



# ΕΙΔΙΚΑ ΘΕΜΑΤΑ ΓΕΝΕΤΙΚΗΣ

## Ενότητα 2<sup>η</sup>: Δομική γονιδιωματική

Τριανταφυλλίδης Α  
Τμήμα Βιολογίας

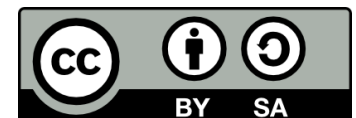


Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



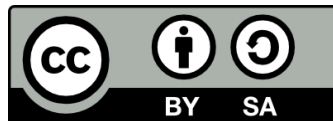
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Άδειες Χρήσης

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



# Χρηματοδότηση

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



# Άδεια χρήσης εικόνων

Ευχαριστούμε θερμά τις Ακαδημαϊκές Εκδόσεις Μπάσδρα για την παραχώρηση του δικαιώματος χρήσης των εξής εικόνων της παρούσης παρουσίασης:

Εικόνες: 4-9, 12

Οι εικόνες αυτές προέρχονται από το βιβλίο Peter Russell, iGenetics: Μια μεντελική προσέγγιση, 1η έκδοση, Ακαδημαϊκές Εκδόσεις Ι. Μπάσδρα και ΣΙΑ Ο.Ε.



# Περιεχόμενα ενότητας

- Στόχοι Μελέτης
- DNA – Ο φορέας της γενετικής πληροφορίας
- Προβλήματα στην αλληλούχιση και μέθοδοι αντιμετώπισης τους
- Χρωμοσωμικοί, Γενετικοί, Φυσικοί χάρτες και Χάρτες αλληλουχίας (και Ασκήσεις)
- Νέα μηχανήματα αλληλούχισης
- Συναρμολόγηση και αποθήκευση δεδομένων
- Το μέλλον των αναλύσεων
- Εύρεση γονιδίων στις αλληλουχίες γονιδιωμάτων



# Στόχοι Μελέτης 1

- Οι στρατηγικές αλλά και οι προκλήσεις κατά την ανάλυση του γονιδιώματος
- Τα προβλήματα που προκύπτουν από την παρουσία πολυμορφισμού στα γονιδιώματα
- Η μελέτη τριών διαφορετικών γονιδιακών χαρτών: των χαρτών σύνδεσης υψηλής ανάλυσης, των φυσικών χαρτών μεγάλης εμβέλειας και των χαρτών αλληλούχισης



# Στόχοι Μελέτης 2

Τα κύρια συμπεράσματα που προκύπτουν από ολοκληρωμένα ή ημι-ολοκληρωμένα προγράμματα αλληλούχισης γονιδιωμάτων διαφόρων ειδών, όπως:

- ο αριθμός και ο τύπος των γονιδίων,
- το ποσοστό των επαναλαμβανόμενων ακολουθιών,
- η οργάνωση και η δομή των γονιδιώματων,
- η εξέλιξή τους, και
- οι μελλοντικές κατευθύνσεις της έρευνας



# Στόχοι Μελέτης 3

Οι τεχνικές και οι μέθοδοι ανάλυσης του γονιδιώματος σε ευρεία κλίμακα, όπως:

- Τα αυτόματα μηχανήματα εύρεσης αλληλουχίας (**sequencers**),
- Οι συσκευές μικροσυστοιχίας DNA (**microarrays**),
- Οι φασματογράφοι μαζών (**mass spectrometers**).





# DNA – Νουκλεϊνικά οξέα Φορείς Γενετικής πληροφορίας (1/5)

Οποιοσδήποτε Φορέας Πληροφορίας πρέπει να έχει τις παρακάτω ιδιότητες

1. Αποθήκευση – Κωδικοποίηση πληροφορίας
2. Διπλασιασμός – Αναπαραγωγή πληροφορίας
3. Μεταγραφή – Μεταφορά πληροφορίας
4. Δημιουργία Ποικιλομορφίας – Εξέλιξη

Το DNA είναι τέτοιο μόριο



# DNA – Νουκλεϊνικά οξέα Φορείς Γενετικής πληροφορίας (2/5)

## Το παράδοξο της τιμής του C

Τιμή του C: Η συνολική ποσότητα DNA σε κάθε απλοειδές κύτταρο

Τιμή σταθερή, χαρακτηριστική για κάθε οργανισμό

Οι πιο εξελιγμένοι οργανισμοί δεν έχουν απαραίτητα και μεγαλύτερη τιμή C, π.χ. ένα πρῶτιστο, *Polychaos dubium* (πρώην *Amoeba dubia*) με γονιδίωμα 670 Gb έχει >200 φορές μεγαλύτερο γονιδίωμα από ότι ο άνθρωπος.

<http://www.genomesize.com/statistics.php>



# DNA – Νουκλεϊνικά οξέα Φορείς Γενετικής πληροφορίας (3/5)

Ούτε ο αριθμός των πρωτεϊνικών γονιδίων σχετίζεται με την πολυπλοκότητα  
Για παράδειγμα ο νηματώδης σκώληκας, παρ' όλο που είναι εξελικτικά κατώτερος  
οργανισμός (πιο απλός) από τη Δροσόφιλα έχει μεγαλύτερο αριθμό τέτοιων γονιδίων.

| Οργανισμός         | # Γονιδίων | Μέγεθος<br>γονιδιώματος (Mb) |
|--------------------|------------|------------------------------|
| <i>E. coli</i>     | 4.288      | 4,64                         |
| Ζύμη               | 6.600      | 12,1                         |
| <i>C. elegans</i>  | 20.000     | 97                           |
| <i>Drosophila</i>  | 15.000     | 170                          |
| <i>Arabidopsis</i> | 25.000     | 120                          |
| Ποντίκι            | ~30.000    | 2.600                        |
| Άνθρωπος           | ~30.000    | 3.200                        |



# DNA – Νουκλεϊνικά οξέα Φορείς Γενετικής πληροφορίας (4/5)

## Τι είναι αυτό που μετράει τελικά?

Οι μεταγραφικοί παράγοντες!

- Ζύμη 300
- Δροσόφιλα 1000
- Άνθρωπος 3000

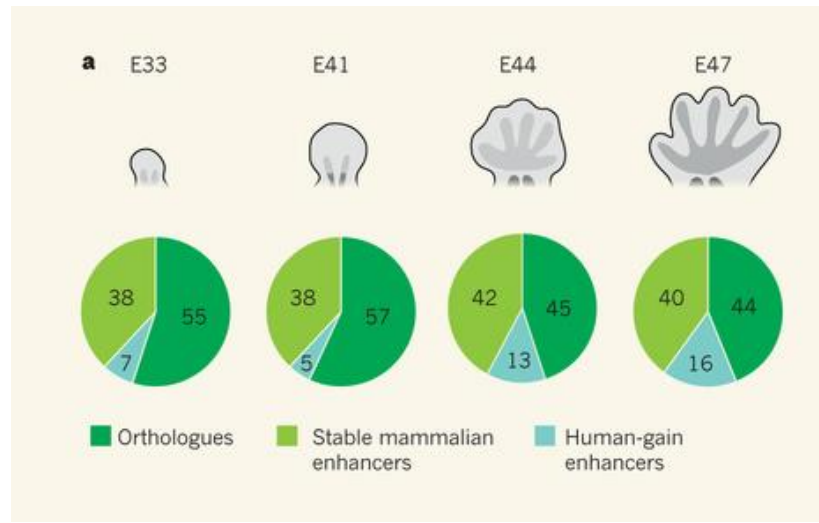
Στους περισσότερο πολύπλοκους οργανισμούς υπάρχουν:

- Πιο πολύπλοκες ρυθμιστικές περιοχές
- Περισσότερες ρυθμιστικές πρωτεΐνες



# DNA – Νουκλεϊνικά οξέα Φορείς Γενετικής πληροφορίας (5/5)

Στην εργασία των Gilad *et al.* (Nature 2006 440 242-245) έγινε ανάλυση των επιπέδων έκφρασης σε 1056 κοινά γονίδια μεταξύ ανθρώπου, χιμπατζή, ουρακοτάγκου και μακάκου. Αποκαλύφθηκε ότι γονίδια που κωδικοποιούν μεταγραφικούς παράγοντες είναι ιδιαίτερα ενεργά στον άνθρωπο. Κάποια άλλα έχουν παρόμοια έκφραση και συνεπώς είναι απαραίτητα για όλα τα πρωτεύοντα.



**Εικόνα 1:** Φαίνεται η εξέλιξη των ενισχυτών των γονιδίων που σχετίζονται με την ανάπτυξη των άκρων, όπως μελετήθηκαν στα είδη: ποντίκι, μακάκος, άνθρωπος. Πέρα από τα κοινά γονίδια, υπάρχουν περισσότερα νέα **στον άνθρωπο** (Cotney *et al.* 2013 Cell, 154, 185-196)



# Ανάλυση γονιδιώματων

Οι ερευνητές στην προσπάθεια τους να μελετήσουν το γονιδίωμα ενός οργανισμού ενδιαφέρονται για δύο κύρια στοιχεία του:

- Την **εύρεση της αλληλουχίας** κατά μήκος όλων των χρωμοσωμάτων ενός απλοειδούς γονιδιώματος.
- Την **αποκάλυψη της ποικιλομορφίας** ή αλλιώς της ύπαρξης διαφορετικών αλληλομόρφων που συναντώνται μεταξύ των διαφορετικών ατόμων του ίδιου είδους ή και στις δύο σειρές ομόλογων χρωμοσωμάτων (αν αναφερόμαστε σε διπλοειδή οργανισμό).



# Προβλήματα στην αλληλούχιση (1/5)

Τα προβλήματα που μπορούν όμως να προκύψουν στην αλληλούχιση είναι τα εξής:

- Τα λάθη κατά την αλληλούχιση
- Ο πολυμορφισμός των γονιδιωμάτων
- Οι επαναλαμβανόμενες αλληλουχίες που εντοπίζονται με δυσκολία
- Τα τμήματα DNA που δεν κλωνοποιούνται και συνεπώς δεν μπορούν και να αλληλουχηθούν



# Προβλήματα στην αλληλούχιση (2/5)

Το μέγεθος ενός χρωμοσώματος ποικίλλει (π.χ. στον άνθρωπο 700kb- 500 Mb). Η τεχνολογία όμως δεν επιτρέπει την επιτυχή εύρεση της ακολουθίας με μια μόνο προσπάθεια (διάβασμα) σε κομμάτια DNA παραπάνω από 600 bp

Πώς είναι δυνατή η επιτυχής αλληλούχιση ενός χρωμοσώματος 500 Mb με ρυθμό 600 βάσεις τη φορά;

Πόσο ακριβές είναι το αποτέλεσμα της αλληλούχισης όταν το ποσοστό λάθους είναι περίπου 1% ανά διάβασμα;

☞ Το HGP αποφάσισε ότι η τελική κατατεθειμένη ακολουθία του ανθρώπινου γονιδιώματος δεν πρέπει να έχει ποσοστό λάθους υψηλότερο από  $1 / 10.000$  bp ή 0,01%





# Προβλήματα στην αλληλούχιση (3/5)

## Πολυμορφισμός

Οι ανώτεροι οργανισμοί είναι διπλοειδείς. Επομένως τα γονιδιώματα περιλαμβάνουν τόσο μητρικά όσο και πατρικά χρωμοσώματα → διαφορετικά αλληλόμορφα των γονέων, δημιουργούν τους **πολυμορφισμούς**

Στον άνθρωπο ο πολυμορφισμός έχει υπολογιστεί ότι είναι 1/500 bp

Πώς είναι δυνατόν να διακρίνουμε τον πολυμορφισμό από τα λάθη της αλληλούχισης;



# Προβλήματα στην αλληλούχιση (4/5)

## Επαναλαμβανόμενες αλληλουχίες

Τα γονιδιώματα των ανώτερων οργανισμών περιλαμβάνουν πλήθος επαναλαμβανόμενων ακολουθιών. Οι ακολουθίες αυτές μπορεί να έχουν μέγεθος μερικές δεκάδες βάσεις, π.χ. (GT) $n$  ή εκατοντάδες κιλοβάσεις.

Όταν δημιουργούνται (με διπλασιασμό ή άνισο διασκελισμό), μοιάζουν μεταξύ τους. Με την πάροδο των χρόνων διαφοροποιούνται εξαιτίας μεταλλάξεων και εκφυλισμού τους. Οι πιο νέες ακολουθίες μοιάζουν πολύ, ενώ οι πιο παλιές διαφέρουν.

Πρόβλημα με νέες ακολουθίες (που μοιάζουν πολύ και είναι διάσπαρτες στο γονιδίωμα) μεγέθους  $> 600$  βάσεις. Αν ένα «διάβασμα» πέσει σε ακολουθία διασπαρμένη σε 10.000 αντίγραφα σε διάφορα χρωμοσώματα, πώς θα βρεθεί ποιο αντίγραφο αναλύεται και σε ποιο χρωμόσωμα βρίσκεται;



# Προβλήματα στην αλληλούχιση (5/5)

## Τμήματα DNA που δεν κλωνοποιούνται

Ιδιαίτερα οι ετεροχρωματινικές περιοχές του DNA δεν κλωνοποιούνται εύκολα σε φορείς κλωνοποίησης.

Αν και η ετεροχρωματίνη περιέχει πολύ λίγα γονίδια αποτελεί περίπου το 30% του συνολικού DNA.

Αν όμως δεν μπορούν να κλωνοποιηθούν αυτές οι περιοχές, πώς θα γνωρίζουμε τι περιέχουν;



# Μέθοδοι αντιμετώπισης προβλημάτων (1/3)

Οι επιστήμονες ακολουθούν την τακτική «**διαίρει και βασίλευε**»

Τα χρωμοσώματα σπάζουν σε πολύ μικρά και αλληλοεπικαλυπτόμενα κομμάτια και κλωνοποιούνται σε βιβλιοθήκες.

