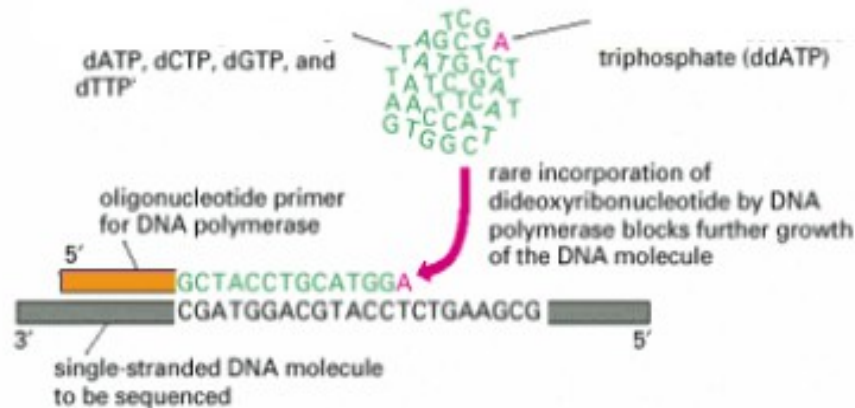
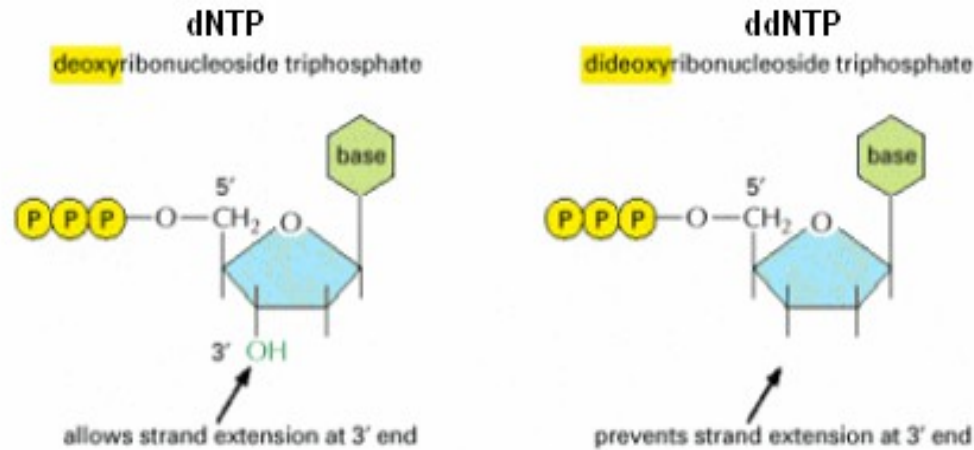
500 bp



Φθορίζουσα Χρωστική  
ή Υβριδισμός



# Αλληλούχιση DNA κατά Sanger

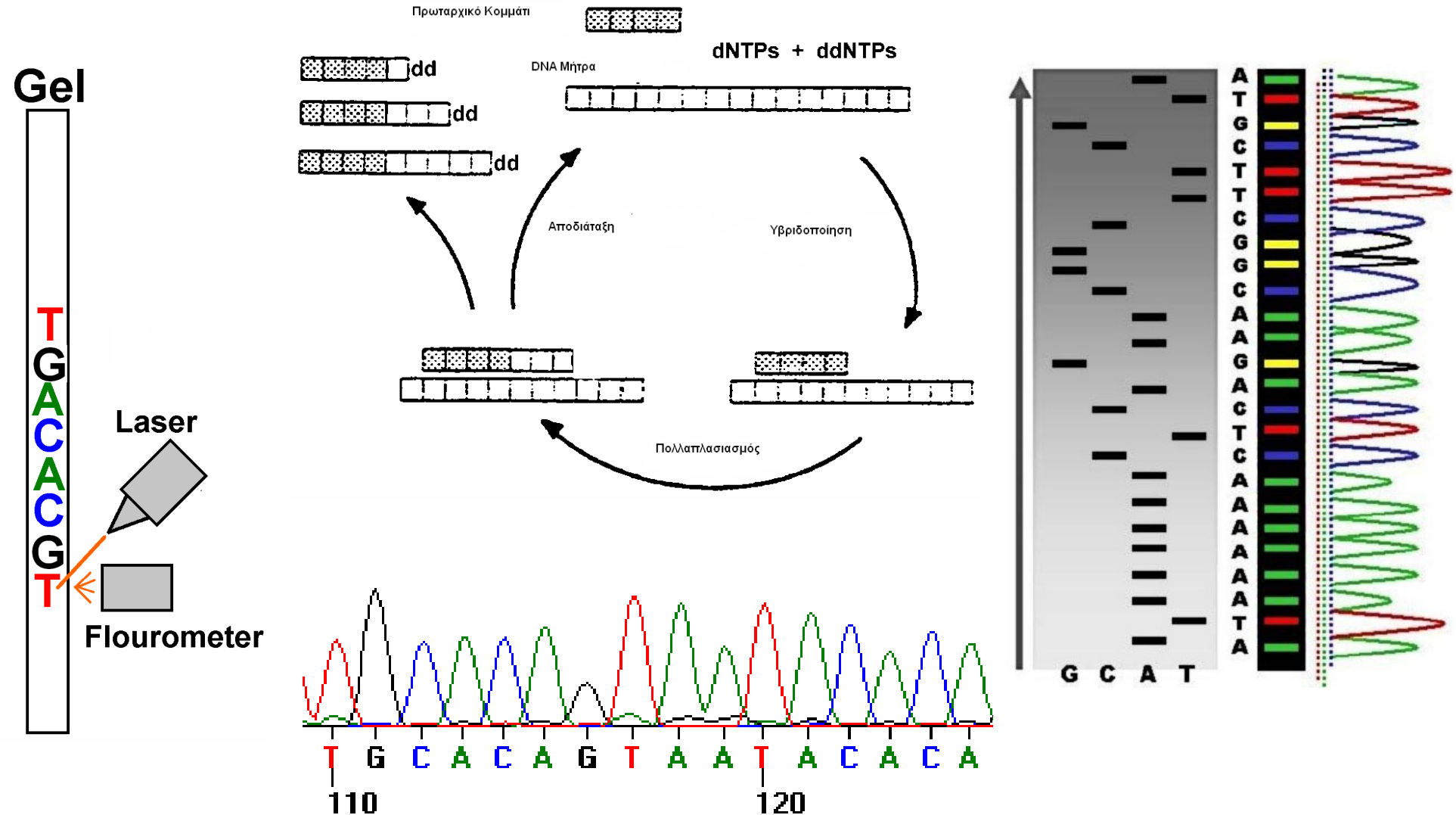


(Alberts *et al*, 2002)



