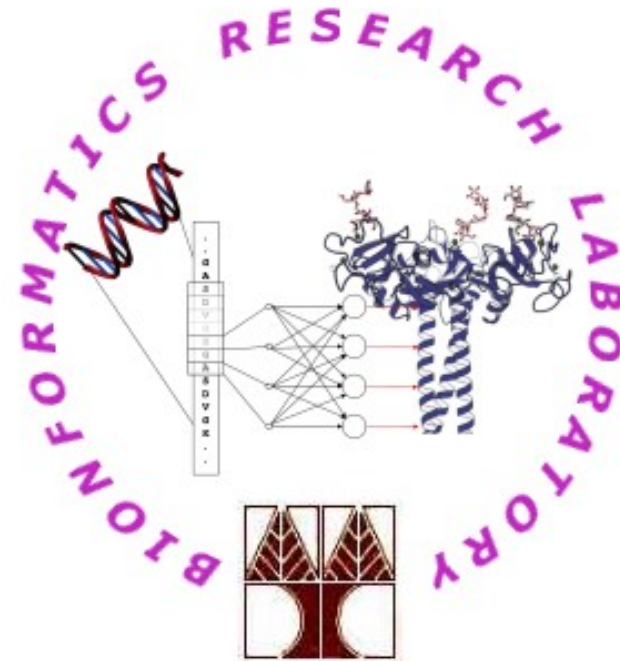


# Βάσεις δεδομένων χαρτογράφησης γονιδιωμάτων



Vasilis Promponas

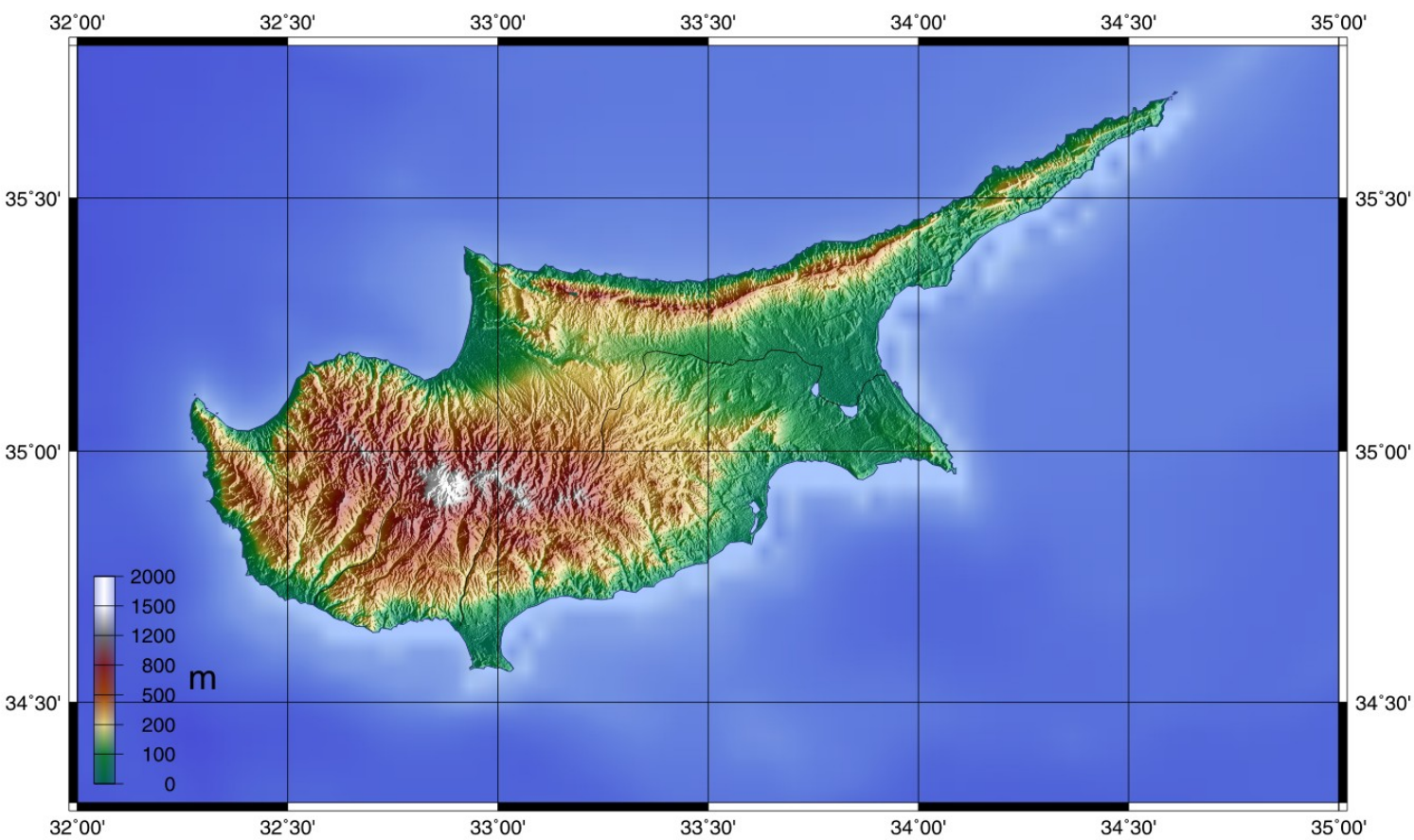
Bioinformatics Research Laboratory