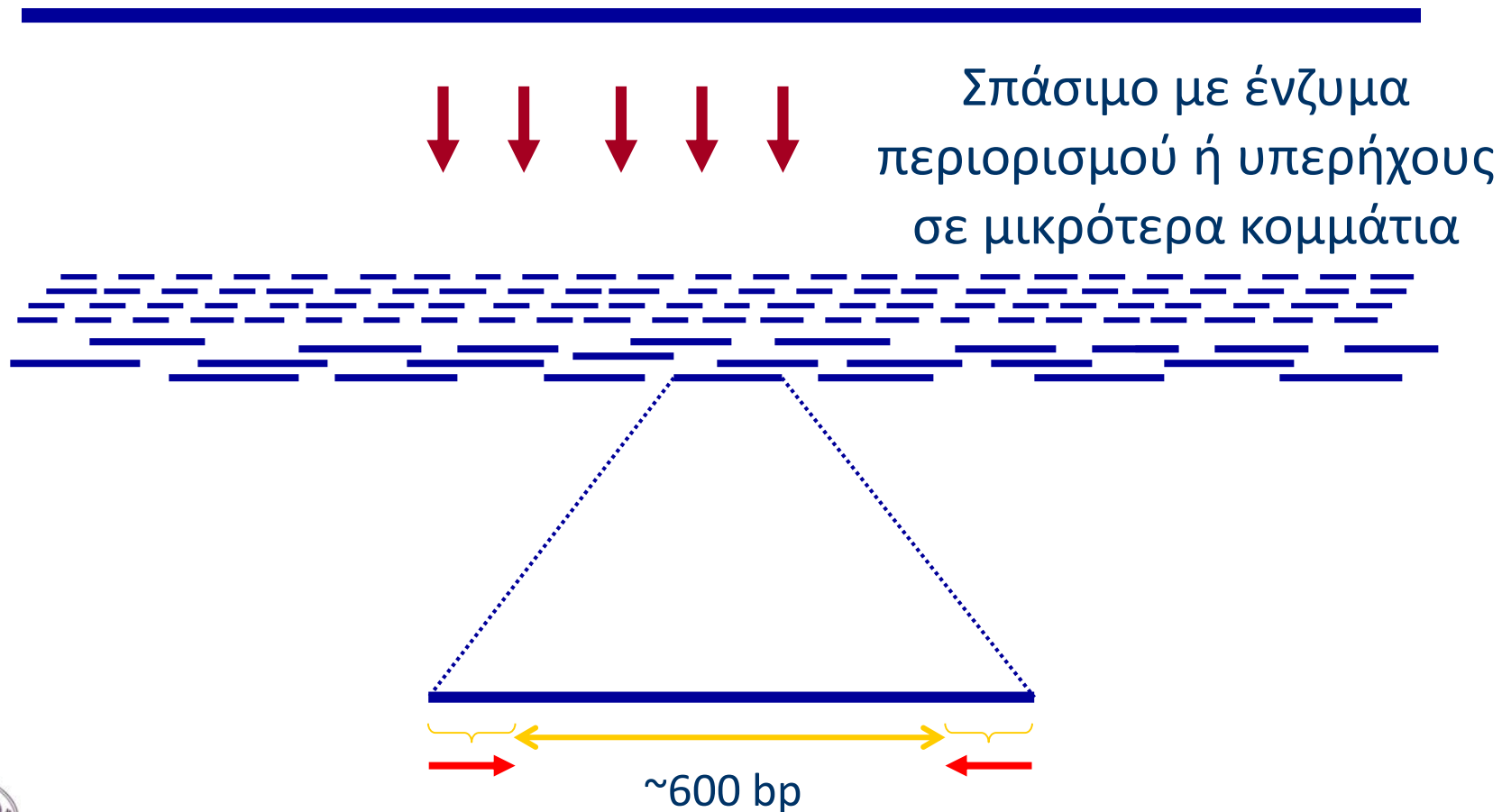
Η αλληλούχιση μπορεί να επιτευχθεί κατά μήκος όλων σχεδόν των κομματιών. Αφού υπάρχει αλληλοεπικάλυψη, είναι δυνατή η επανασυναρμολόγηση των επιμέρους κομματιών στο συνολικό χρωμόσωμα.

Η επανασυναρμολόγηση αυτή δεν είναι εύκολη υπόθεση και σίγουρα δε γίνεται χωρίς τη χρήση υπολογιστών.



# Μέθοδοι αντιμετώπισης προβλημάτων (2/3)

## Γονιδίωμα



# Μέθοδοι αντιμετώπισης προβλημάτων (3/3)

Για να επιτευχθεί ποσοστό λάθους  $1/10.000$ , κάθε τμήμα ακολουθίας DNA πρέπει να αλληλουχηθεί πολλαπλές φορές από διαφορετικούς κλώνους. Χρειάζεται να αλληλουχηθούν τουλάχιστον 10 διαφορετικοί κλώνοι → δεκαπλάσια κάλυψη της αλληλουχίας.

Τα λάθη εμφανίζονται σε  $1/10$  διαβάσματα, ενώ οι πολυμορφισμοί εμφανίζονται σε  $5/10$  διαβάσματα. Ταυτόχρονα όμως, δεκαπλασιάζονται και τα δεδομένα αλληλουχίας που επεξεργάζονται οι υπολογιστές για να συναρμολογήσουν το γονιδίωμα.



# Δύο Βασικές Μέθοδοι Αλληλούχισης

Ολική γονιδιωματική προσέγγιση (whole genome shotgun sequencing) → Τυχαία πέψη και αλληλούχιση τμημάτων DNA (Εταιρεία Celera). Είναι πιο γρήγορη, αλλά όχι τόσο ακριβής.

Ιεραρχημένη προσέγγιση (hierarchical shotgun sequencing) → Αλληλούχιση τμημάτων DNA που οι ερευνητές είχαν ήδη τοποθετήσει σε φυσικούς χάρτες (Πρόγραμμα HGP). Είναι εξαιρετικά ακριβής, αλλά απαιτεί ιδιαίτερα δαπανηρή και χρονοβόρα προετοιμασία.



# Τεχνικές για ανάλυση γονιδιώματος 1

**Κλωνοποίηση:** DNA μεγέθους 500 - 1.000.000 bp κλωνοποιείται σε πλήθος φορέων. Το σύνολο των κλωνοποιημένων τμημάτων DNA αποτελεί μια βιβλιοθήκη.

Σε μια βιβλιοθήκη με καλή κάλυψη θα πρέπει οι ακολουθίες DNA να συναντώνται τουλάχιστον σε 10 ανεξάρτητους κλώνους. Έτσι θα είναι δυνατό να διαβαστεί το γονιδίωμα 10 ανεξάρτητες φορές και να αποφευχθούν τα λάθη.





# Τεχνικές για ανάλυση γονιδιώματος 2

**Υβριδισμός:** Χρησιμοποιείται για να εντοπιστεί η ακριβής θέση τμημάτων DNA σε μια βιβλιοθήκη. Στηρίζεται στην συμπληρωματικότητα μιας αλληλουχίας DNA στόχου με ένα τμήμα DNA από τη βιβλιοθήκη.

Ταυτοποιούνται οι κλώνοι που περιλαμβάνουν κάποιο γονίδιο ή αποτελούν μέρος χρωμοσώματος. Έτσι, γίνεται δυνατή η χαρτογράφηση και η συνέχιση της αλληλούχησης.



# Τεχνικές για ανάλυση γονιδιώματος 3

**Ενίσχυση με PCR:** Ενίσχυση τμημάτων DNA πάνω από 1.000.000 φορές. Μέγεθος από 1 Kb έως και 20 Kb.

Απαιτούνται εκκινητές ειδικοί για τη συγκεκριμένη ακολουθία, αλλιώς ενισχύονται ταυτόχρονα πολλαπλά διαφορετικά τμήματα και δημιουργούνται προβλήματα στην αλληλούχιση.

**Αλληλούχιση DNA:** Αυτόματα μηχανήματα (sequencers) διαβάζουν ακολουθίες κλωνοποιημένων τμημάτων DNA ποικίλων μεγεθών (από 50 bp έως >600 bp).



# Τεχνικές για ανάλυση γονιδιώματος 4

## Υπολογιστικά εργαλεία

Υπάρχουν 4 διαφορετικά προγράμματα υπολογιστών, αυτά που :

1. Ελέγχουν αν η ακολουθία μοιάζει με ήδη αλληλουχηθέντα τμήματα DNA (που υπάρχουν δηλ. σε βάσεις δεδομένων)
2. Ανακαλύπτουν τις αλληλεπικαλύψεις μεταξύ των τμημάτων DNA και τα συναρμολογούν σε μια σειρά
3. Υπολογίζουν το ρυθμό λάθους στις διάφορες ακολουθίες
4. Ανακαλύπτουν και ταυτοποιούν γονίδια στα χρωμοσώματα

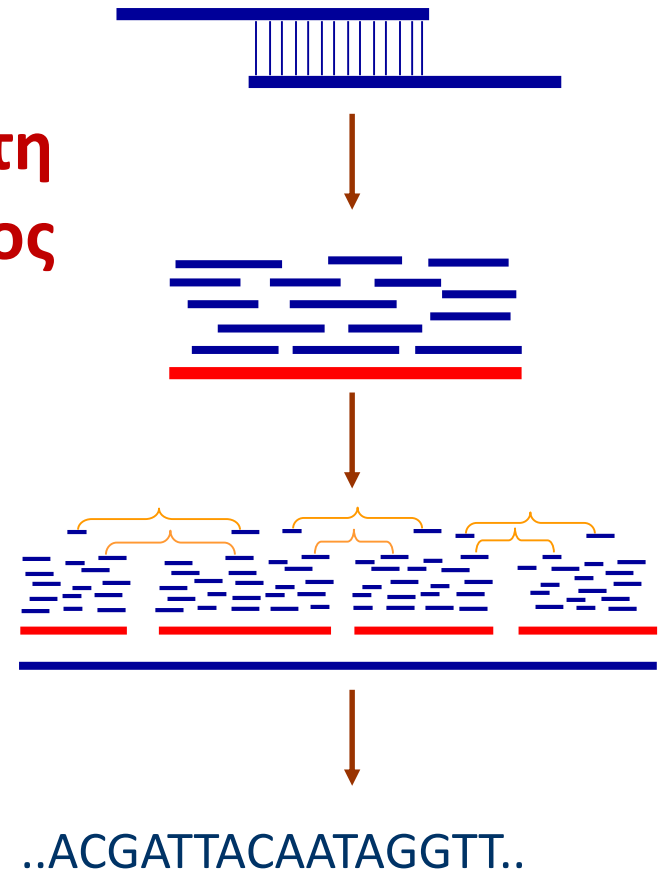


# Υπολογιστικά εργαλεία

Υπολογιστικά προγράμματα για τη συναρμολόγηση του γονιδιώματος

PHRED, PHRAP, LUCY,  
TGICL, CAP3, GAP4

<http://seqanswers.com/wiki/Software/list>,  
<http://www.jcvi.org/cms/research/software/>



# Χρωμοσωματικοί Χάρτες (1/2)

Δείχνουν τη θέση των γονιδίων, των γενετικών δεικτών, των κεντρομερών, των τελομερών και άλλων θέσεων κατά μήκος των χρωμοσωμάτων.

Συμπεριλαμβάνουν τους γενετικούς χάρτες (χάρτες σύνδεσης), φυσικούς χάρτες και χάρτες αλληλουχίας **σε ευρεία κλίμακα**.

Η διαδικασία ανάλυσης του γονιδιώματος περιλαμβάνει τρία στάδια: Πρώτα τη δημιουργία ενός χάρτη σύνδεσης. Μετά τη δημιουργία ενός φυσικού χάρτη. Τέλος, τη δημιουργία του χάρτη της αλληλουχίας από το συνδυασμό των παραπάνω.



# Χρωμοσωματικοί Χάρτες (2/2)

**Γενετικοί χάρτες (χάρτες σύνδεσης)** – δείχνουν τη σχετική θέση των γονιδίων χωρίς να είναι γνωστή η πραγματική απόσταση. Βασίζονται στη συχνότητα ανασυνδυασμού μεταξύ γειτονικών γονιδίων/γενετικών δεικτών.

**Φυσικοί χάρτες** – δείχνουν την πραγματική απόσταση μεταξύ των γονιδίων σε bp.

**Χάρτες αλληλουχίας** – αποτελούνται από τη συνολική αλληλουχία των νουκλεοτιδίων σε ένα γονίδιο ή γενικά περιοχή του DNA.

Δείτε εδώ χάρτες διάφορους τύπους χαρτών στο κριθάρι:

<http://pgrc.ipk-gatersleben.de/kuenzel/barleymap.html>



# Γενετικοί Χάρτες Υψηλής Ανάλυσης (1/8)

- Δημιουργούνται με βάση τη συχνότητα ανασυνδυασμού
- Απόσταση 1 centiMorgan (cM) αντιστοιχεί σε απόσταση που δίνει ποσοστό ανασυνδυασμού 1%
- Στους απλούς χάρτες σύνδεσης οι γενετιστές χαρτογραφούν τη θέση μόνο ενός μικρού αριθμού γονιδίων που βρίσκονται αρκετά κοντά μεταξύ τους, μέσω απλών διασταυρώσεων, όπου εξετάζονται κάποιοι εύκολα αναγνωρίσιμοι μορφολογικοί ή βιοχημικοί φαινότυποι
- Στους χάρτες υψηλής ανάλυσης όμως δεν επαρκούν οι γνωστοί φαινότυποι για να μπορέσουμε να βρούμε γονίδια που είναι πυκνά τοποθετημένα κατά μήκος των χρωμοσωμάτων
- Επίσης, μπορεί να προκύψουν προβλήματα με φαινότυπους, που δεν καθορίζονται μονογονιδιακά
- Οι ελεγχόμενες διασταυρώσεις δύο και τριών σημείων δεν είναι πάντα δυνατές λόγω ηθικών διλημάτων, ειδικά στον άνθρωπο
- Για αυτό είναι απαραίτητη η χρήση των **ΓΕΝΕΤΙΚΩΝ ΔΕΙΚΤΩΝ** που ελέγχονται για το μοριακό φαινότυπο τους δηλ. το γενότυπο τους



# Γενετικοί Χάρτες Υψηλής Ανάλυσης (2/8)

## ΓΕΝΕΤΙΚΟΙ ΔΕΙΚΤΕΣ

- Δεν έχουν απαραίτητα κάποιο λειτουργικό ρόλο
- Δεν μεταγράφονται (δεν είναι δηλαδή γονίδια, πάντα)
- Παρουσιάζουν υψηλό πολυμορφισμό (πολλά αλληλόμορφα)
- Υπάρχουν εκατομμύρια τέτοιοι DNA πολυμορφισμοί διεσπαρμένοι σε όλα τα χρωμοσώματα

Δύο είναι οι κύριοι γενετικοί δείκτες που χρησιμοποιούνται στη χαρτογράφηση υψηλής ανάλυσης

- Το **μικροδορυφορικό DNA** (microsatellite DNA, Simple Sequence Repeats-SSRs)
- Οι **απλοί νουκλεοτιδικοί πολυμορφισμοί** (Single Nucleotide Polymorphisms-SNPs)





# Γενετικοί Χάρτες Υψηλής Ανάλυσης (3/8)

## Μικροδορυφορικό DNA-SSRs

Αποτελείται από πολλαπλές (4-50) επαναλήψεις 2-6 νουκλεοτιδίων (π.χ. AGAGAGAG... ή GCAGCAGCAGCA...)