# Αυτόματη Αλληλούχιση DNA με Φθορίζουσες Χρωστικές

Κάθε ddNTP είναι συνδεδεμένο με διαφορετική φθορίζουσα χρωστική





# Προσδιορισμός Αλληλουχίας: Αλληλούχιση DNA

---



# Γονιδίωμα

---

- Κάθε κύτταρο φέρει τις πληροφορίες που χρειάζεται το ίδιο και κατ' επέκταση και ο οργανισμός στον οποίο ανήκει κωδικοποιημένες σε μόρια DNA.
- Στα ευκαρυωτικά κύτταρα το DNA φυλάσσεται στον πυρήνα του κυττάρου και αυτό το DNA αποτελεί το γονιδίωμά του.
- Η ποσότητα του DNA που περιέχεται στα κύτταρα ενός οργανισμού είναι σταθερή και χαρακτηριστική του είδους του.
- Στον πυρήνα όλων των σωματικών κυττάρων του ανθρώπου υπάρχουν  $6 \cdot 10^9$  ζεύγη νουκλεοτιδίων που αν τα ξεδιπλώσουμε εκτείνονται σε μήκος περίπου 2m.
- Αυτό το DNA μπορούμε να το παρατηρήσουμε στο οπτικό μικροσκόπιο μόνο στη φάση όπου διαιρείται το κύτταρο, όπου και συμπυκνώνεται και παίρνει τη μορφή χρωμοσωμάτων. Σε ένα φυσιολογικό ανθρώπινο κύτταρο παρατηρούμε 46 χρωμοσώματα.

# Προσδιορισμός της Αλληλουχίας του Ανθρώπινου Γονιδιώματος The Human Genome Project

---

- Παγκόσμια προσπάθεια που άρχισε το 1990 και ολοκληρώθηκε το 2003.
- Το πρόγραμμα του ανθρώπινου γονιδιώματος διεξαγόταν παράλληλα από ένα δημόσιο (IHGSC) και ένα ιδιωτικό φορέα (Celera Genomics).
- Στόχος η αναγνώριση όλων των γονιδίων του ανθρώπου και η χαρτογράφησή τους πάνω στα 46 χρωμοσώματα.
- Στόχος επίσης ο προσδιορισμός των αλληλουχιών των παραπάνω γονιδίων.
- Στόχοι ήταν η κατασκευή γονιδιακών και χρωμοσωμικών χαρτών του ανθρώπινου γονιδιώματος.
- Παράλληλος στόχος ήταν η μελέτη των γονιδιωμάτων άλλων οργανισμών- μοντέλων.
- Κατάθεση των πληροφοριών σε βάσεις δεδομένων έτσι ώστε να είναι εύκολα προσβάσιμες σε όλους και να είναι δυνατή η ανάλυσή τους.

# Προσδιορισμός της Αλληλουχίας του Ανθρώπινου Γονιδιώματος The Human Genome Project



# Υπερβολική Αύξηση Νουκλεοτιδικών Αλληλουχιών

---

## Αυτοματοποίηση του DNA Sequencing



- Με τη μέθοδο Sanger και τη χρήση των αυτοματοποιημένων μεθόδων που εισήγαγε η Celera Genomics αποκρυπτογραφήθηκε το Human Genome (κ όχι μόνον).
- Αρχικά η ολοκλήρωση του Human Genome Project είχε προσδιοριστεί στο 2005
- Η ακολουθία ολοκληρώθηκε την άνοιξη του 2003
- Υψηλό κόστος Sanger sequencing
- Next-Generation-Sequencing (NGS): Αυτοματοποίηση της διαδικασίας → Επίσπευση ( $10^6$  nucleotides/ cycle) & Χαμηλότερο κόστος





# Γιατί Βιοπληροφορική?

- Παλιά

Βρες τη λέξη   C  A  T  



- Σήμερα

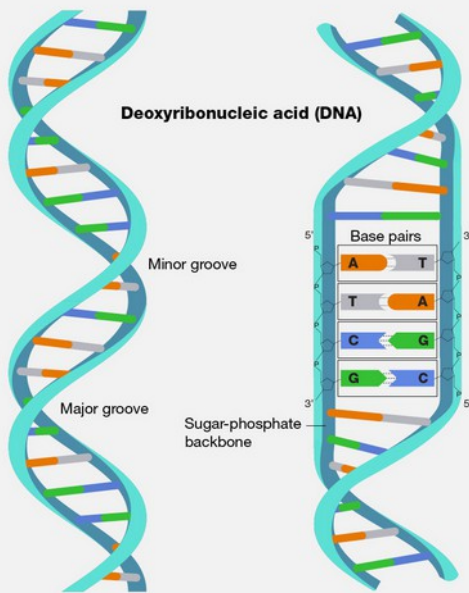
Βρείτε τις λέξεις

ACTTTATATTTTATTTTGGAGCTTGATCTAGAATAATTGGAACCTT  
CTTTAAGAAATATTAATTCGAATTTAGGTCATCCAGGTTCCCTT  
AATTGGAAATGACCAAATTTTGTAAACAGCTCATGC  
ATTTATTATAATTTTTTTTAAATTAATTGGAGGA  
TTTGGAAATGATTAGTTTAATAAGGCACCAGATATAG  
CTTTCCCTCGAATGAATAATATAAGTTACTTCCTCCTGCT  
TTAATACTTTTATTAACAAGTAGAAGAAAGTGGAGCTGGA  
ACAGGATGAACAGTTTATCCTCTCATCTATTATTGCTCATG  
GAGGAGCATCTGTTGACTTAGCTTTTCTCTTCATTTAGCAGG  
AATTTCTTCTATTTTAGGAGCTGTAAATTTTATTACAACGTGAATT  
AATATACGATCTATTGGTATTATGATCGAATACCTTTATTTG  
TTTGATCAGTTGCTATTACAGCCTTATTACTTTTATTATCTTTACCA  
GTTTTAGCTGGAGCAATTACAATATTATTAACNGATCGAAATTTA  
AATACATCATTTTTTG

# Ας εξοικειωθούμε λίγο με την ορολογία

## Γονίδιο και Γονιδίωμα

Γονίδιο ονομάζεται η βασική μονάδα της κληρονομικότητας.