Department of Biological Sciences

University of Cyprus

BIO003

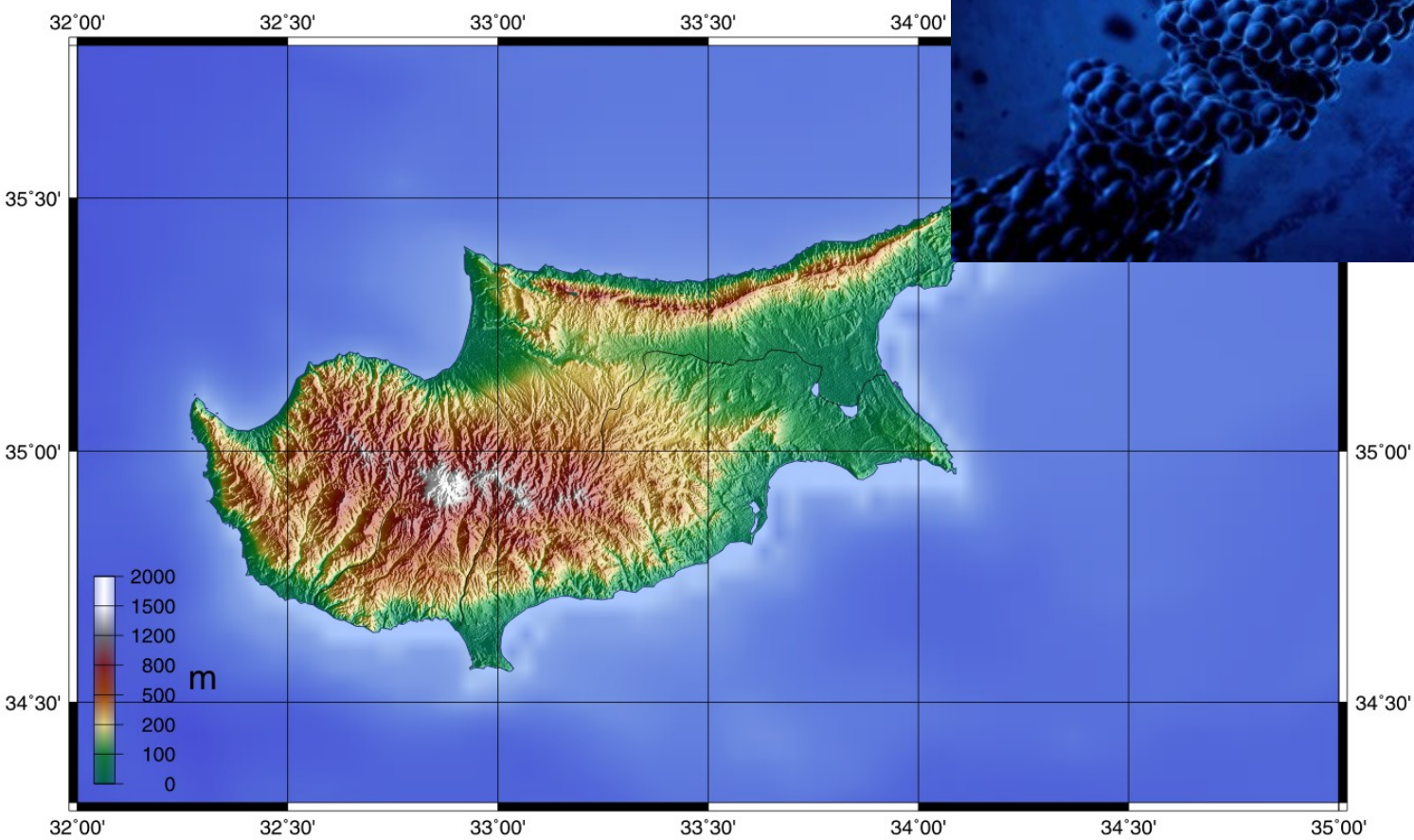
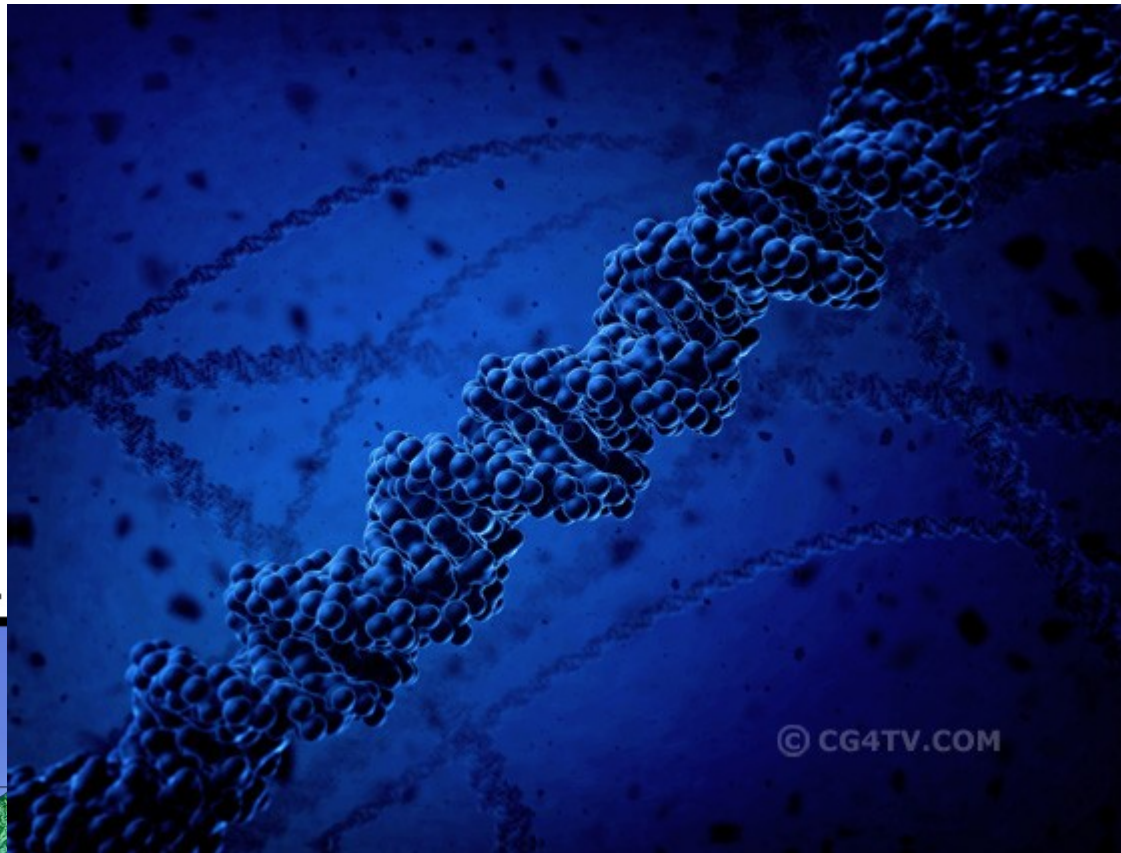
Εισαγωγή στη Βιοπληροφορική