Παρουσιάζουν πολλά αλληλόμορφα, επειδή κατά την αντιγραφή του DNA γίνονται λάθη και χάνεται ή προστίθεται μια επανάληψη (π.χ. AG ή GCA αντίστοιχα)

Τα αλληλόμορφα καθορίζονται με βάση τον αριθμό των επαναλήψεων, ύστερα από ηλεκτροφόρηση

Συνήθως συναντάται 1 τόπος /2-30 Kb στο γονιδίωμα



# Γενετικοί Χάρτες Υψηλής Ανάλυσης (4/8)

## Απλοί νουκλεοτιδικοί πολυμορφισμοί-SNPs

- Είναι απλές μεταλλάξεις σημείου
- Χρησιμοποιούνται τα τελευταία χρόνια
- Συχνότητα 1/100 έως 1/300 βάσεις
- Γενοτύπηση κοστίζει 20-30 λεπτά (αυτόματα μηχανήματα με βάση GEL ή PCR)



# Γενετικοί Χάρτες Υψηλής Ανάλυσης (5/8)

## Μικροδορυφορικό DNA:

TGAACCGTTACTCG**TATATATATA**GCGTATGCT



TGAACCGTTACTCG**TATATATATATATATA**GCGTATGCT

Ο πολυμορφισμός προκύπτει από το διαφορετικό αριθμό επαναλήψεων του δινουκλεοτίδιου TA

## Απλοί νουκλεοτιδικοί πολυμορφισμοί

CGTTACT**T**GGTAACG



CGTTACT**G**GGTAACG

Πολυμορφισμός: όπου υπάρχει απλή αντικατάσταση ενός από τα τέσσερα νουκλεοτίδια



# Γενετικοί Χάρτες Υψηλής Ανάλυσης (6/8)

## Γενοτύπηση SNPs

Έχουν αναπτυχθεί πολλές διαφορετικοί μέθοδοι για τη γενοτύπηση των Απλών Νουκλεοτιδικών Πολυμορφισμών, κάποιες από τις οποίες αναφέρονται παρακάτω:

- ◆ Μέθοδοι υβριδισμού (μικροσυστοιχίες, TaqMan, Molecular Beacons)
- ◆ Αλληλομορφο-ειδική PCR
- ◆ Ανάλυση με ενζυμα περιορισμού

☞ Κάθε μέθοδος διαθέτει τόσο **πλεονεκτήματα** όσο και **μειονέκτηματα**. Έτσι κάθε φορά είναι στο χέρι του ερευνητή να επιλέξει την κατάλληλη μέθοδο για την επίτευξη των στόχων του.

<http://www.nature.com/nrg/journal/v2/n12/full/nrg1201-930a.html>



# Γενετικοί Χάρτες Υψηλής Ανάλυσης (7/8)

## Laboratory Methods for High-Throughput Genotyping

Howard J. Edenberg and Yunlong Liu

*Genetics of Complex Human Diseases: A Laboratory Manual* (eds. Al-Chalabi and Almasy). CSHL Press, Cold Spring Harbor, NY, USA, 2009.

<http://cshprotocols.cshlp.org/content/2009/11/pdb.top62.long>:

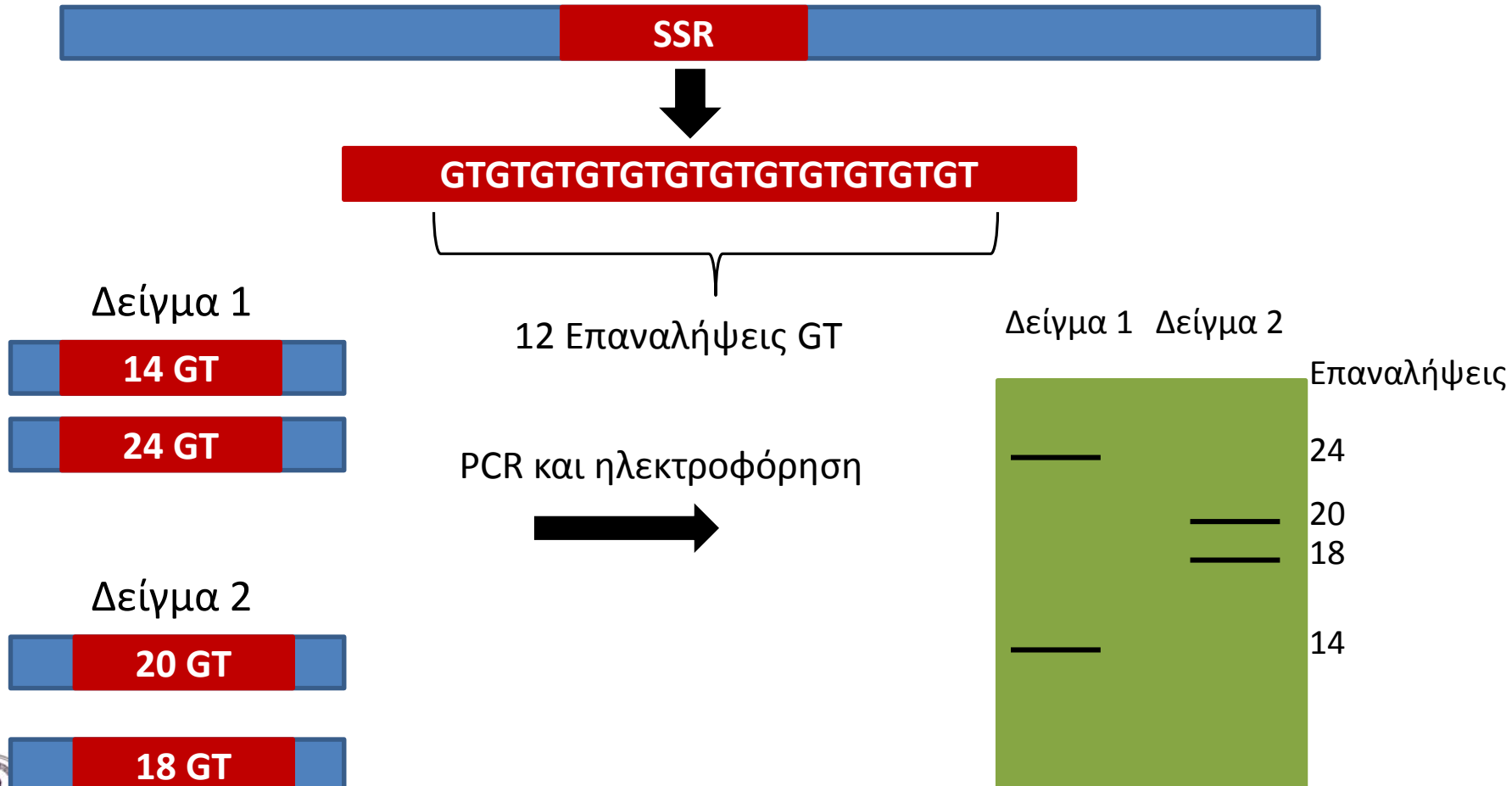
Figure 1: Διαφορετικές μέθοδοι (Σειριακές/παράλληλες) γενοτύπησης SNPs για τους διάφορους τύπους ερευνητικών προγραμμάτων

Figure 2: Παραδείγματα αποτελεσμάτων αυτόματης γενοτύπησης SNPs με το μηχάνημα MassArray



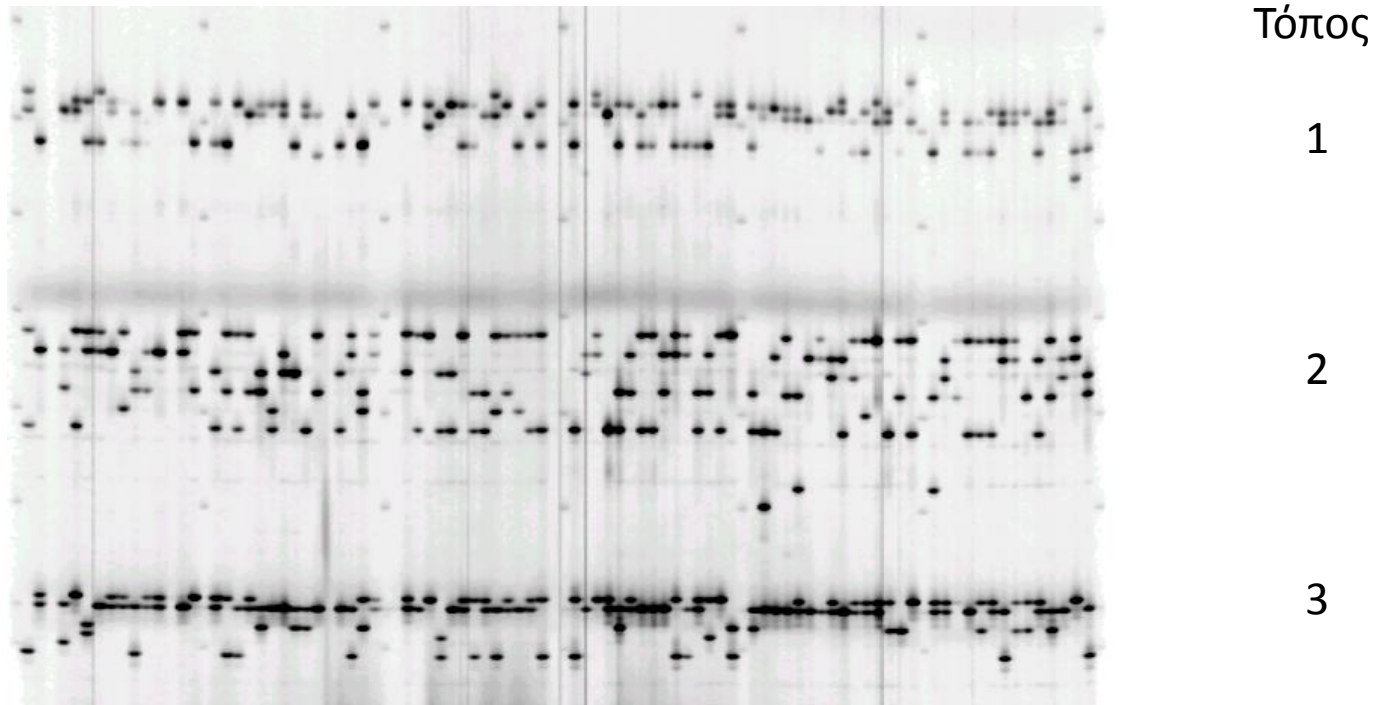
# Γενετικοί Χάρτες Υψηλής Ανάλυσης (8/8)

## Γενοτύπηση SSRs



# Γενετική Χαρτογράφηση με SSRs (1/3)

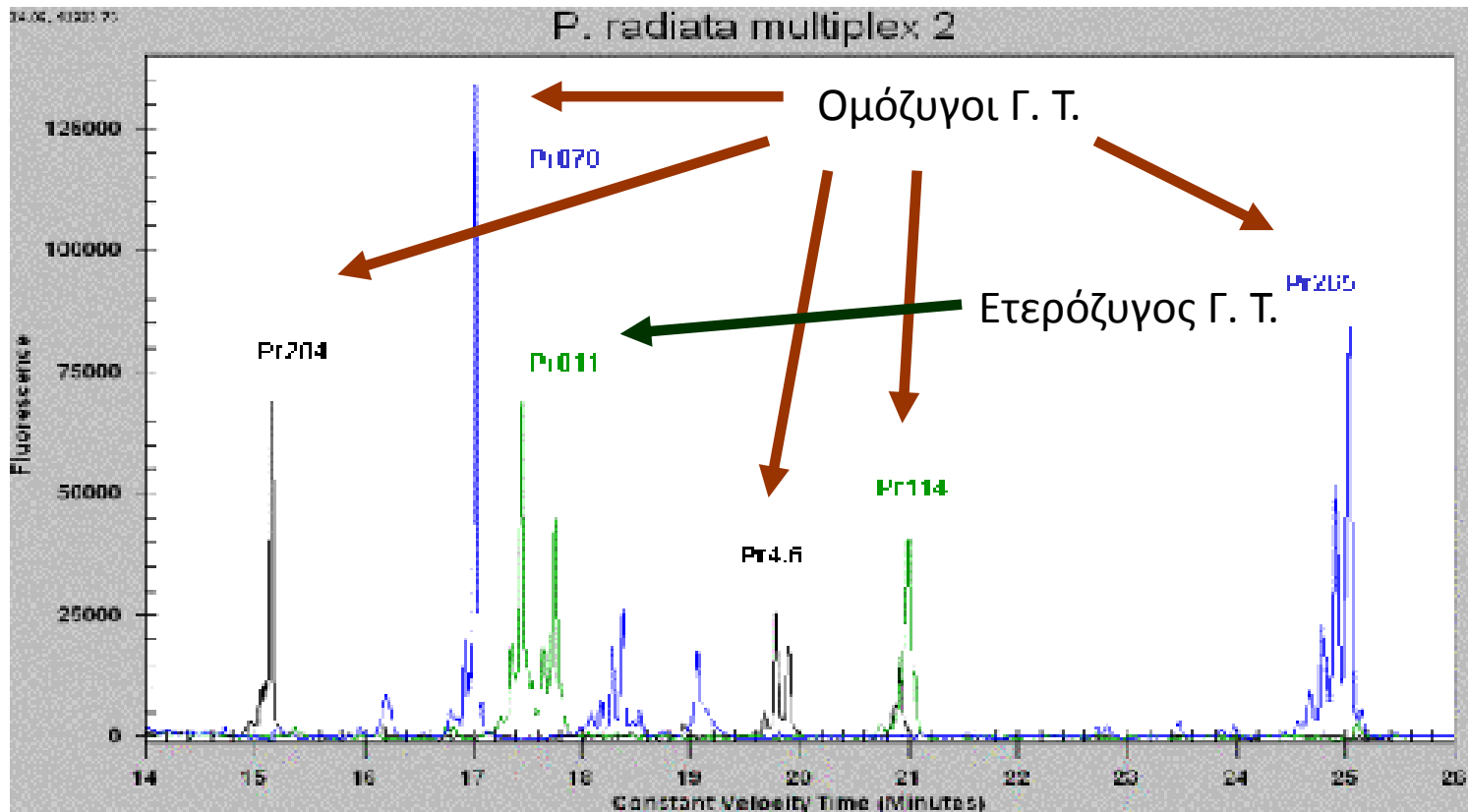
**Εικόνα 2:** Αποτελέσματα ηλεκτροφόρησης τριών τόπων SSR σε πηκτή πολύ-ακρυλαμίδης. Σε κάθε στήλη έχουν ηλεκτροφορηθεί τα αποτελέσματα από PCR σε ένα άτομο για τρεις τόπους. Υπάρχουν σχεδόν 100 άτομα σε αυτή την πηκτή (δηλ 100 άτομα X 3 τόπους = 300 γενότυποι)



# Γενετική Χαρτογράφηση με SSRs (2/3)

Πολύ πιο γρήγορη διαδικασία είναι η γενοτύπηση εκατοντάδων ως χιλιάδων SSRs με τη βοήθεια αυτόματων μηχανήματων.