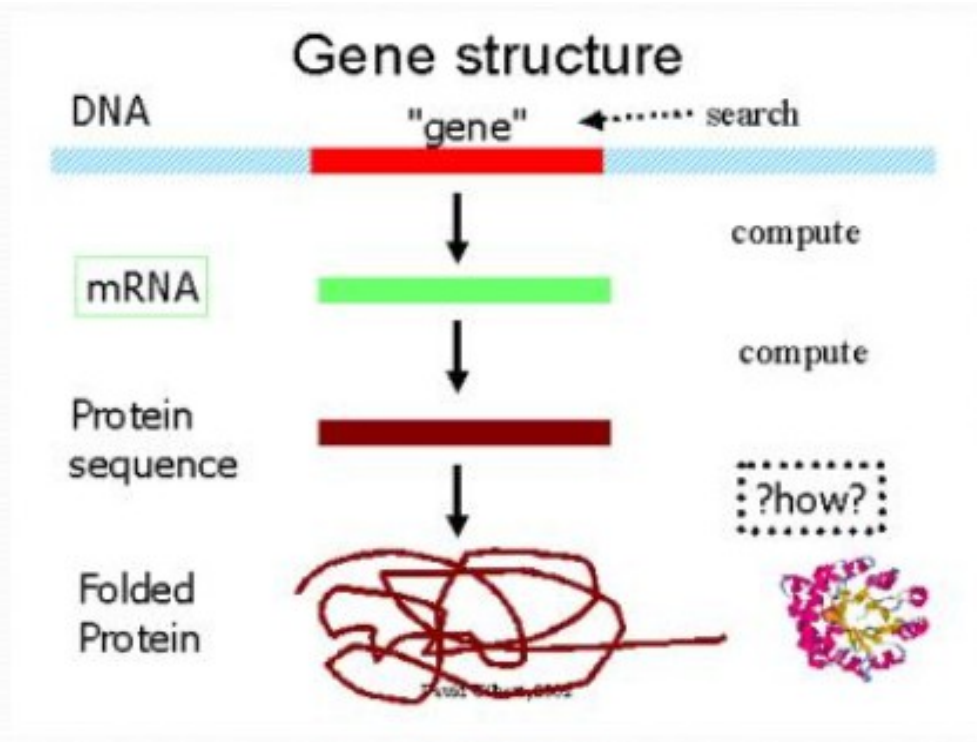
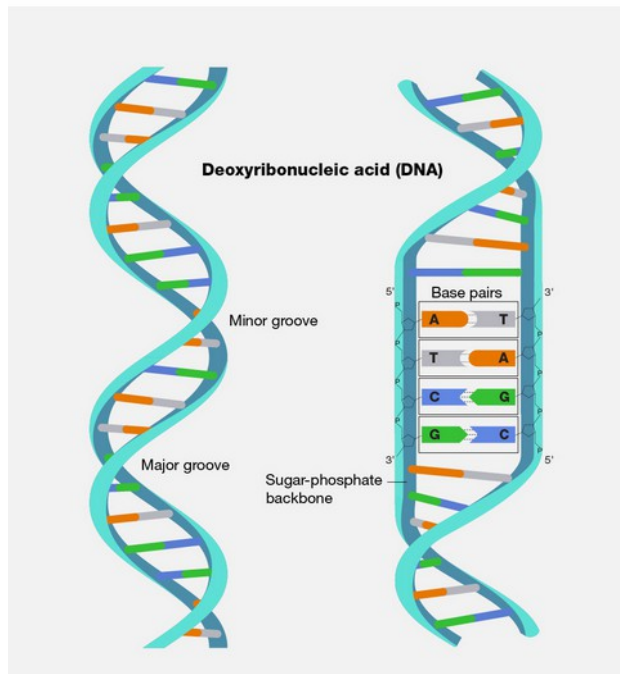
Αποτελεί τμήμα του γονιδιώματος με έκταση κατά μέσο όρο μερικές χιλιάδες νουκλεοτίδια.

Το γονιδίωμα δεν περιέχει μόνο γονίδια αλλά και αλληλουχίες που δεν αντιστοιχούν σε γονίδια.

Παλαιότερα γονίδιο ονομαζόταν ο τρόπος ελέγχου μιας ιδιότητας ενός οργανισμού.

Σύμφωνα με τη σύγχρονη μοριακή προσέγγιση γονίδιο είναι η αλληλουχία DNA που μετατρέπεται σε μόριο RNA, το οποίο, είτε καταλήγει στη σύνθεση μιας πρωτεΐνης, είτε ελέγχει κατά κάποιο τρόπο τη σύνθεση των πρωτεϊνών του

# Από το Γονίδιο στην πρωτεΐνη



Το Κεντρικό Δόγμα της Μοριακής Βιολογίας

$\text{DNA} \rightleftharpoons \text{RNA} \rightarrow \text{πρωτεΐνες}$  ή

$\text{νουκλεϊκά οξέα} \rightarrow \text{πρωτεΐνες}$

# Πρωτεΐνες

---

- Είναι και αυτές βιολογικά μακρομόρια που αποτελούνται από μικρότερα μόρια, τα αμινοξέα, που ενώνονται και σχηματίζουν μια πολυπεπτιδική αλυσίδα.
- Οι πρωτεΐνες ενός κυττάρου ή ενός οργανισμού έχουν ρόλο δομικό (π.χ. κερατίνη) ή λειτουργικό (π.χ. αιμοσφαιρίνη).
- Ο ρόλος μιας πρωτεΐνης είναι απόλυτα εξαρτημένος από τη διαμόρφωσή της στον χώρο.

# Πρωτεΐνες

Η τρισδιάστατη δομή μιας πρωτεΐνης καθορίζεται από τα ακόλουθα 4 επίπεδα οργάνωσης:

- πρωτοταγής δομή: η σειρά διάταξης των αμινοξέων στην αλυσίδα της πρωτεΐνης (συμβολοσειρά από 20 γράμματα- αμινοξέα)
- δευτεροταγής δομή: τοπική αναδίπλωση α- έλικας ή β-πτυχωτής επιφάνειας
- τριτοταγής δομή: η συνολική διαμόρφωση της αλυσίδας στο χώρο που καθορίζεται από τις επιμέρους τοπικές αναδιπλώσεις
- τεταρτοταγής δομή: η τελική διαμόρφωση στο χώρο για πρωτεΐνες που αποτελούνται από 2 ή περισσότερες αλυσίδες