[http://en.wikipedia.org/wiki/Image:Cyprus\\_topo.png](http://en.wikipedia.org/wiki/Image:Cyprus_topo.png)

**Εισαγωγή στη Βιοπληροφορική**

[http://3danimation.e-spaces.com/3d\\_images/hi\\_rez/dna.jpg](http://3danimation.e-spaces.com/3d_images/hi_rez/dna.jpg)



[http://en.wikipedia.org/wiki/Image:Cyprus\\_topo.png](http://en.wikipedia.org/wiki/Image:Cyprus_topo.png)

**Εισαγωγή στη Βιοπληροφορική**

# ΣΥΝΟΨΗ

- Εισαγωγή

- Χαρτογράφηση και Αλληλουχία DNA

- Τύποι Χαρτών

- Το ανθρώπινο γονιδίωμα και οι χάρτες ..

- Συζήτηση ..

## Τύποι Χαρτών

- Κυτταρογενετικοί Χάρτες
- Γενετικοί Χάρτες
- Φυσικοί Χάρτες
- Χάρτες Περιεχομένου STS
- Χάρτες Υβριδισμού με Ακτινοβολία
- Χάρτες Μεταγράφων
- Συγκριτικοί Χάρτες
- Ενοποιημένοι Χάρτες

# Προς αποφυγή παρεξηγήσεων ...



# MOLECULAR STRUCTURE OF NUCLEIC ACIDS

**A Structure for Deoxyribose Nucleic Acid**



J. D. WATSON  
F. H. C. CRICK

Medical Research Council Unit for the  
Study of the Molecular Structure of  
Biological Systems,  
Cavendish Laboratory, Cambridge.  
April 2.

BIO003

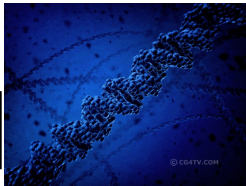
Εισαγωγή στη Βιοπληροφορική

# Εισαγωγή

- Γονιδιωματικός Χάρτης??



Ορισμός: Ένα σύνολο δεικτών με γνωστές σχετικές θέσεις [στο γονιδίωμα/χρωμόσωμα]



- Δείκτης??