**Εικόνα 3:** Το προφίλ ενός ατόμου σε έξι τόπους SSR

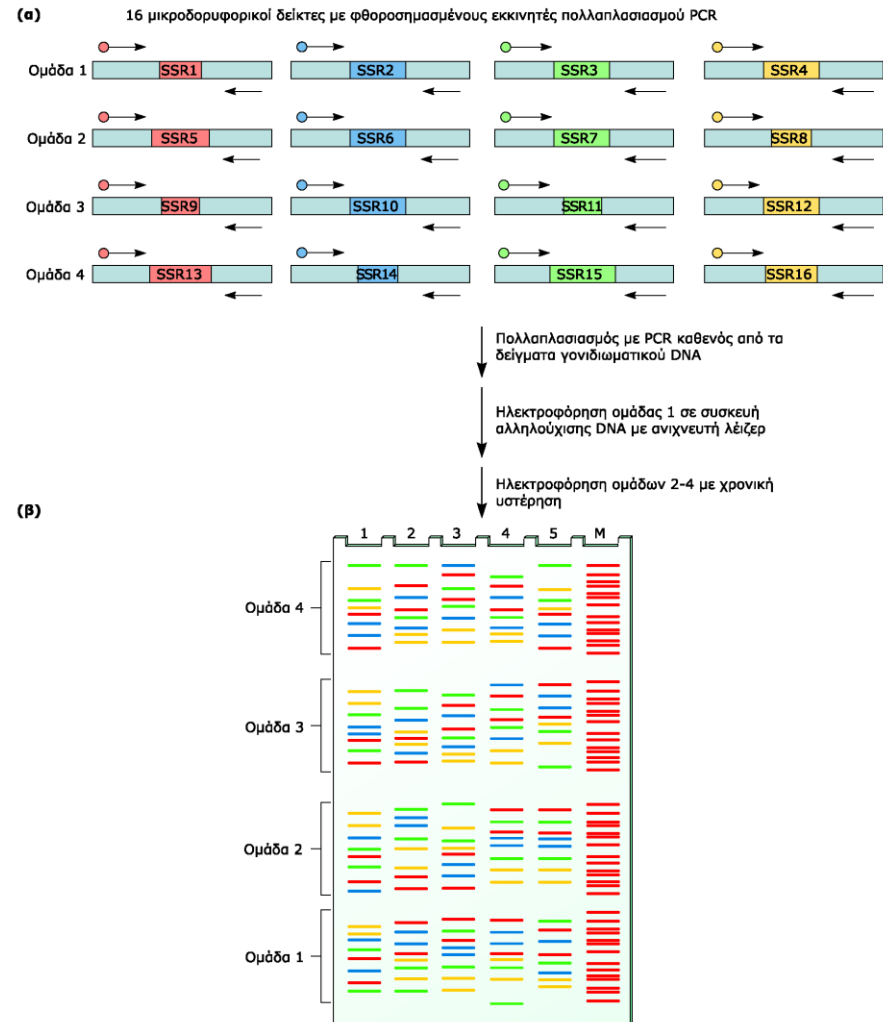




# Γενετική Χαρτογράφηση με SSRs (3/3)

**Εικόνα 4:** Γενοτύπηση εκατοντάδων ως χιλιάδων SSRs.

Στα αυτόματα μηχανήματα μπορούν να χρησιμοποιηθούν εκκινητές επισημασμένοι με διαφορετικές φθορίζουσες χρωστικές. Ακόμα και αν το μέγεθος του προϊόντος PCR για 2 διαφορετικούς τύπους SSR είναι το ίδιο, το σήμα σε διαφορετικό μήκος κύματος βοηθάει ώστε να διαφοροποιηθούν.



# Γενετικοί Χάρτες Υψηλής Ανάλυσης

- Είναι εύκολη η ανάλυση σύνδεσης σήμερα, καθώς υπάρχουν πλήθος δεδομένων σε ειδικές βάσεις και ιστοσελίδες για γενετικούς δείκτες
- Στον άνθρωπο και στο ποντίκι υπάρχουν πάνω από 150.000 και 50.000 SSRs και πάνω από 40.000.000 SNPs (230 εκ. και 70 εκ. SNPs αντίστοιχα μη ελεγμένα).
- Οι ερευνητές χρησιμοποιούν περίπου 500 SSR για βρουν τη σχετική θέση ενός γονιδίου, δηλαδή περίπου 1 SSR / 6 Mb.
- Σήμερα με τη χρήση των SNPs υπάρχουν τουλάχιστον χιλιάδες τέτοιοι γενετικοί δείκτες, δηλαδή έχουμε μια αύξηση χίλιες φορές στη δυνατότητα να χαρτογραφήσουμε ένα γονίδιο

<http://hapmap.ncbi.nlm.nih.gov/>

[http://www.ncbi.nlm.nih.gov/SNP/snp\\_summary.cgi](http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi)



# Άσκηση 1 – Λύση

• Δώστε το γενότυπο τριών ατόμων σε δύο μικροδορυφορικούς (SSR) τύπους DNA και σε δύο τύπους SNP. Τα άτομα αυτά πρέπει να έχουν διαφορετικούς γενοτύπους.

Σημειώστε τα αλληλόμορφα που παρατηρείτε.

Αλληλόμορφα Τύπου SNP1 : A, T

Αλληλόμορφα Τύπου SNP2 : A, C

Αλληλόμορφα Τύπου SSR1 :  $(GC)_3$ ,  $(GC)_4$ ,  $(GC)_5$ ,  $(GC)_7$ ,  $(GC)_{10}$

Αλληλόμορφα Τύπου SSR2 :  $(AAT)_2$ ,  $(AAT)_3$ ,  $(AAT)_4$ ,  $(AAT)_7$

|                | Τόπος SNP1 | Τόπος SNP2 | Τόπος SSR1        | Τόπος SSR2       |
|----------------|------------|------------|-------------------|------------------|
| <b>Άτομο 1</b> | AT         | AA         | $(GC)_7(GC)_4$    | $(AAT)_2(AAT)_4$ |
| <b>Άτομο 2</b> | AA         | CC         | $(GC)_{10}(GC)_3$ | $(AAT)_3(AAT)_2$ |
| <b>Άτομο 3</b> | TT         | AC         | $(GC)_5(GC)_3$    | $(AAT)_4(AAT)_7$ |

Πού υπάρχει υψηλότερος πολυμορφισμός σε τύπους SNP ή SSR;

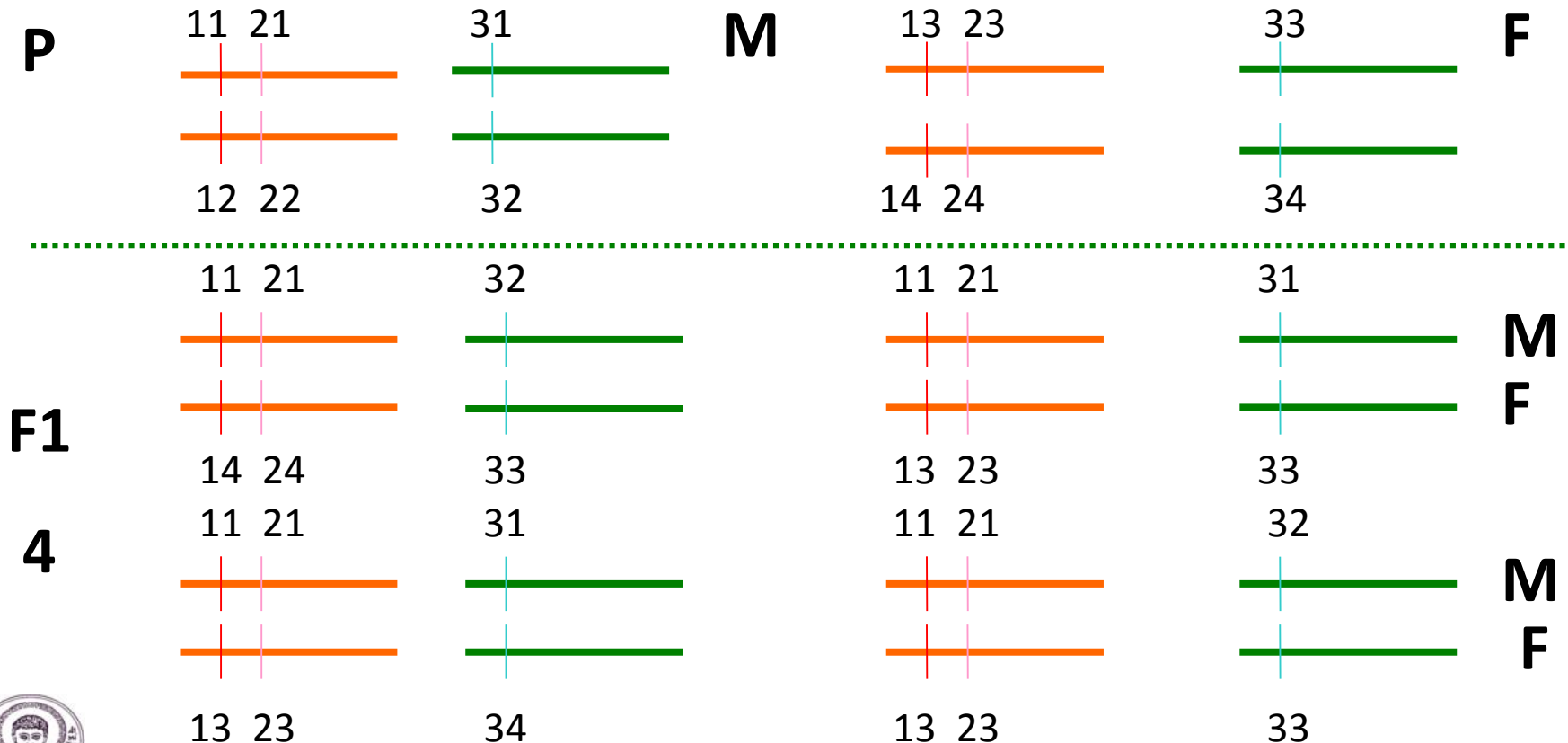
Υψηλότερος πολυμορφισμός παρατηρείται σε τύπους SSR, καθώς έχουν περισσότερα αλληλόμορφα

\* Τα SNP είναι συνήθως γενετικοί τόποι με δύο αλληλόμορφα



# Γενετική Χαρτογράφηση με SSRs

Το επόμενο βήμα είναι η παρακολούθηση διαχωρισμού πολύ μεγάλου αριθμού γενετικών δεικτών και συν-κληρονόμησης τους σε γενεαλογικά δένδρα. Αλληλόμορφα διαφορετικών τόπων που εμφανίζονται μαζί σε γενεαλογικά δένδρα, αποκαλύπτουν την ύπαρξη σύνδεσης των τόπων (π.χ. το 1.1 και το 2.1 παρακάτω)



# Γενετικοί Χάρτες Υψηλής Ανάλυσης

Έτσι, σχηματίζεται χάρτης σύνδεσης υψηλής ανάλυσης που περιλαμβάνει τη σχετική θέση χιλιάδων, πολύ κοντινών γενετικών δεικτών.

[http://www.funpecrp.com.br/gmr/year2007/vol3-6/gmr0340\\_full\\_text.html](http://www.funpecrp.com.br/gmr/year2007/vol3-6/gmr0340_full_text.html)

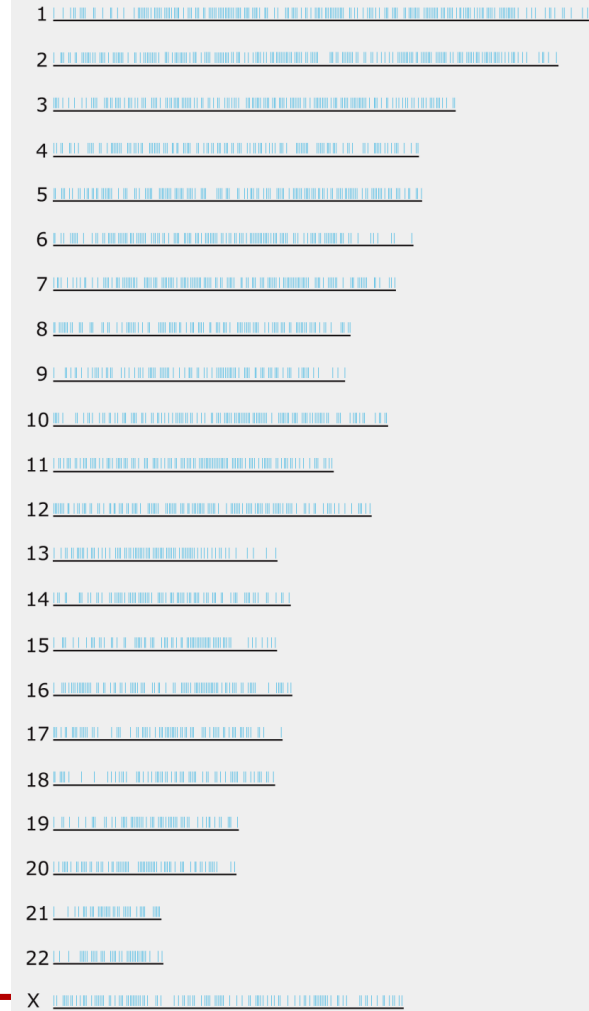


# Γενετικοί Χάρτες Υψηλής Ανάλυσης

Για τον άνθρωπο χρησιμοποιήθηκαν ~5000 δείκτες σε 500 άτομα από 40 οικογένειες 3 γενεών καθεμία

$500 \times 5000 = 2.500.000$  ΓΕΝΟΤΥΠΗΣΕΙΣ!!!