## Primary Structure

VHLTPEEKSAVTALWGKVVNDE  
VGGEALGRLLVVPWTQRFES  
FGDLSTPDVAMGNPKVKAHGKK  
VLGAFSDGLAHLNLIKGTFTLS  
ELHCDKLHVDPENFRLLGNVLV  
CVLAHHFGKEFTPPVQAAVQKV  
VAGVANALAHKYH

amino acids (1hb - Haemoglobin B)

1hb (Chain: Tau-R David Gilbert, 2001)

19

## Secondary Structure



alpha helix



beta sheets

1hb (Chain: Tau-R David Gilbert, 2001)

20

## Tertiary Structure



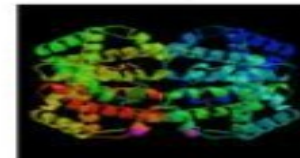
1hbAAS



1hbABO



## Quaternary Structure



# Από το Γονίδιο στην Πρωτεΐνη

Συνοπτικά λοιπόν έχουμε

Αλληλουχία  
νουκλεοτιδίων  
γονιδίου

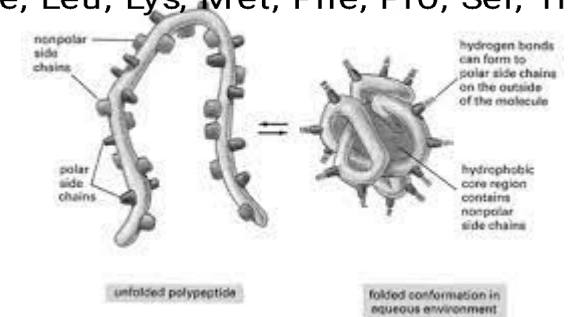
Αλληλουχία  
αμινοξέων  
πρωτεΐνης

Τρισδιάστατη  
διαμόρφωση  
πρωτεΐνης

Βιολογικός  
Ρόλος της  
πρωτεΐνης

ATTGCTGAAATGCCGATTTTGGCCAGAGAGAGAGATAGATAAAAGGGATT

Ala, Arg, Asp, Asn, Cys, Glu, Gln, Gly, Hsi, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thi





# Στόχοι της Μοριακής Βιολογίας

---

Αλληλούχιση και σύγκριση γονιδιωμάτων διαφορετικών οργανισμών-  
Εξελικτική πορεία

Αναγνώριση γονιδίων και πρωτεϊνών- Σύνδεση με λειτουργία κυττάρου

Διερεύνηση της γονιδιακής έκφρασης- Φαινότυποι κυττάρων και  
οργανισμών

Αναγνώριση Μεταλλάξεων και Πολυμορφισμών- Γενετικές ασθένειες

# Στόχοι της Βιοπληροφορικής

---



# **Το Πρόγραμμα του Ανθρώπινου Γονιδιώματος**

---

**Βάσεις Δεδομένων  
Βασικά Εργαλεία & Αλγόριθμοι**

# Γονιδιώματα & Βιοπληροφορική

---

Διάδραση Βιολογίας & Υπολογιστών

Ανάλυση πρωτεϊνών, γονιδίων και γονιδιωμάτων με χρήση αλγορίθμων και βάσεων δεδομένων.

**Genomics ονομάζεται η ανάλυση γονιδιωμάτων.**

Σύγχρονα εργαλεία βιοπληροφορικής χρησιμοποιούνται για την ανάλυση δισεκατομμυρίων βάσεων DNA που προκύπτουν από τα επιμέρους προγράμματα αλληλούχισης γονιδιωμάτων διαφόρων οργανισμών.