Ορισμός: Οποιοδήποτε στοιχείο του γονιδιώματος το οποίο μπορεί να αναγνωρισθεί  
**\*\*ΜΟΝΟΣΗΜΑΝΤΑ\*\***

# Εισαγωγή

- Χαρτογράφηση?
  - Τίνος?
  - Πώς?
  - Γιατί?



# Εισαγωγή

- Χαρτογράφηση?
  - **Τίνος?**
    - Του ... γονιδιώματος! (ή συγκεκριμένων περιοχών του)
  - Πώς?
  - Γιατί?

# Εισαγωγή

- Χαρτογράφηση?

- Τίνος?

- Πώς?

- Με διάφορα μέσα

- Αρκεί να μπορούμε να μετρήσουμε παρατηρήσιμες ιδιότητες κατά μήκος των χρωμοσωμάτων

- Γιατί?

Κυτταρογενετικοί Χάρτες

Γενετικοί Χάρτες

Φυσικοί Χάρτες

Χάρτες Περιεχομένου STS

Χάρτες Υβριδισμού με Ακτινοβολία

Χάρτες Μεταγράφων

Συγκριτικοί Χάρτες

Ενοποιημένοι Χάρτες

# Εισαγωγή

- Χαρτογράφηση?
  - Τίνος?
  - Πώς?
  - Γιατί?
    - Το γονιδίωμα του ανθρώπου (και όχι μόνο) αλληλουχήθηκε
      - Καθοδήγηση αλληλούχησης (clone-2-clone, WGS)
      - Όχι σε ποσοστό 100%
      - Quality Control
    - Χαρτογράφηση άλλων γονιδιωμάτων
    - Ανακάλυψη νέων γονιδίων – Μελέτη γενετικών περιοχών

# Γονιδιωματική χαρτογράφηση

(Γενικά)

- Στους ευκαρυωτικούς οργανισμούς ο απλούστερος τρόπος “κατακερμάτισης” του γονιδιώματος είναι στα χρωμοσώματα από τα οποία αποτελείται
- Το χρωμοσωμικό περιεχόμενο ενός οργανισμού ονομάζεται **Καρυότυπος (Caryotype)**

# Γονιδιωματική χαρτογράφηση

(Λεπτομέρειες ...)

- Ακόμη και τα χρωμοσώματα του ίδιου οργανισμού (γενικά) διαφέρουν
  - στο απόλυτο μέγεθος
    - π.χ. Human: Chr1  $279 \times 10^6$  Chr21  $45 \times 10^6$
  - στη θέση των κεντρομεριδίων
- Σε διαφορετικούς οργανισμούς
  - στο πλήθος τους (π.χ. άνθρωπος 46, πίθηκος 48)
  - σχετικό μέγεθος (π.χ. διπλασιασμοί)
  - εύρος και κατανομή ετεροχρωματίνης

# Χαρτογράφηση και Αλληλοχία

- Γονιδιακοί χάρτες  $\neq$  αλληλουχίες?
  - Χμμ...

# Χαρτογράφηση και Αλληλοχία

- Γονιδιωματικοί χάρτες  $\neq$  αλληλουχίες?
  - ΟΧΙ!!!
    - Η αλληλουχία μιας περιοχής αποτελεί ένα χάρτη υψηλής ευκρίνειας για τη συγκεκριμένη περιοχή
- Οι γονιδιωματικοί χάρτες βοηθούν στην ολοκλήρωση του προσδιορισμού της αλληλουχίας αλλά και στη μελέτη γενικών χαρακτηριστικών του γονιδιώματος



# Δείκτες

(Markers)

- Μονοσήμαντα αναγνωρίσιμο τμήμα του DNA
- Διάφοροι τύποι (1-400 bases)
- Μέθοδοι ανίχνευσης
  - Μέθοδοι PCR [sequence tagged sites-STSSs]
  - Μέθοδοι υβριδισμού

# Κυτταρογενετικοί χάρτες

Cytogenetic (or cytologic) maps

- Αναφέρονται στη “ζωνοποίηση” χρωμοσωμάτων που έχουν υποστεί χρώση και είναι ορατή στο οπτικό μικροσκόπιο
  - Εναλλασσόμενες φωτεινές-σκοτεινές ζώνες
  - Παραδοσιακή τεχνική:
    - Χρωμοσωμική χρωστική (π.χ. μείγμα Giemsa)
    - Σημασμένο DNA probe