**Εικόνα 5:** Γενετικός χάρτης γονιδιώματος ανθρώπου υψηλής πυκνότητας με 5.264 μικροδορυφόρους (Dib *et al.* 1996 *Nature* 380: 152-154)



# Άσκηση 2

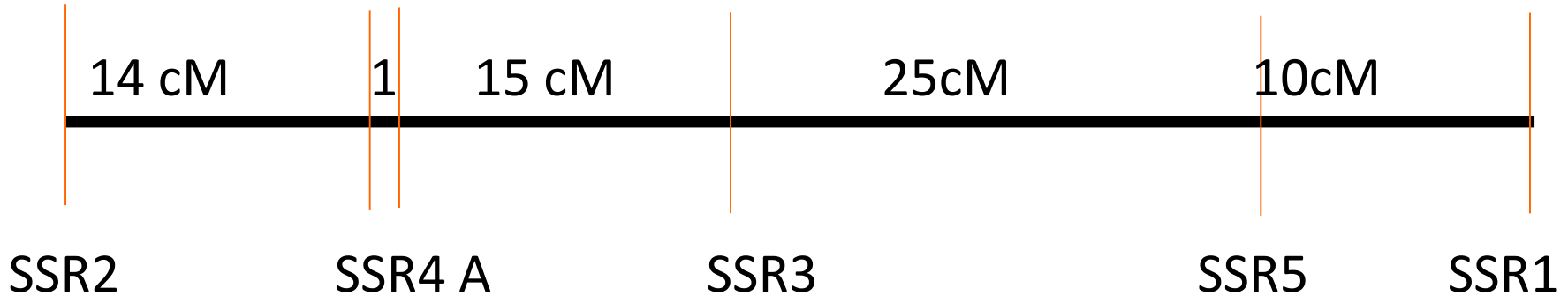
Γίνεται έλεγχος σε κλώνους μιας βιβλιοθήκης για να βρεθεί η ακριβής θέση ενός γονιδίου A. Με τη βοήθεια SSR δεικτών υπολογίστηκαν οι παρακάτω συχνότητες ανασυνδυασμού:

| Τόποι   | % Ανα/σμού | Τόποι      | % Ανα/σμού | Τόποι      | % Ανα/σμού |
|---------|------------|------------|------------|------------|------------|
| A, SSR1 | 50         | SSR1, SSR2 | 50         | SSR2, SSR4 | 14         |
| A, SSR2 | 15         | SSR1, SSR3 | 35         | SSR2, SSR5 | 50         |
| A, SSR3 | 15         | SSR1, SSR4 | 50         | SSR3, SSR4 | 16         |
| A, SSR4 | 1          | SSR1, SSR5 | 10         | SSR3, SSR5 | 25         |
| A, SSR5 | 40         | SSR2, SSR3 | 30         | SSR4, SSR5 | 41         |

Σχεδιάστε ένα γενετικό χάρτη για τους δείκτες SSR και το γονίδιο A.



# Άσκηση 2 – Λύση



Αν το ανθρώπινο γονιδίωμα είναι 3,3 Gb και ο γενετικός χάρτης του ανθρώπου περιλαμβάνει 3300 cM τι μέγεθος κομματιού περιλαμβάνεται στον κλώνο αυτό? ( $65 \text{ cm} = 65 \text{ Mb}$ )

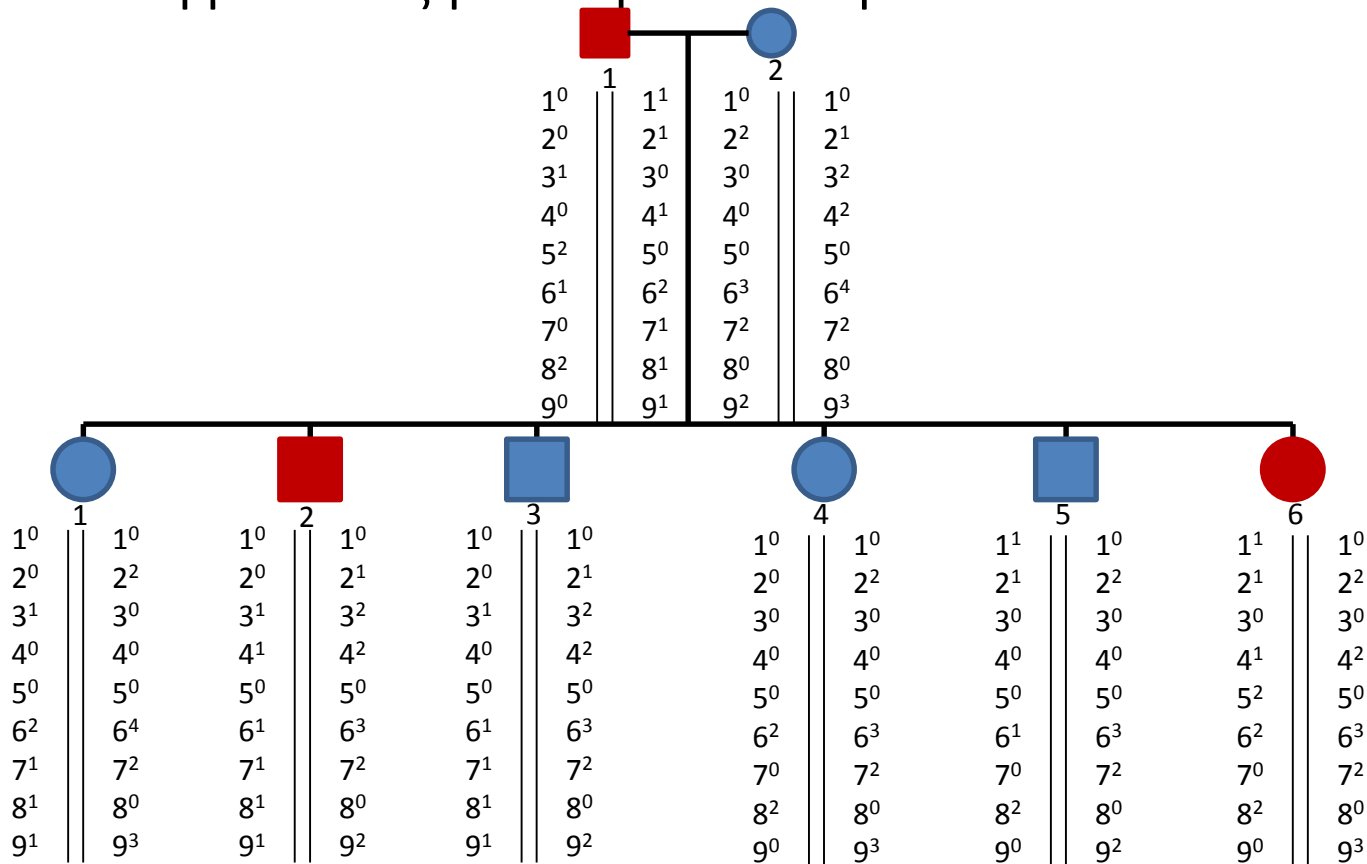
Ποια είναι η απόσταση σε bp μεταξύ του γονιδίου A και του πλησιέστερου SSR? ( $1 \text{ cm} = 1 \text{ Mb}$ )





# Άσκηση 3

Η ασθένεια Best σχετίζεται με σταδιακή τύφλωση. Είναι αυτοσωμική ασθένεια και εδράζεται στο χρωμόσωμα 11. Χρησιμοποιήθηκαν 9 SSR δείκτες σε προσπάθεια να βρεθεί η ακριβής θέση του γονιδίου με βάση τη μελέτη σε γενεαλογικό δένδρο. Ποιος από τους 9 δείκτες βρίσκεται πιο κοντά στη μετάλλαξη που προκαλεί την ασθένεια?



# Άσκηση 3 - Λύση

---

Το αλληλόμορφο  $4^0$  είναι αυτό που εμφανίζεται στα άτομα που πάσχουν και μόνο σε αυτά



# Φυσικοί Χάρτες Ευρείας Κλίμακας

Ένας **Φυσικός Χάρτης** αποτελείται από τη συνένωση αλληλοεπικαλυπτόμενων τμημάτων DNA από ενθέματα βιβλιοθηκών που τοποθετούνται με συγκεκριμένη κατεύθυνση πάνω σε καθένα από τα χρωμοσώματα του γονιδιώματος. Δηλαδή για να σχηματιστεί ο φυσικός χάρτης ενός χρωμοσώματος πρέπει να μπου οι κλώνοι που έχουν προκύψει από αυτό το χρωμόσωμα σε μια σειρά. Η διαδικασία απαιτεί τη χρήση υπολογιστών.

Με βάση τους φυσικούς χάρτες υπολογίζεται ακριβώς ο αριθμός των ζευγών βάσεων (bp, Kb, Mb) που απέχει ένας γονιδιακός τόπος ή μια θέση DNA από τα γειτονικά της στοιχεία σε ένα συγκεκριμένο χρωμόσωμα.

Για κάθε οργανισμό μπορεί να υπολογιστεί η συσχέτιση φυσικού χάρτη και χάρτη σύνδεσης. Στον άνθρωπο η αντιστοιχία είναι  $1 \text{ cM} \sim 1 \text{ Mb}$  και στο ποντίκι είναι  $1 \text{ cM} \sim 2 \text{ Mb}$ .

Meyers et al 2004 Nat. Rev. Genet 5. 578-588



# Φυσικοί Χάρτες Ευρείας Κλίμακας

## Φορείς Κλωνοποίησης για Βιβλιοθήκες

Τα ενθέματα πρέπει να είναι όσο το δυνατόν μεγαλύτερα και σταθερά. Δεν πρέπει να σπάζουν ή να ανασυνδυάζονται.

Μέχρι τώρα έχουν χρησιμοποιηθεί τα:

- Τεχνητά χρωμοσώματα ζύμης (yeast artificial chromosomes YACs): Δέχονται μεγάλα ενθέματα (έως 1 Mb), αλλά είναι ασταθή και δύσκολα στο χειρισμό.
- Τεχνητά χρωμοσώματα βακτηρίων (bacterial artificial chromosomes BACs). Δέχονται ένθεμα από 50-300 Kb και είναι πιο σταθερά συνεπώς χρησιμοποιήθηκαν περισσότερο.

Πόσα BACs (μέσου μήκους 200 Kb) χρειάζονται για να περιλαμβάνουν ολόκληρο το ανθρώπινο γονιδίωμα;

$$3 \times 10^9 \text{ bp το γονιδίωμα} / 200 \text{ Kb BAC} = 15.000 \text{ BACs.}$$



# Φυσικοί Χάρτες Ευρείας Κλίμακας

|          | Φορέας  | Μέγεθος τμημάτων μερικής πηγής του γονιδιοματικού DNA | Κλώνος μεγάλου μήκους                                      | Αριθμός αντιγράφων        | Αριθμός κλώνων που απαιτείται για κάλυψη του ανθρώπινου γονιδιώματος 1x |
|----------|---|---|--|---------------------------|---|
| Κοσμίδιο | <p>Έναρξη αντιγραφής πλασμίδιου πολλαπλών αντιγράφων</p> <p>Θέση COS</p> <p>Πολυσυνδέτης</p> <p>Amp<sup>R</sup></p>   | 35-45 kb  | <p>Η ένθεση ανθρώπινου DNA αποτελεί ~80% του κοσμίδιου</p> | 50-100                    | ~75.000   |
| Φοσμίδιο | <p>Έναρξη αντιγραφής παράγοντα F</p> <p>Θέση COS</p> <p>Πολυσυνδέτης</p> <p>Cam<sup>R</sup></p> <p>Amp<sup>R</sup></p>  | 35-45 kb  | <p>Η ένθεση ανθρώπινου DNA αποτελεί ~80% του φοσμίδιου</p> | Ένα αντίγραφο ανά κύτταρο | ~75.000   |
| BAC      | <p>Έναρξη αντιγραφής παράγοντα F</p> <p>Πολυσυνδέτης</p> <p>Cam<sup>R</sup></p> <p>Amp<sup>R</sup></p>  | 100-200 kb  | <p>Η ένθεση ανθρώπινου DNA αποτελεί ~90% του BAC</p>       | Ένα αντίγραφο ανά κύτταρο | 15.000-30.000   |
| PAC      | <p>Έναρξη αντιγραφής φάγου P1</p> <p>Πολυσυνδέτης</p> <p>Kan<sup>R</sup></p> <p>Amp<sup>R</sup></p> <p>Iox</p>  | 100-200 kb  | <p>Η ένθεση ανθρώπινου DNA αποτελεί ~90% του PAC</p>       | Ένα αντίγραφο ανά κύτταρο | 15.000-30.000   |
| YAC      | <p>Έναρξη αντιγραφής</p> <p>Κεντρομερές</p> <p>Επιλέξιμος δείκτης 2</p> <p>Θέση κλωνοποίησης</p> <p>Αmp<sup>R</sup></p> <p>Τελομερές</p> <p>Θέση κοπής</p> <p>Τελομερές</p> | ~200-1.000 kb   | <p>Ένθεση ανθρώπινου DNA</p> <p>Tel M1 Ori Cen M2 Tel</p>  | Ένα αντίγραφο ανά κύτταρο | ~3.000-15.000   |

**Εικόνα 6: Φορείς κλωνοποίησης τμημάτων μεγάλου μήκους: Τα συστήματα αυτά είναι ουσιώδους σημασίας για τη μελέτη του γονιδιώματος**



# Φυσικοί Χάρτες Ευρείας Κλίμακας

| Χάρτης                 | Αριθμός BACs που συναρμολογήθηκαν | Κάλυψη των BACs |
|------------------------|-----------------------------------|-----------------|
| Άνθρωπος               | 283.287                           | 15x             |
| Ποντίκι                | 305.716                           | 33x             |
| <i>D. melanogaster</i> | 10.253                            | 14x             |
| <i>A. thaliana I</i>   | 20.206                            | 17x             |
| <i>A. thaliana II</i>  | 9.389                             | 7,2x            |
| Ρύζι I                 | 21.087                            | 6,9x            |
| Ρύζι II                | 65.287                            | 20x             |
| Σόγια                  | 78.001                            | 9,6x            |
| Αρουραίος              | 189.689                           | 13,1x           |
| Ζαχαρόχορτο            | 22.233                            | 4x              |
| <i>B. japonicum</i>    | 4.608                             | 77x             |

## Φορείς Κλωνοποίησης για Βιβλιοθήκες

Αριθμός BACs που χρειάστηκαν για τη δημιουργία φυσικών χαρτών σε διάφορα είδη

Meyers et al 2004

Nat. Rev. Genet 5. 578-588



# Φυσικοί Χάρτες Ευρείας Κλίμακας

ΣΚΟΠΟΣ → να βρεθεί η σχετική θέση των ενθεμάτων των κλώνων πάνω στα χρωμοσώματα.

Μπορεί να επιτευχθεί με μια:

Τεχνική από πάνω προς τα κάτω → γίνεται υβριδισμός (FISH) με ολόκληρα τα ενθέματα επάνω στον καρυότυπο του οργανισμού

Τεχνική από κάτω προς τα πάνω → η σχετική θέση των κλώνων ταυτοποιείται με τη χρήση 1) **χαρτών σύνδεσης** ή 2) **ενζύμων περιορισμού** ή 3) αλληλουχικών θέσεων σημαίων-ετικέτες θέσεων αλληλουχίας (**Sequence Tagged Sites–STS**).