# Εφαρμογές Βιοπληροφορικής Ανάλυσης των Γονιδιωμάτων Genomics

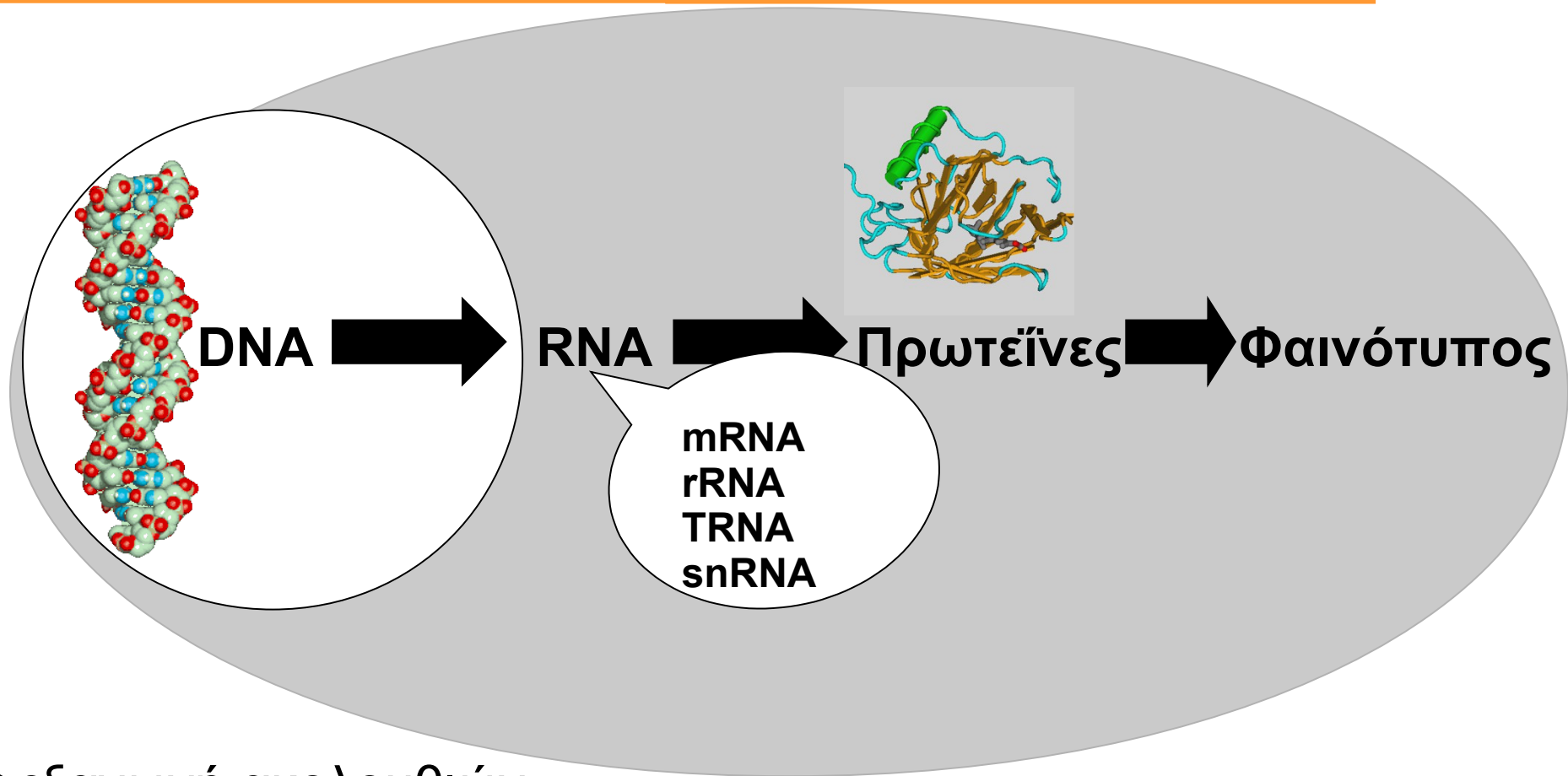
---

**Κύτταρο:** Ταυτοποίηση, επιβεβαίωση & πρόγνωση βιομορίων

**Οργανισμός:** Εξήγηση & πρόγνωση πειραματικών δεδομένων

**Δέντρο της Ζωής:** Αποκαλύπτοντας τις λειτουργίες των γονιδιωμάτων & την εξέλιξη της ζωής

# Κεντρικό Δόγμα Μοριακής Βιολογίας



- εξαγωγή ακολουθιών
- σύγκριση 2 ακολουθιών
- σύγκριση 1 ακολουθίας με άλλες καταχωρημένες σε βάσεις δεδομένων

Γονιδίωμα ➡ Μεταγράφομα ➡ Πρωτέωμα

Κεντρικό Δόγμα της Βιοπληροφορικής & της Γονιδιωματικής



# Τι είναι το ανθρώπινο γονιδίωμα?

---

- Το συνολικό DNA που υπάρχει στον πυρήνα του κυττάρου
- Γονίδια που κωδικοποιούν πρωτεΐνες- mRNA
- Γονίδια που κωδικοποιούν rRNA, tRNA, snRNA που συμμετέχουν στην σύνθεση πρωτεϊνών
- Μη εκφραζόμενες περιοχές DNA – Μη γονίδια

# Επιτεύγματα του Human Genome Project

- Αποκάλυψη της αλληλουχίας των  $\sim 3 \cdot 10^9$  νουκλεοτιδίων που αποτελούν το ανθρώπινο DNA.
- Η σειρά των νουκλεοτιδίων στο DNA (As, Gs, Cs, and Ts) καθορίζει την αμινοξική σύσταση των πρωτεϊνών.
- Αναγνώριση όλων των ( $\sim 25.000$  γονιδίων) που κωδικοποιούν πρωτεΐνες.
- Το ανθρώπινο DNA στην πλειοψηφία του ( $\sim 95\%$ ) περιλαμβάνει μεταξύ των γονιδίων μη εκφραζόμενες αλληλουχίες.
- Σύγκριση του αριθμού των γονιδίων διαφορετικών οργανισμών

*Fugu rubripes* (pufferfish): 20,000 to 25,000  
*Arabidopsis thaliana* (thale cress): 26,000  
*Caenorhabditis elegans* (worm): 19,000  
*Drosophila melanogaster* (fly): 13,000

} **Genes**

- Η διαφορά μεταξύ των γονιδιωμάτων διαφορετικών ανθρώπων  $\sim 1\%$
- 50% του ανθρώπινου γονιδιώματος: επαναλαμβανόμενο DNA (Retrotransposones: LINEs, Alu, LTR)

# Τι άλλο μάθαμε από το Human Genome Project

- Αναγνώριση των φυσιολογικών & μεταλλαγμένων αλληλουχιών των γονιδίων που κωδικοποιούν πρωτεΐνες.
- Οι πλούσιες σε G-C περιοχές του γονιδιώματος αντιστοιχούν σε περιοχές που κωδικοποιούν πρωτεΐνες, ενώ οι περιοχές που είναι πλούσιες σε A-T είναι συνήθως ρυθμιστικές περιοχές ή περιοχές εσωνίων.
- Οι  $\sim 1,4 \cdot 10^6$  πολυμορφισμοί του DNA εξηγούν την ποικιλομορφία μεταξύ των ανθρώπων και δίνουν ώθηση σε νέες τεχνολογίες διάγνωσης, πρόληψης και θεραπείας των ασθενειών.
- Είναι δύσκολο να βρεθεί ο ρόλος των περιοχών του DNA που αντιστοιχούν σε μη μεταφραζόμενα RNA.
- Επίσης περιοχές DNA με άγνωστη λειτουργία και εσώνια χιλιάδων ζ.β.
- Μεγάλος αριθμός Βάσεων Δεδομένων, όπου αποθηκεύονται δεδομένα.
- Μεγάλος όγκος πληροφορίας → Ανάπτυξη εργαλείων για την διεξαγωγή αναλύσεων.

# Το Πρόγραμμα του Ανθρώπινου Γονιδιώματος :

## 8 Στόχοι

---

- 1] Ανθρώπινες DNA αλληλουχίες
- 2] Αναγνώριση πολυμορφισμών αλληλουχιών του Ανθρώπινου Γονιδιώματος
- 3] Λειτουργική Γονιδιωματική
- 4] Συγκριτική Γονιδιωματική
- 5] ELSI: ethical, legal & social issues\*
- 6] Ανάπτυξη τεχνολογιών αλληλούχησης
- 7] Βιοπληροφορική & Υπολογιστική Βιολογία
- 8] Εκπαίδευση & Ανθρώπινο δυναμικό

\* ELSI: Σε ποιον ανήκει η πληροφορία ? Ποιος έχει πρόσβαση σ' αυτή?  
Σε ποιο βαθμό τα γονίδια επηρεάζουν τη συμπεριφορά του ανθρώπου?  
Ποια η σχέση γονιδίων και φυλών?