# Κυτταρογενετικοί χάρτες

Cytogenetic (or cytologic) maps

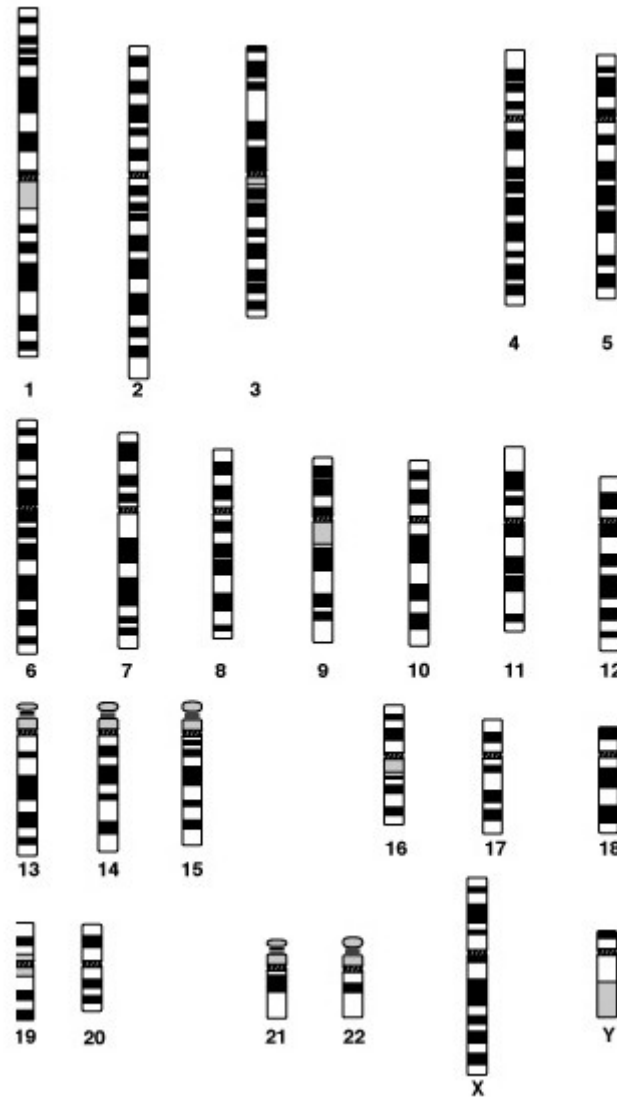
- Κάθε χρωμόσωμα => μοναδικό μοτίβο
  - Καρυότυπος: Επιτρέπει τον εντοπισμό (σημαντικών) χρωμοσωμικών ανωμαλιών (εξαλείψεις, διπλασιασμοί κλπ) με μεγάλη έκταση
  - Με χρήση Fluorescent *in situ* hybridisation (FISH) μπορούμε να έχουμε μια χονδρική εικόνα για τη θέση διαφόρων αλληλουχιών
  - Παρέχουν τη δυνατότητα να διαχωρίσουμε (και να ονοματίσουμε) χρωμοσωμικές περιοχές

# Κυτταρογενετικοί χάρτες

## Μειονεκτήματα

- Παρέχουν (με χαμηλή ανάλυση) τη δυνατότητα να διαχωρίσουμε (και να ονοματίσουμε) χρωμοσωμικές περιοχές
  - π.χ. αλληλουχίες που χωρίζονται από  $< 3\text{M}$  βάσεις ( $3\text{Mbp}$ ) εμφανίζονται στην ίδια περιοχή
- Η ανίχνευση των κυτταρογενετικών θέσεων εξαρτάται από υποκειμενικά κριτήρια, π.χ.
  - τεχνολογία ανίχνευσης – μεθοδολογία
  - καθορισμός ορίων των ζωνών

# Ανθρώπινος κυτταρογενετικός χάρτης



<http://www.ncbi.nlm.nih.gov/Class/MLACourse/Original8Hour/Genetics/cytogeneticmap.gif>

BIO003

Εισαγωγή στη Βιοπληροφορική

# Χάρτες γενετικών συνδέσεων

(Genetic linkage maps)

- Αποτελούν απεικόνιση της “απόστασης” μεταξύ δύο στοιχείων του DNA
- Βασίζονται στη συχνότητα ανασυνδυασμού (recombination) που συμβαίνει κατά τη μείωση
  - Γονίδια που συν-κληρονομούνται έχουν μικρή φυσική απόσταση (... βρίσκονται κοντά ...)
  - Λογικά: “Δεν υφίσταται “απόσταση” μεταξύ γονιδίων που βρίσκονται σε διαφορετικά χρωμοσώματα
- ΙΣΤΟΡΙΑ:
  - Πρωτοχρησιμοποιήθηκε τη δεκαετία του 1910 (διασταυρώσεις *Drosophila melanogaster*)

# Παρένθεση

(Ανασυνδυασμός - Recombination)