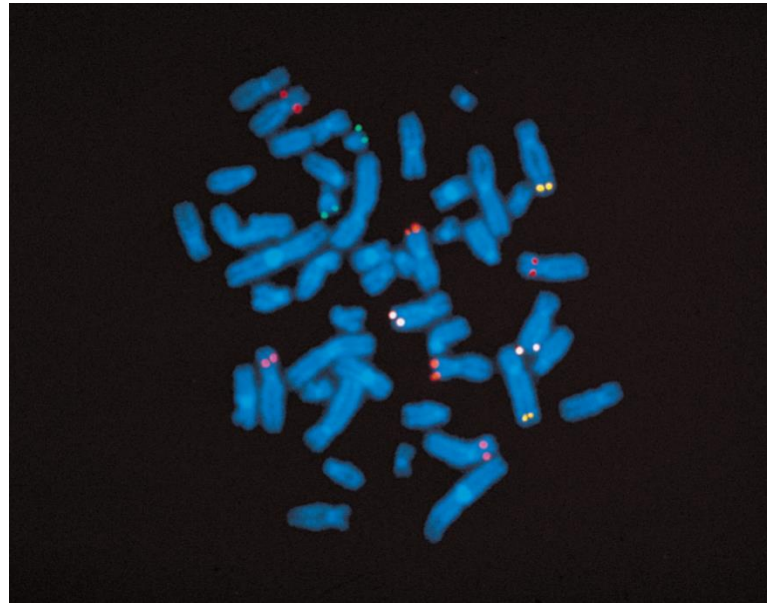


# Φυσικοί Χάρτες Ευρείας Κλίμακας ↓

## Τεχνική FISH (Fluorescence in situ hybridisation)

Βασίζεται στη χρήση ανιχνευτών που είναι επισημασμένοι με φθορίζουσες χρωστικές και χρησιμοποιούνται για υβριδισμό σε αποδιεταγμένα μεταφασικά χρωμοσώματα

Εύκολη  
χαρτογράφηση  
αλλά χαμηλής  
ανάλυσης



Εικόνα 7: Τεχνική FISH

<https://www.youtube.com/watch?v=nm8Ai1CI9Is>





# Φυσικοί Χάρτες Ευρείας Κλίμακας ↑

Για να σχηματιστούν οι φυσικοί χάρτες ευρείας κλίμακας χρησιμοποιούνται ποικίλες προσεγγίσεις:

- 1) Χάρτες σύνδεσης
- 2) Ένζυμα περιορισμού
- 3) Αλληλουχικές θέσεις σημαίες (STS)

☞ Οι ερευνητές συνήθως χρησιμοποιούν έναν συνδυασμό αυτών των μεθόδων



# Φυσικοί Χάρτες Ευρείας Κλίμακας ↑

## 1. Χάρτες Σύνδεσης

Χρησιμοποιείται η προϋπάρχουσα πληροφορία που αφορά τη σχετική θέση των γενετικών δεικτών πάνω στα χρωμοσώματα.

Απαιτούνται ομοιόμορφα κατανεμημένοι γενετικοί δείκτες με απόσταση μεταξύ τους όχι παραπάνω από 1 cM.

Οι γενετικοί δείκτες χρησιμοποιούνται ως ανιχνευτές υβριδισμού.



# Φυσικοί Χάρτες Ευρείας Κλίμακας ↑

## Περπάτημα στα χρωμοσώματα

- Αν 2 κοντινοί δείκτες M1 και M2, υβριδίζονται σε κλώνους YAC (μεγέθους 1 Mb, γM1 και γM2), πιθανώς οι κλώνοι επικαλύπτονται
- Χρησιμοποιούνται τα άκρα του ενός από τους δύο κλώνους, για να βρεθεί η σχετική θέση / επικάλυψη των κλώνων
- Με χρήση πολλών συνεχόμενων δεικτών M1, M2, M3, M4, M5, M6, M7 ... ενώνονται οι κλώνοι σε μία συνεχή σειρά
- Αν οι γενετικοί δείκτες δεν είναι ομοιόμορφα κατανεμημένοι ή ένας κλώνος έχει πολύ μεγάλο μήκος, οι κλώνοι δεν αποτελούν μια συνεχή σειρά
- Οι επικαλυπτόμενοι κλώνοι που συναρμολογούν ένα συνεχές κομμάτι γονιδιώματος αποτελούν ένα (συναρμολόγημα) **contig**
- Η ένωση των contigs δημιουργεί τα **scaffolds**  
<http://oregonstate.edu/dept/biochem/hhmi/hhmiclasses/bb451/figslett/FigBC.html>



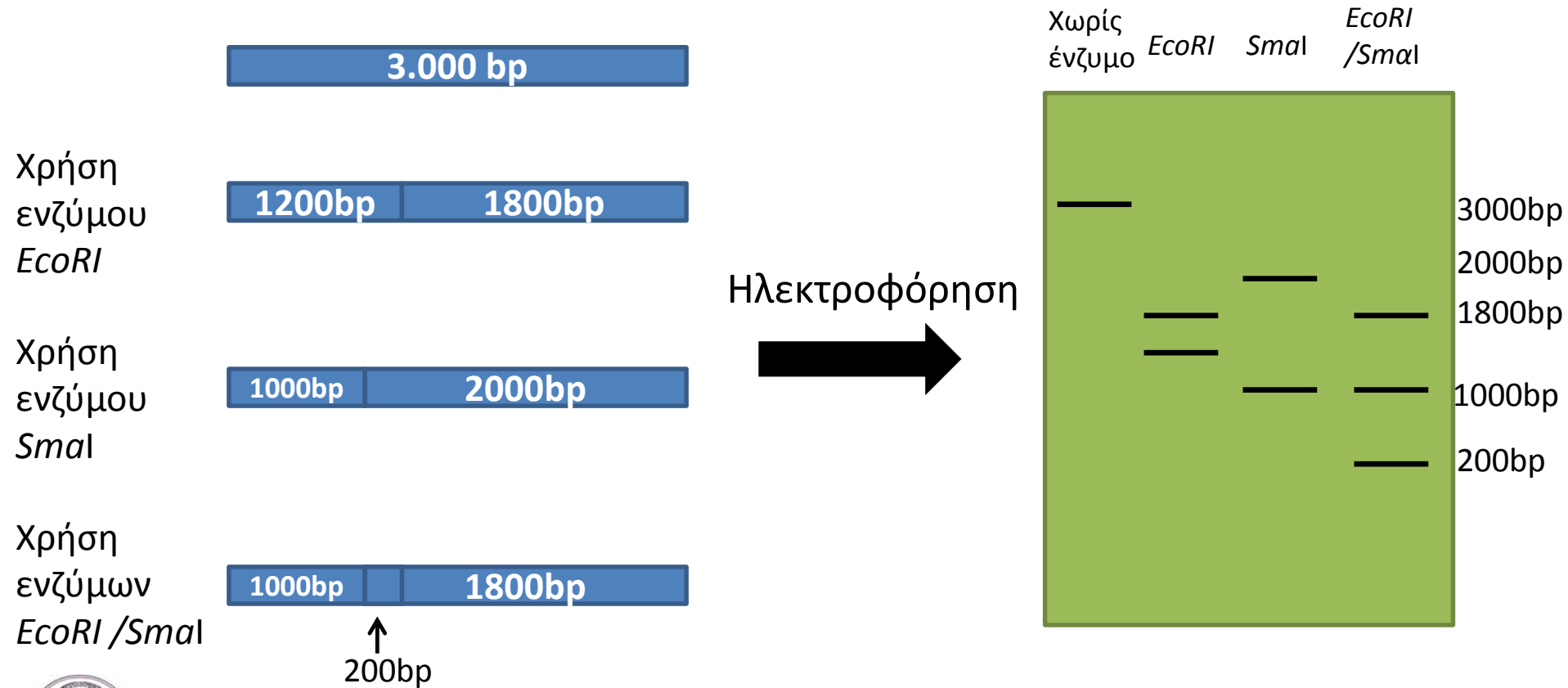
# Σύγκριση της μεθόδου FISH με τους χάρτες σύνδεσης για την κατασκευή φυσικών χαρτών

- Η FISH επιτρέπει τη χαρτογράφηση οποιουδήποτε κλώνου ανεξαρτήτως αν υπάρχει πολυμορφισμός, ενώ η ΑΣ απαιτεί πολυμορφισμό
- Η FISH χρειάζεται μια μικρή μόνο ποσότητα DNA του κλώνου, ενώ η ΑΣ απαιτεί ποσότητα πληροφοριών γενοτύπησης από πολλά άτομα
- Η FISH δίνει άμεσες και συγκρίσιμες πληροφορίες για τη σχετική θέση κλώνων στα χρωμοσώματα ακόμα και όταν αυτά έχουν τροποποιηθεί μέσω μετατοπίσεων
- Η FISH εφαρμόζεται σε μοναδικούς κλώνους, ενώ η ΑΣ συγκρίνει πάντα ένα κλώνο με έναν άλλο
- **Μειονέκτημα** → δεν μπορεί να δώσει ακριβή θέση κλώνων
- Συνήθως χρησιμοποιείται για να βρεθεί περίπου η θέση ενός κλώνου σε περιοχή 4-8 Mb ενός χρωμοσώματος
- Μετά η τεχνική ανάλυσης σύνδεσης επιτρέπει να φθάσουμε σε επίπεδο cM



# Φυσικοί Χάρτες Ευρείας Κλίμακας ↑

Η χαρτογράφηση με τη χρήση ενζύμων περιορισμού βασίζεται στην ικανότητα των ενζύμων περιορισμού να κόβουν τμήματα DNA σε συγκεκριμένες θέσεις



# Φυσικοί Χάρτες Ευρείας Κλίμακας ↑

## 2. Ένζυμα Περιορισμού

Τα χρωμοσώματα πακετάρονται σε βιβλιοθήκες από κοσμίδια.

Τα ενθέματα των κοσμιδίων πέπτονται με ένζυμα περιορισμού.

Δημιουργείται ένα χαρακτηριστικό **αποτύπωμα DNA** που αποτελείται από το σύνολο των τμημάτων συγκεκριμένου μεγέθους, που προκύπτουν από πέψη.

Εφόσον δύο κοσμίδια επικαλύπτονται σε ένα τμήμα τους, θα προκύψουν και κάποια τμήματα από τις πέψεις, που θα είναι κοινά στα αντίστοιχα αποτυπώματα.

Τα δεδομένα από τις πέψεις είναι δεκάδες χιλιάδες, άρα είναι απαραίτητη η χρήση ηλεκτρονικών υπολογιστών.

[http://www.nature.com/nrg/journal/v5/n8/fig\\_tab/nrg1404\\_F1.html](http://www.nature.com/nrg/journal/v5/n8/fig_tab/nrg1404_F1.html)



# Φυσικοί Χάρτες Ευρείας Κλίμακας ↑

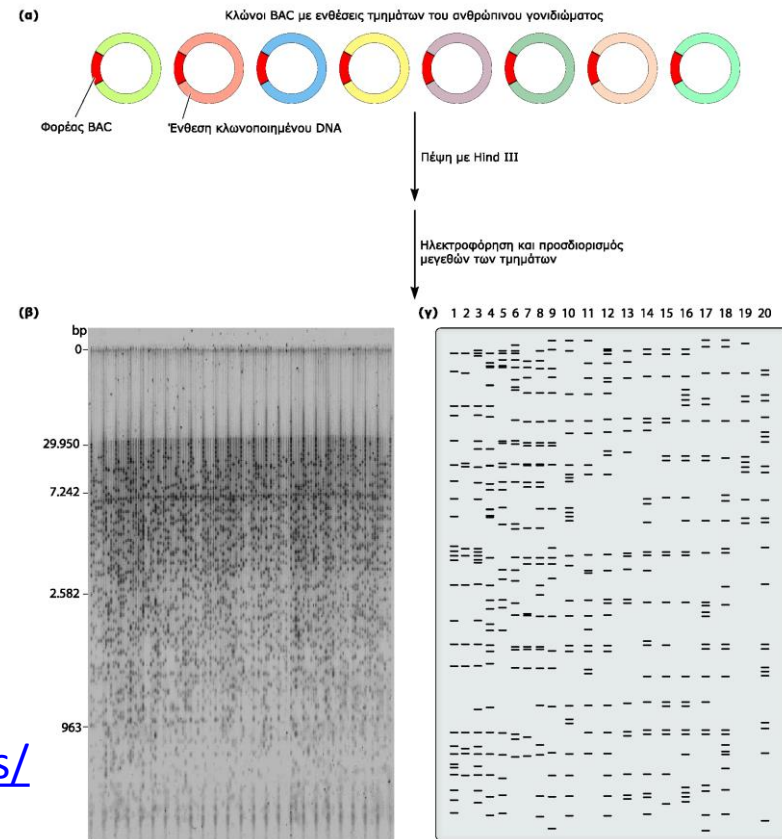
## 2. Ένζυμα Περιορισμού

Η ανάλυση αποτυπωμάτων χρησιμοποιείται για την χαρτογράφηση και τη συναρμολόγηση κλώνων τεχνητών χρωσωμάτων βακτηρίων (BAC).