# Ολοκληρωμένα Γονιδιώματα

| Species and Common Name                                     | Estimated Total Size of Genome (bp)* | Estimated Number of Protein-Encoding Genes* |
|-------------------------------------------------------------|--------------------------------------|---------------------------------------------|
| <i>Saccharomyces cerevisiae</i> (unicellular budding yeast) | 12 million                           | 6,000                                       |
| <i>Trichomonas vaginalis</i>                                | 160 million                          | 60,000                                      |
| <i>Plasmodium falciparum</i> (unicellular malaria parasite) | 23 million                           | 5,000                                       |
| <i>Caenorhabditis elegans</i> (nematode)                    | 95.5 million                         | 18,000                                      |
| <i>Drosophila melanogaster</i> (fruit fly)                  | 170 million                          | 14,000                                      |
| <i>Arabidopsis thaliana</i> (mustard; thale cress)          | 125 million                          | 25,000                                      |
| <i>Oryza sativa</i> (rice)                                  | 470 million                          | 51,000                                      |
| <i>Gallus gallus</i> (chicken)                              | 1 billion                            | 20,000-23,000                               |
| <i>Canis familiaris</i> (domestic dog)                      | 2.4 billion                          | 19,000                                      |
| <i>Mus musculus</i> (laboratory mouse)                      | 2.5 billion                          | 30,000                                      |
| <i>Homo sapiens</i> (human)                                 | 2.9 billion                          | 20,000-25,000                               |

\* There may be other estimates in the literature, but most estimates approximate those listed here.

**Και πολλά ακόμη... (>100.000 species)**

# Επιμέρους Ενότητες

---

- Ορισμός Βιοπληροφορικής
- Εφαρμογές Βιοπληροφορικής

**Μέρος 1**

- Το Πρόγραμμα του γονιδιώματος του Ανθρώπου & άλλων οργανισμών
- Εισαγωγή στις Βιολογικές Βάσεις Δεδομένων

**Μέρος 2**

- Εξάγοντας πληροφορίες για γονίδια & πρωτεΐνες
- Στοίχιση Ακολουθιών: Πολλαπλή & Κατά Ζεύγη

**Μέρος 3**

- Πρακτική Άσκηση

**Μέρος 4**



# Εισαγωγή Στις Βάσεις Βιολογικών Δεδομένων

---

Τι είναι μια βάση δεδομένων;

- Μια συλλογή από...
  - » δομημένα
  - » που δύναται να ψαχθούν (index) -> πίνακας περιεχομένων
  - » περιοδικά ανανεώσιμα (release) -> καινούργια έκδοση

...δεδομένα

- Cross-referenced (hyperlinks) -> συνδέσεις με άλλες βάσεις
- Συμπεριλαμβάνει εργαλεία (λογισμικά πακέτα) για πρόσβαση και ανανέωση στη βάση, καταχώρηση, διαγραφή πληροφοριών κλπ.

Τα δεδομένα δε δημοσιεύονται πλέον με τον παραδοσιακό τρόπο, αλλά κατατίθενται απευθείας σε βάσεις δεδομένων

Απαραίτητα εργαλεία για έρευνα όπως οι κλασικές δημοσιεύσεις !

# Κατηγορίες βιολογικών βάσεων δεδομένων

---

- Αλληλουχίες (DNA, Πρωτεϊνών) -> Πρωτοταγής δομή
- Γονιδιωματική
- Μεταλλάξεις/Πολυμορφισμοί
- Περιοχές πρωτεϊνών /Κατάταξη σε οικογένειες-> Δευτεροταγής δομή
- Πρωτεωμική (δισδιάστατες ηλεκτροφορήσεις, φασματομετρία μάζας)
- Τρισδιάστατη δομή -> Τριτοταγής & Τεταρτοταγής δομή
- Μεταβολισμός
- Ασθένειες και Γενετικές Ανωμαλίες
- Βιβλιογραφία
- Διάφορα άλλα

# Τι πρέπει να έχει μια Βάση Δεδομένων για Ακολουθίες

---

- Ακολουθίες !!
- Μοναδικούς αριθμούς πρόσβασης ακολουθιών [Accession numbers (AC)]
- Αναφορές
- Ταξινομικά Δεδομένα
- Χαρακτηρισμός- Επανάκτηση (Annotation /Curation)
- Λέξεις- Κλειδιά
- Διασυνδεόμενες Αναφορές

**MINIMUM  
REQUIREMENTS !!!**

# Τι είναι το Accession Number?

---

**Μοναδικός κωδικός που συνοδεύει μια ακολουθία.**

Παραδείγματα RBP4 (Retinol- binding protein 4):

X02775      GenBank genomic DNA sequence

NT\_030059 Genomic contig

Rs7079946 dbSNP (single nucleotide polymorphism)

} **DNA**

N91759.1    An expressed sequence tag (1 of 170)

NM\_006744 RefSeq DNA sequence (from a transcript)

} **RNA**

NP\_007635 RefSeq protein

AAC02945    GenBank protein

P02753      UniProt protein

1KT7          Protein Data Bank structure record

} **Πρωτεΐνη**

# Στη ζούγκλα των διαφορετικών Accession Number

| Βάση Δεδομένων          | Ενδεικτικό Accession Number |
|-------------------------|-----------------------------|
| GenBank<br>ENA          | KR711248.1<br>KT584548      |
| UniProtKB               | P02753                      |
| RefSeq nucleotide       | NM_006744.4                 |
| RefSeq protein          | NP_006735.2                 |
| RefSeq prediction       | XM_032502550.1              |
| PDB (protein structure) | 5NU7                        |

# Μετατροπή από τη μια μορφή στην άλλη

The screenshot shows the 'Gene ID Conversion Tool' interface. The top navigation bar includes links: Home, Start Analysis, Shortcut to DAVID Tools, Technical Center, Downloads & APIs, Term of Service, DAVID Forum, and About Us. The main content area is titled 'Gene ID Conversion Tool' and features a 'Gene List Manager' on the left and a 'Submit' button in the center. The 'Gene List Manager' has tabs for 'Upload', 'List', and 'Background'. Under the 'List' tab, there's a section 'Select to limit annotations by one or more species' with a dropdown menu showing '- Use All Species -' and 'Homo sapiens(171)'. Below this is a 'Select' button. Another section 'List Manager' shows a list with 'demolist1' selected. At the bottom of the 'List Manager' are buttons: 'Use', 'Rename', 'Remove', 'Combine', and 'Show Gene List new!'. In the center, there's a 'Submit' button. To the right, a dropdown menu is open, showing 'DAVIDID (Default)' at the top, followed by 'DAVIDID (Default)' and 'ENTREZ\_GENE\_ID' (which is highlighted). Below these are various gene identifier types: AFFY\_ID, GENE\_SYMBOL, GENPEPT\_ACCESSION, GI\_ACCESSION, PIR\_ACCESSION, PIR\_ID, PIR\_NREF\_ID, REFSEQ\_GENOMIC, REFSEQ\_MRNA, REFSEQ\_PROTEIN, REFSEQ\_RNA, UNIGENE, UNIPROT\_ACCESSION, UNIPROT\_ID, and UNIREF100\_ID. Three numbered annotations are present: (1) points to the 'Homo sapiens(171)' selection; (2) points to the 'ENTREZ\_GENE\_ID' selection in the dropdown; (3) points to the 'Submit' button. A text box in the center says 'Convert the gene list being selected in left panel to'.