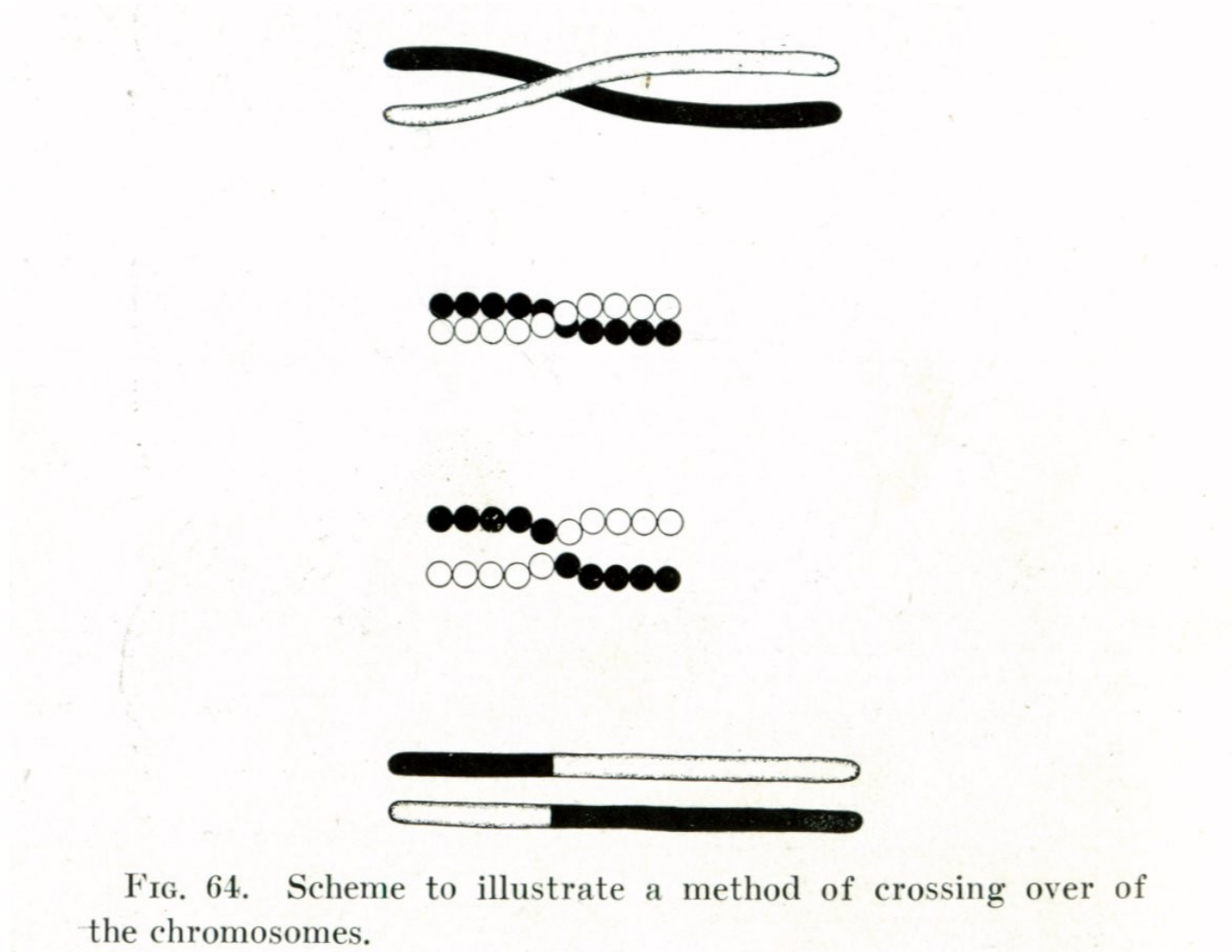


FIG. 64. Scheme to illustrate a method of crossing over of the chromosomes.



## ... και λίγα μαθηματικά ...

- Δύσκολη ποσοτική εκτίμηση του πλήθους των γεγονότων ανασυνδυασμού
- Δύο κύριες προσεγγίσεις
  - Παραμετρική ανάλυση σύνδεσης
  - Μη-παραμετρική ανάλυση σύνδεσης

# Παραμετρική Ανάλυση Σύνδεσης

- Στατιστική μέθοδος “Μέγιστης Πιθανοφάνειας” (Maximum Likelihood)
  - Για 2 δείκτες A, B σύγκριση της πιθανοφάνειας
    - Έστω  $\theta$  ο λόγος ανασυνδυασμού
    - οι A, B συνδέονται
    - οι A, B είναι μη-συνδεδεμένοι
    - Υπολογισμός lod-score ( $\text{lod} \equiv \log$  of the odds)
      - $$\text{LOD} = \log_{10} \frac{L(\theta = \bar{\theta})}{L(\theta = 0.5)}$$
      - τιμή κατωφλίου  $\text{lod} > 3$
- Η εκτιμήτρια ML υπολογίζεται με κόπτο ...
  - Μικρός αριθμός δεικτών
  - Μικρό μέγεθος γενεαλογικού δέντρου

# Ερωτήσεις

- Τι σημαίνει  $\text{lod} > 3$  για τις αντίστοιχες πιθανοφάνειες;
- Μη παραμετρική ανάλυση σύνδεσης (non-parametric linkage analysis - NPL):
  - να δώσετε σε μία σελίδα το σκεπτικό πίσω από αυτήν την προσέγγιση
  - να κατονομάσετε 3 πακέτα λογισμικού που πραγματοποιούν ανάλυση NPL

[Οι απαντήσεις σας με το επόμενο πακέτο εργασιών για το σπίτι ...]