**Εικόνα 8:** Αποτέλεσμα ηλεκτροφόρησης από πέψη BACs (αριστερά) και η επεξήγηση των πέψεων (δεξιά)

**Mapping DNA fingerprints**

<http://www.yourgenome.org/downloads/animations.shtml>



# Φυσικοί Χάρτες Ευρείας Κλίμακας ↑

## 2. Ένζυμα Περιορισμού

Θέσεις κοπής  
ενζύμων  
περιορισμού

Contig



Κλώνος 1



Κλώνος 2



Κλώνος 3



Κλώνος 4



Κλώνος 5



Κλώνος 6



Χάρτης  
συναρμολόγησης ενός  
contig μέσω  
επικαλυπτόμενων  
κλώνων (ως φορείς  
μπορούν να  
χρησιμοποιηθούν είτε  
YAC είτε BAC)

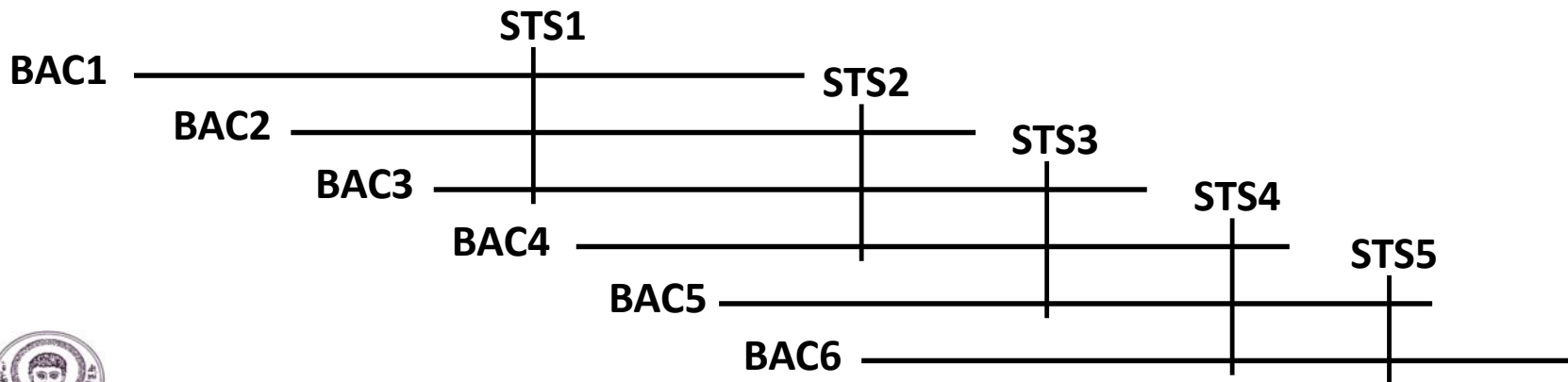




# Φυσικοί Χάρτες Ευρείας Κλίμακας ↑

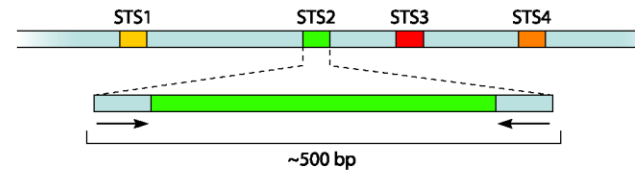
## 3. Αλληλουχικές θέσεις σημαίες (STS)

- Επιλέγεται μια αλληλουχία που είναι μοναδική και χαρακτηριστική για συγκεκριμένη περιοχή του χρωμοσώματος (**sequence tagged site, STS**). (π.χ. SSRs / SNPs).
- Δεν χρειάζονται δεδομένα από γενεαλογικά δένδρα
- Σχεδιάζονται εκκινητές για την ενίσχυση της STS μέσω PCR
- Όλες οι θέσεις STS έχουν μια συγκεκριμένη «διεύθυνση» πάνω στο χρωμόσωμα.
- Ελέγχονται όλοι οι κλώνοι BAC που αντιστοιχούν στο γονιδίωμα (~ 15.000-20.000) αν ενισχύουν αυτή τη θέση σημαία. Αν ναι, σημαίνει ότι οι κλώνοι είναι επικαλυπτόμενοι.
- Αν χρησιμοποιηθούν και άλλες τέτοιες θέσεις σημαίες, σιγά σιγά οι κλώνοι θα μπουν σε μια σειρά.

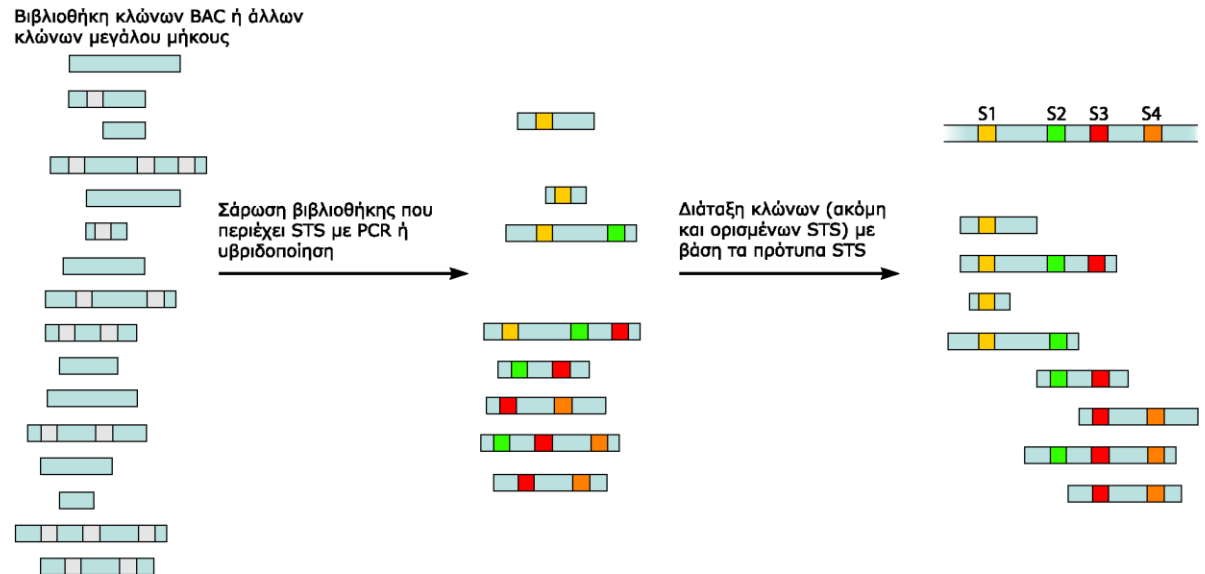


# Φυσικοί Χάρτες Ευρείας Κλίμακας ↑

## 3. Αλληλουχικές θέσεις σημαίες (STS)



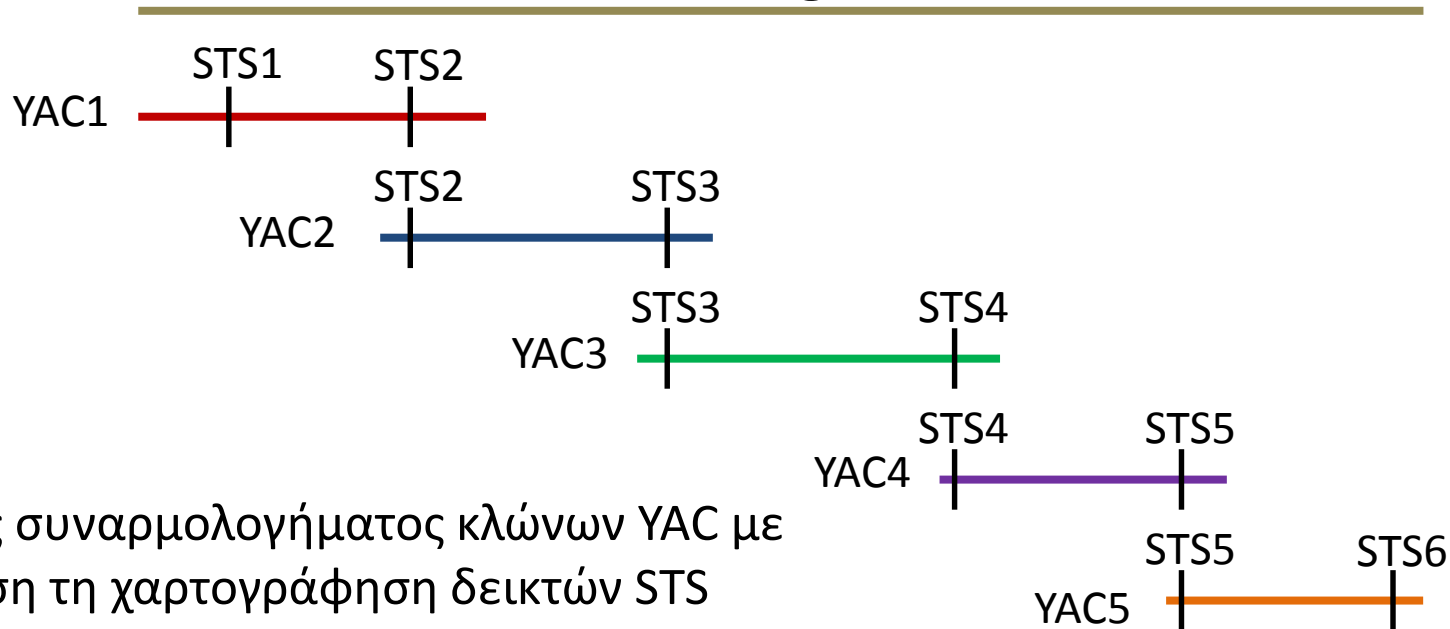
**Εικόνα 9:** Χαρτογράφηση με βάση την ανίχνευση STS σε κλώνους BAC



# Φυσικοί Χάρτες Ευρείας Κλίμακας ↑

## 3. Αλληλουχικές θέσεις σημαίες (STS)

### Contig



Χάρτης συναρμολογήματος κλώνων YAC με  
βάση τη χαρτογράφηση δεικτών STS

<http://genome.cshlp.org/content/7/7/673/F4.expansion.html>



# Άσκηση 4

Παρακάτω φαίνεται ένας χάρτης contig του χρωμοσώματος 3 στο *Arabidopsis*



Αν ένα EST υβριδίζεται με τους γονιδιωματικούς κλώνους C, D και E αλλά όχι με τους άλλους κλώνους σε ποιο τμήμα του χρωμοσώματος 3 βρίσκεται αυτό το EST? Αν κλώνος ενός γονιδίου υβριδίζεται μόνο με τους γονιδιωματικούς κλώνους C και D σε ποιο τμήμα του χρωμοσώματος 3 βρίσκεται? Αν ένα STS υβριδίζεται μόνο με έναν γονιδιωματικό κλώνο σε ποια τμήματα του χρωμοσώματος 3 μπορεί να βρίσκεται?



# Άσκηση 4 - Λύση

Αν ένα EST υβριδίζεται με τους γονιδιωματικούς κλώνους C, D και E αλλά όχι με τους άλλους κλώνους σε ποιο τμήμα του χρωμοσώματος 3 βρίσκεται αυτό το EST? *Το EST είναι μια θέση λίγων εκατοντάδων βάσεων πάνω σε κλώνων χιλιάδων βάσεων. Συνεπώς βρίσκεται στο τμήμα 5.*

Αν κλώνος ενός γονιδίου υβριδίζεται μόνο με τους γονιδιωματικούς κλώνους C και D σε ποιο τμήμα του χρωμοσώματος 3 βρίσκεται? Στο τμήμα 4.

Αν ένα STS υβριδίζεται μόνο με έναν γονιδιωματικό κλώνο σε ποια τμήματα του χρωμοσώματος 3 μπορεί να βρίσκεται? Στα 1,6 και 10.



# Άσκηση 5

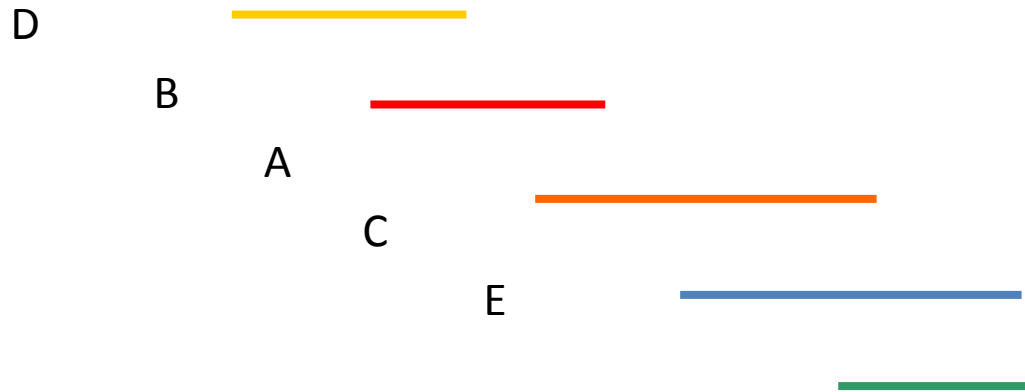
Πέντε γονιδιωματικοί DNA κλώνοι (A-E) από φορείς YAC ελέγχθηκαν με υβριδισμό για την παρουσία 6 STS. Τα αποτελέσματα φαίνονται στον παρακάτω πίνακα. Ποια είναι η σειρά των θέσεων STS στο χρωμόσωμα? Σχεδιάστε τον contig χάρτη που προκύπτει από τη συνένωση των κλώνων.

|   | STS1 | STS2 | STS3 | STS4 | STS5 | STS6 |
|---|------|------|------|------|------|------|
| A | +    | -    | +    | +    | -    | -    |
| B | +    | -    | -    | -    | +    | -    |
| C | -    | -    | +    | +    | -    | +    |
| D | -    | +    | -    | -    | +    | -    |
| E | -    | -    | +    | -    | -    | +    |



# Άσκηση 5 - Λύση

STS markers: 2 5 1 4 3 6



# Άσκηση 6

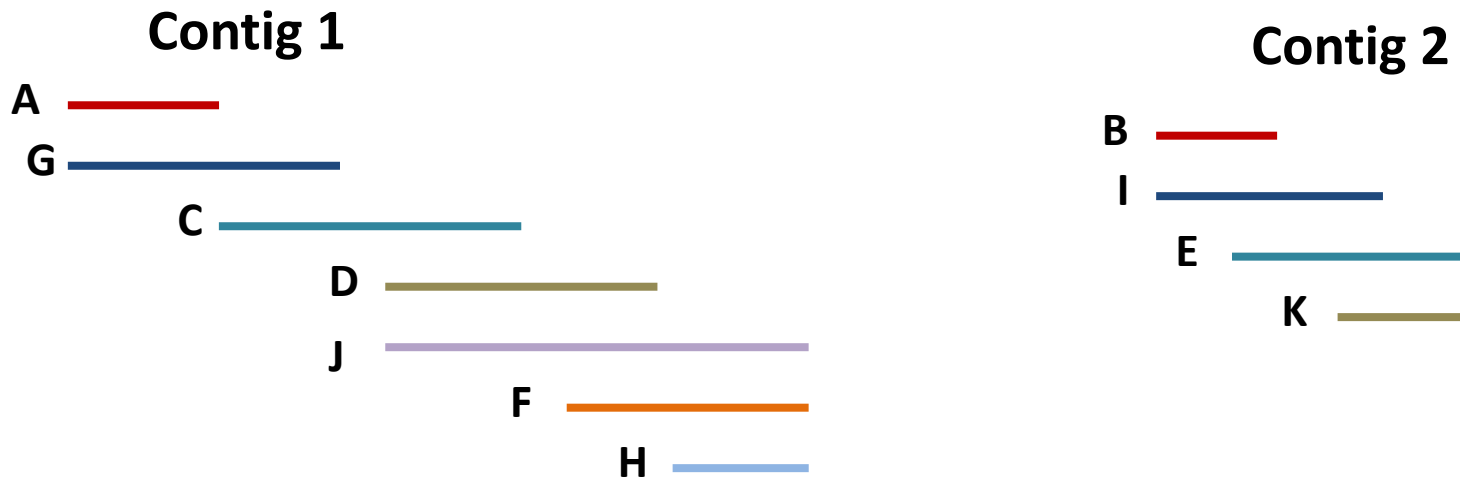
11 γονιδιωματικοί κλώνοι (A-K) από το χρωμόσωμα 4 της *Drosophila melanogaster* χρησιμοποιήθηκαν για να σχηματίσουν χάρτη από contigs με αλληπάλληλους υβριδισμούς. Τα αποτελέσματα φαίνονται στον παρακάτω πίνακα. Το (+) δείχνει υβριδισμό μεταξύ των κλώνων. Πόσα είναι τα contigs? Ποια η σειρά των κλώνων σε αυτά τα contigs?

|   | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| K | - | - | - | - | + | - | - | - | - | - | + |
| J | - | - | + | + | - | + | - | + | - | + |   |
| I | - | + | - | - | + | - |   | - | + |   |   |
| H | - | - | - | - | - | + | - | + |   |   |   |
| G | + | - | + | - | - | - | + |   |   |   |   |
| F | - | - | - | + | - | + |   |   |   |   |   |
| E | - | - | - | - | + |   |   |   |   |   |   |
| D | - | - | + | + |   |   |   |   |   |   |   |
| C | - | - | + |   |   |   |   |   |   |   |   |
| B | - | + |   |   |   |   |   |   |   |   |   |
| A | + |   |   |   |   |   |   |   |   |   |   |





# Άσκηση 6 - Λύση



# Χάρτες αλληλουχίας (1/10)

Οι **Χάρτες Αλληλουχίας** δείχνουν τη σειρά των νουκλεοτιδίων σε ένα κλωνοποιημένο τμήμα DNA

Δύο είναι οι στρατηγικές που ακολουθούνται:

1. Ιεραρχημένη στρατηγική αλληλούχισης (**hierarchical shotgun sequencing strategy**)
2. Ολική-γονιδιωματική στρατηγική αλληλούχισης (**whole-genome shotgun sequencing strategy-WGS**)

**Shotgun**: τα επικαλυπτόμενα τμήματα DNA που κλωνοποιούνται, προκύπτουν τυχαία από σπάσιμο ενθεμάτων BAC ή ολόκληρου του γονιδιώματος με τη βοήθεια υπερήχων ή μερικών πέψων με ένζυμα περιορισμού (όπως ακριβώς μια καραμπίνα χτυπάει με τα σκάγια της σε πολλά σημεία).