**Gene ID Conversion Tool**  
DAVID Bioinformatic Resources 2007, NIAID/NIH

Home Start Analysis Shortcut to DAVID Tools Technical Center Downloads & APIs Term of Service DAVID Forum About Us

Upload List Background

**Gene List Manager**

Select to limit annotations by one or more species [Help](#)

- Use All Species -  
Homo sapiens(171)

Select

List Manager [Help](#)

demolist1

Select List to:

Use Rename  
Remove Combine  
Show Gene List new!

**Gene ID Conversion Tool**

Convert the gene list being selected in left panel to

Submit

DAVIDID (Default) v  
DAVIDID (Default)  
ENTREZ\_GENE\_ID  
AFFY\_ID  
GENE\_SYMBOL  
GENPEPT\_ACCESSION  
GI\_ACCESSION  
PIR\_ACCESSION  
PIR\_ID  
PIR\_NREF\_ID  
REFSEQ\_GENOMIC  
REFSEQ\_MRNA  
REFSEQ\_PROTEIN  
REFSEQ\_RNA  
UNIGENE  
UNIPROT\_ACCESSION  
UNIPROT\_ID  
UNIREF100\_ID

1 Select genes in the gene list and its species to be converted

2 Choose the destination gene identifier type to be converted to

3 Submit the work to engine

<https://david.ncifcrf.gov/conversion.jsp?VFROM=NA>

# Βιολογικές βάσεις δεδομένων(>1000!!)

Οι δέκα πιο σημαντικές βάσεις βιοπληροφορικής \*

---

|                  |                                                                                         |                                 |
|------------------|-----------------------------------------------------------------------------------------|---------------------------------|
| GenBank/DDJB/ENA | <a href="https://www.ebi.ac.uk/ena">https://www.ebi.ac.uk/ena</a>                       | Νουκλεοτιδικές αλληλουχίες      |
| Ensembl          | <a href="http://www.ensembl.org">www.ensembl.org</a>                                    | Γονιδιώματα                     |
| PubMed           | <a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a>                          | Αναφορές βιβλιογραφίας          |
| RefSeq-NR        | <a href="https://www.ncbi.nlm.nih.gov/refseq/">https://www.ncbi.nlm.nih.gov/refseq/</a> | Πρωτεϊνικές αλληλουχίες         |
| UniProt          | <a href="https://www.uniprot.org/">https://www.uniprot.org/</a>                         | Πρωτεϊνικές αλληλουχίες         |
| InterPro         | <a href="https://www.ebi.ac.uk/interpro/">https://www.ebi.ac.uk/interpro/</a>           | Συντηρημένες περιοχές πρωτεϊνών |
| OMIM             | <a href="https://www.omim.org/">https://www.omim.org/</a>                               | Γενετικές ασθένειες ανθρώπου    |
| BRENDA           | <a href="https://www.brenda-enzymes.org/">https://www.brenda-enzymes.org/</a>           | Ένζυμα                          |
| PDB              | <a href="https://www.rcsb.org/">https://www.rcsb.org/</a>                               | Δομές πρωτεϊνών                 |
| KEGG             | <a href="https://www.genome.jp/kegg/">https://www.genome.jp/kegg/</a>                   | Μεταβολικά μονοπάτια            |

\* σύμφωνα με το «Bioinformatics for dummies»



# Δύο Τύποι Βάσεων Δεδομένων για Νουκλεοτιδικές Ακολουθίες

---

## Με κέντρο το Γονίδιο

Άμεση πρόσβαση στις  
πληροφορίες σχετικά με  
την ακολουθία καθενός  
γονιδίου

π.χ. **RefSeq**

## Με κέντρο το Γονιδίωμα

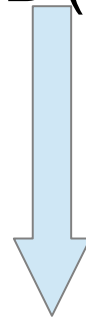
Πληροφορίες σχετικά με τις  
ακολουθίες των γονιδίων, την  
σχετική τους θέση, τον  
προσανατολισμό τους στο  
δίκλωνο DNA, τις βιοχημικές  
τους λειτουργίες...

Συστήματα διαχείρισης  
πληροφοριών που συνδέονται με  
εξειδικευμένες συλλογές  
ακολουθιών και εργαλεία  
αναζήτησης (browsing tools)  
π.χ. **Ensembl**

# Βάσεις Δεδομένων Γονιδιωματική

---

- Περιλαμβάνουν πληροφορίες για: γονίδια, γονιδιακούς τόπους (χαρτογράφηση γονιδίων), ονοματολογία γονιδίων και συνδέσεις (links) προς βάσεις δεδομένων για ακολουθίες. Συνήθως οι ακολουθίες δεν περιλαμβάνονται!
- Υπάρχουν διαθέσιμες βάσεις για καθένα είδος οργανισμού ειδικά. Μεγαλύτερη έμφαση δίνεται σε οργανισμούς σημαντικούς στην έρευνα των επιστημών υγείας.
- Παραδείγματα: OMIM (human), MGI (mouse), FlyBase (Drosophila), SGD (yeast), MaizeDB (maize), κλπ.