# Γενετική απόσταση

- Ο λόγος ανασυνδυασμού μετατρέπεται μέσω μιας “συνάρτησης χαρτογράφησης” σε απόσταση
  - αθροιστική μονάδα απόστασης 1 centiMorgan (1cM)
  - $\Delta(A,B) = 1\text{cM} \rightarrow 1\%$  πιθανότητα να εμφανιστεί ανασυνδυασμός μεταξύ των δύο δεικτών
    - Η αντίστοιχη φυσική απόσταση εξαρτάται από τον οργανισμό
      - *Homo sapiens*: 1cM ~ 1Mb
      - *Drosophila melanogaster*: 1cM ~ 0.5Mb

# Χάρτες γενετικών συνδέσεων

(Genetic linkage maps)

- Μειονεκτήματα
  - Απαιτεί ύπαρξη φαινοτύπου
  - Υποθέτει ότι όλες οι περιοχές του γονιδιώματος έχουν την ΙΔΙΑ πιθανότητα ανασυνδυασμού
    - Αυτό δεν ισχύει
    - Recombination “HOT SPOTS”
  - Για ανθρώπινο γενετικό χάρτη
    - Θα κάνουμε διασταυρώσεις ?!?!?!?
      - όμως ... έχουμε δεδομένα από οικογένειες
        - Φαινοτυπικά
        - Πολυμορφισμοί (SNPs, VNTRs, μικροδορυφόροι)

# Φυσικοί Χάρτες

(Physical maps)

- Επιτρέπουν ΑΜΕΣΗ μέτρηση αποστάσεων μεταξύ γενετικών δεικτών
- Είναι συνήθως το πρώτο βήμα στη μελέτη των γονιδιωμάτων
- Κύριοι τύποι
  - Περιοριστικοί χάρτες
  - Χάρτες υβριδισμού με ακτινοβολία
  - Χάρτες περιεχομένου STS

# Περιοριστικοί Χάρτες

(Restriction maps)

- Περιοριστικά ένζυμα “κόβουν” το γονιδιωματικό DNA
  - αναγνωρίζουν μικρές, συγκεκριμένες αλληλουχίες
  - έχουν μεγάλη ειδικότητα (π.χ. 5'-GCGGCCGC-3', NotI)
  - ευαισθησία σε πολυμορφισμούς
    - π.χ. η NotI δεν αναγνωρίζει την 5'-GCGGACGC-3'



## .. και τώρα εσείς ..

- **Ερώτηση:** Με δεδομένο ότι οι βάσεις του DNA έχουν ισοπίθανη εμφάνιση, πόσες περιοχές αναγνώρισης του NotI περιμένετε να υπάρχουν στο ανθρώπινο γονιδίωμα?
- **Ερώτηση:** Πειραματικά, το NotI “κόβει” το ανθρώπινο γονιδίωμα κατά μέσο όρο κάθε 10Mbp. Γιατί???

# Χάρτες υβριδισμού με ακτινοβολία

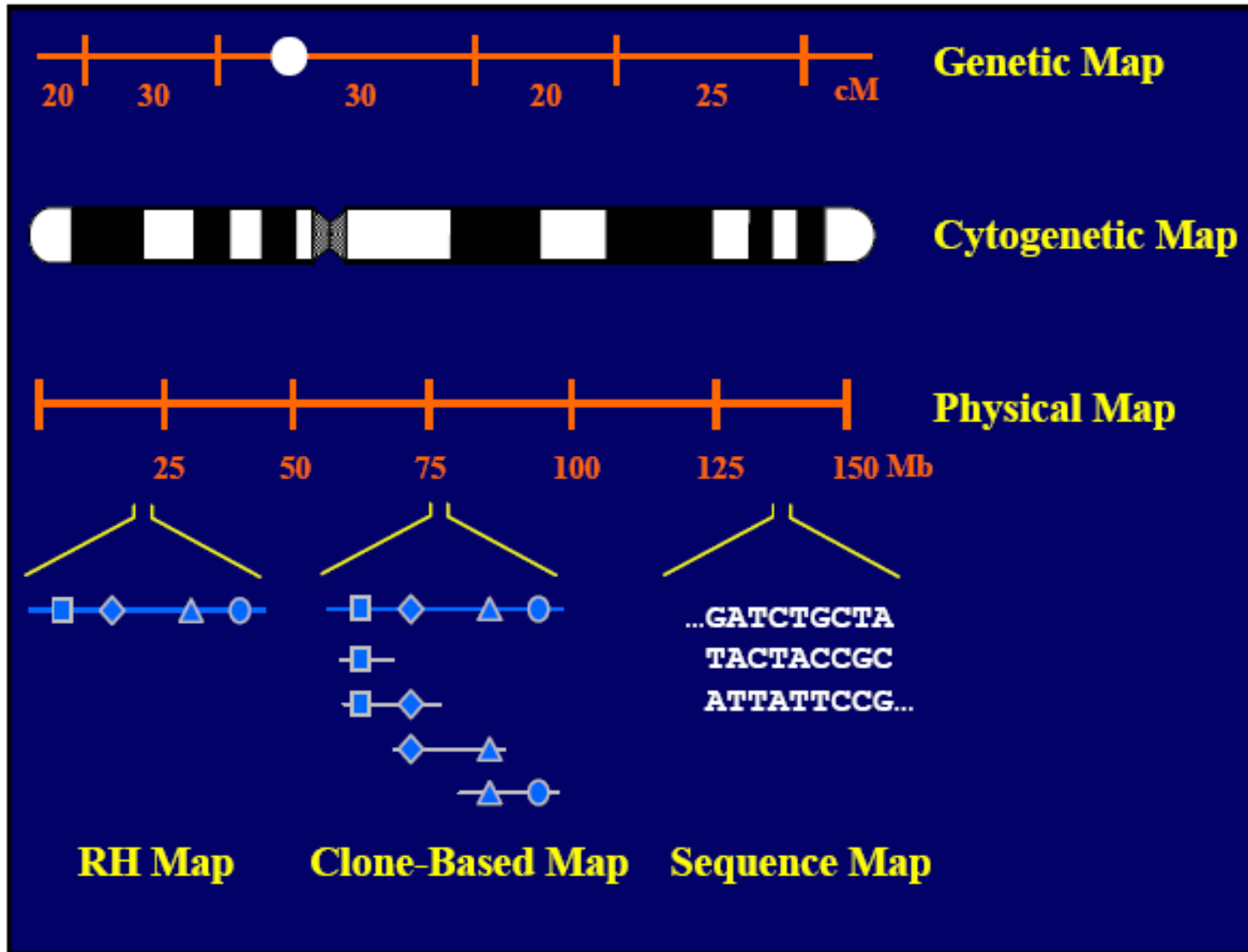
(Radiation hybrid maps)

- Χρησιμοποιούμε ακτινοβολία (X-rays) για να δημιουργήσουμε τυχαίες θραύσεις στο γονιδίωμα που μελετάμε
- Το μήκος των θραυσμάτων είναι αντιστρόφως ανάλογο της δόσης
- Τα κύτταρα πεθαίνουν, αλλά τα θραύσματα του DNA εισάγονται σε ζωντανά κύτταρα από κυτταροκαλλιέργειες
- Ελέγχεται η θέση δεικτών με υβριδισμό ή PCR
- Δείκτες στο ίδιο θραύσμα -> μικρή απόσταση

# Χάρτες περιεχομένου STS

(STS content maps)

- Σχετικά απλή τεχνική-ανάλυση σχετικά μεγάλων περιοχών (>1Mb)
- STS (Sequence tagged site)
  - ένα τμήμα DNA (100-200bp) το οποίο δημιουργείται με PCR με βάση γνωστούς primers
- Βασίζεται σε έλεγχο δεικτών STS με PCR έναντι βιβλιοθήκης κλώνων με μεγάλους ενθέτες



Εικόνα από: Techniques for Genome Mapping and Sequencing, Eric Green, NHGRI

[http://www.genome.gov/COURSE2003/PDF/CTGA\\_lect1.pdf](http://www.genome.gov/COURSE2003/PDF/CTGA_lect1.pdf)

# Συγκριτικοί Χάρτες

(Comparative maps)

- Ταυτοποίηση συντηρημένων χρωμοσωμικών τμημάτων μεταξύ διαφορετικών ειδών (π.χ. άνθρωπος-ποντίκι)
- Χρησιμότητα
  - Οργανισμοί μοντέλα
    - Βιολογικές διεργασίες
    - Ασθένειες
    - ...

# Ενοποιημένοι χάρτες

(Integrated maps)

- Συνδυασμός πληροφοριών από διάφορες προσεγγίσεις χαρτογράφησης
- Ενοποιημένη-συμπληρωματική πληροφορία
- Δυσκολία
  - Οι διάφοροι χάρτες δεν έχουν τις ίδιες “μονάδες” μέτρησης ...

# Διαδικτυακές πηγές δεδομένων

- The Genome DataBase (GDB)
  - <http://www.gdb.org>
- Mouse Genome Informatics
  - <http://www.informatics.jax.org>

# Συζήτηση ...

## Χρήσιμες διαδικτυακές πηγές:

NHGRI mini course: Current topics in Genome Analysis  
<http://www.genome.gov/COURSE2003/handouts.cfm>

## Διδακτικό υλικό:

Βιοπληροφορική (Baxevanis, Ouellette) Κεφάλαιο 6,

<http://troodos.biol.ucy.ac.cy/BRL/courses/BIO003/index.html>