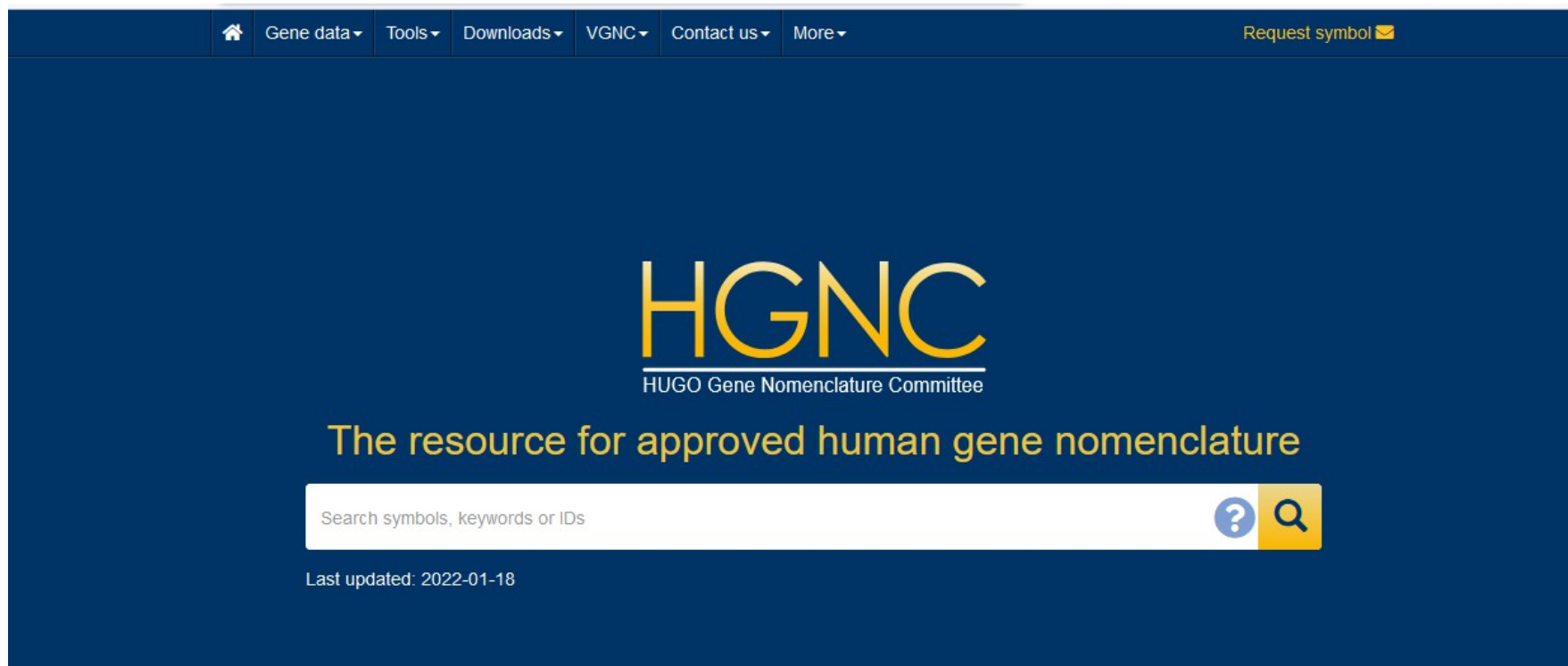


<http://www.informatics.jax.org/>

---

Κοινή μεταξύ των επιστημόνων ονοματολογία και  
ορολογία για τα γονίδια και τις οικογένειές τους

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



# Πρακτική Άσκηση- Μέρος Α

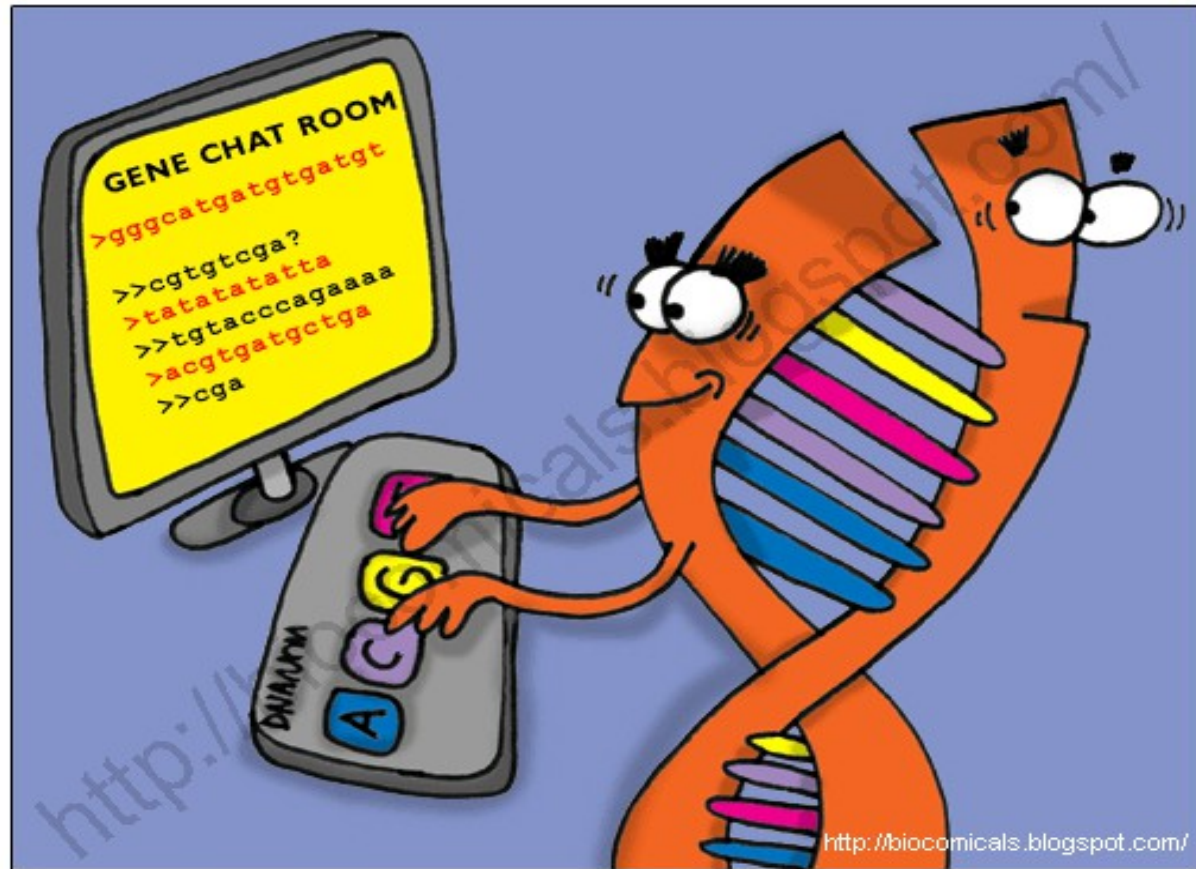
ΕΠΙΣΚΕΥΤΕΙΤΕ ΤΟΝ ΙΣΤΟΤΟΠΟ:

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

και επιλέγοντας ένα γονίδιο της επιλογής σας βρείτε για αυτό

- το εγκεκριμένο σύμβολό του και την κοινά αποδεκτή ονομασία του
- συνώνυμα- παλαιότερα ονόματα και σύμβολα
- τη βιολογική του λειτουργία (functional annotation)
- αν κωδικοποιεί πρωτεΐνη ή έστω μια πολυπεπτιδική αλυσίδα και αν ναι πως ονομάζεται το πρωτεϊνικό προϊόν του.

# Ευχαριστώ



**mkapasa@med.uoa.gr**