# Χάρτες αλληλουχίας (2/10)

## Ιεραρχημένη στρατηγική αλληλούχισης

<https://www.youtube.com/watch?v=-gVh3z6MwdU>

Χρησιμοποιήθηκε από το δημόσιο τομέα στα πλαίσια του HGP:

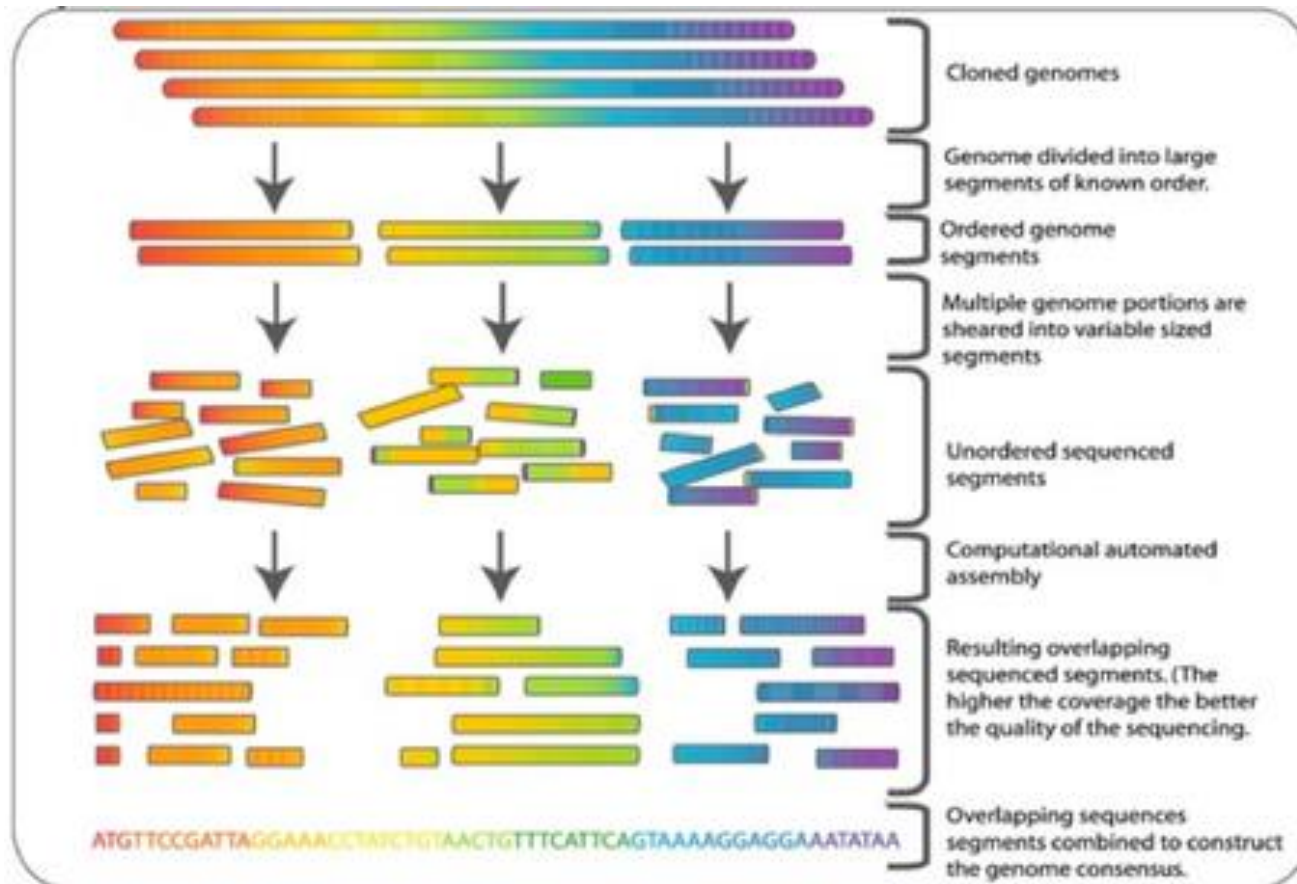
1. Δημιουργείται μια γονιδιωματική βιβλιοθήκη σε BACs
2. Δημιουργείται χάρτης από επικαλυπτόμενα BACs
3. Επιλέγονται όσο το δυνατόν ελάχιστα BACs για να έχουμε πλήρη κάλυψη του γονιδιώματος
4. Τα BACs σπάζουν σε κομμάτια μεγέθους ~2 Kb που κλωνοποιούνται σε πλασμίδια, <http://www.genome.gov/Edkit/flash/section3.html>
5. Ακολουθεί αλληλούχιση στα πλασμίδια μέχρις ότου επιτευχθεί μια 10πλάσια κάλυψη του BAC μέσω αλληλούχισης

[http://www.yourgenome.org/landing\\_hgp.shtml](http://www.yourgenome.org/landing_hgp.shtml)



# Χάρτες αλληλουχίας (3/10)

Εικόνα 10: Ιεραρχημένη στρατηγική αλληλούχισης



# Χάρτες αλληλουχίας (4/10)

## Ιεραρχημένη στρατηγική αλληλούχισης

- Πλεονέκτημα: χρειάζονται να αλληλουχηθούν ένας μικρός αριθμός πλασμιδίων/ BAC
- Μειονέκτημα: προϋποθέτει μια αρκετά χρονοβόρα και δαπανηρή προεργασία, για να δημιουργηθούν οι φυσικοί χάρτες των BACs και οι πλασμιδιακές βιβλιοθήκες για περίπου 15.000-20.000 BACs.



# Χάρτες αλληλουχίας (5/10)

## Ολική Γονιδιωματική Στρατηγική

[https://www.youtube.com/watch?v=vg7Y5EeZsjk&list=UUfe1\\_Tt2cyejbU3g0Hdo5lw](https://www.youtube.com/watch?v=vg7Y5EeZsjk&list=UUfe1_Tt2cyejbU3g0Hdo5lw)

Χρησιμοποιήθηκε από την ιδιωτική εταιρεία Celera: Ολόκληρο το γονιδίωμα σπάζει τρεις φορές για να κατασκευαστεί

- Μια πλασμιδιακή βιβλιοθήκη με ενθέματα ~ 2 Kb. Χρησιμοποιείται για 6πλάσια κάλυψη του γονιδιώματος
- Μια πλασμιδιακή βιβλιοθήκη με ενθέματα ~ 10 Kb. Χρησιμοποιείται για 3πλάσια κάλυψη του γονιδιώματος
- Μια BAC βιβλιοθήκη με ενθέματα ~ 200 Kb. Χρησιμοποιείται για μονή κάλυψη του γονιδιώματος

Χρησιμοποιούνται ενθέματα διαφορετικών μεγεθών επειδή βοηθάνε στην συναρμολόγηση τμημάτων μεταξύ των οποίων υπάρχουν κενά. Αν δύο αλληλουχίες δεν έχουν συναρμολογηθεί σε μία (υπάρχει κενό) αλλά ανακαλύπτονται και σε ένα μεγάλο ένθεμα τότε τα τμήματα αυτά σίγουρα αποτελούν μέρος ενός contig. Υπενθυμίζεται ότι σε οποιαδήποτε ένθεμα βιβλιοθήκης μπορούν να διαβαστούν μόνο η αρχή και το τέλος (από 600 βάσεις το πολύ), όχι το ενδιάμεσο.



# Χάρτες αλληλουχίας (6/10)

## Ολική Γονιδιωματική Στρατηγική



Λογική συναρμολόγησης γονιδιωματικών διαβασμάτων από βιβλιοθήκες διαφορετικών ενθεμάτων: Στην κάτω σειρά φαίνονται οι συναρμολογήσεις από τα ενθέματα των μικρών βιβλιοθηκών (2 και 10 kb) οπότε έχουν προκύψει δύο ανεξάρτητα contigs (A και B). Παρ' όλο που αυτά δεν έχουν ενωθεί σε ένα contig η αλληλεπικάλυψη του καθενός με τα δύο άκρα από τον κλώνο 200 kb (πάνω σειρά) μας επιτρέπει να καταλάβουμε ότι αποτελούν μέρος ενός συνεχόμενου χρωμοσωματικού τμήματος, δηλαδή μας επιτρέπουν την καλύτερη συναρμολόγηση.



# Χάρτες αλληλουχίας (7/10)

## Ολική Γονιδιωματική Στρατηγική

### Πλεονεκτήματα:

- Δεν χρειάζεται να κατασκευαστεί φυσικός χάρτης
- Απαιτεί τη δημιουργία μόνο μιας BAC και δύο πλασμιδιακών βιβλιοθηκών
- Βασίζεται σε μια και μόνο αυτοματοποιημένη και ώριμη πια τεχνική –την εύρεση της πρωτοδιάταξης

### Μειονέκτημα:

Το πρόβλημα που δημιουργούν οι επαναλαμβανόμενες ακολουθίες στη συναρμολόγηση.

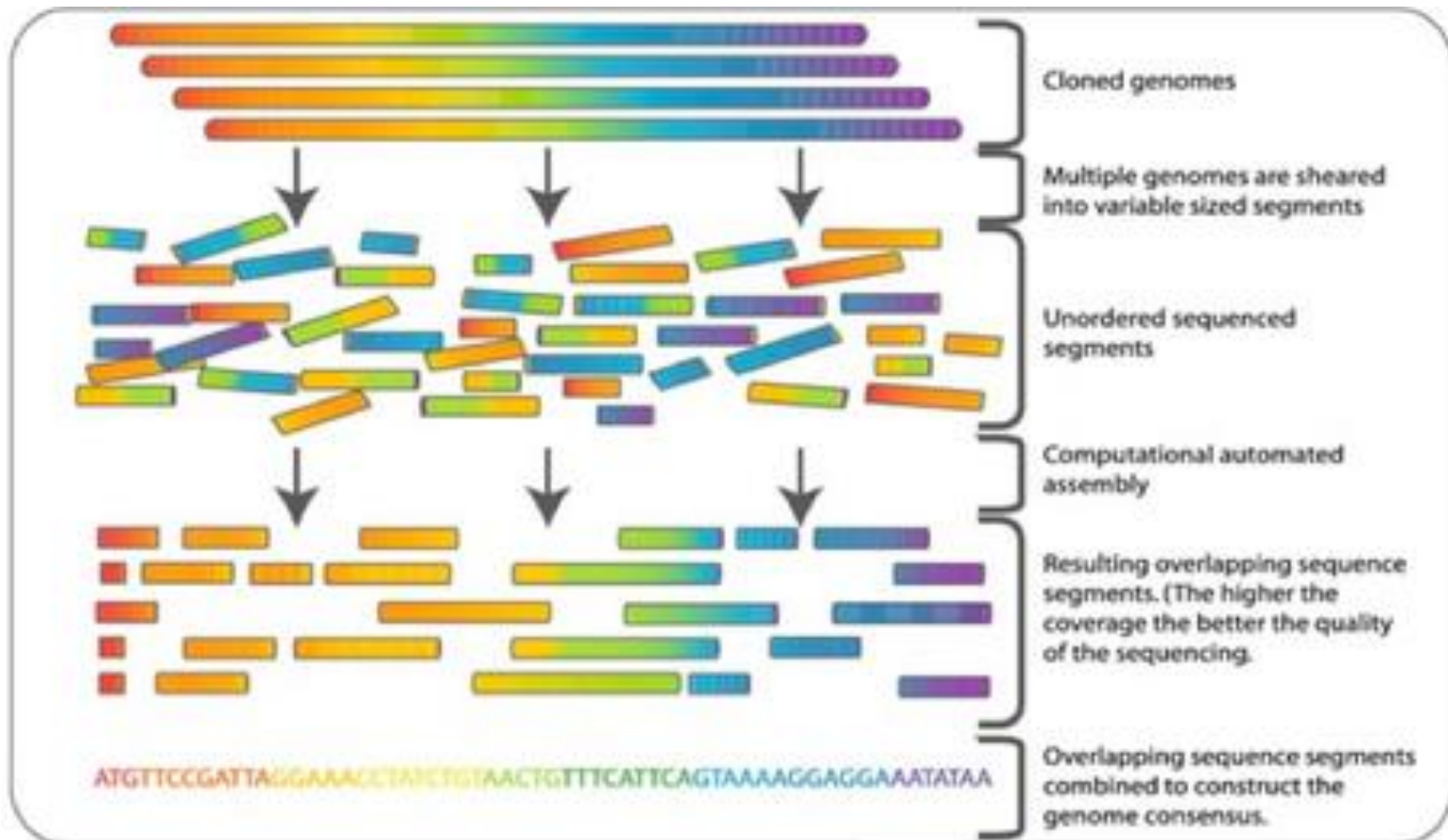
Ο δημόσιος τομέας πάντως κατηγόρησε τη Celera ότι δεν θα είχε καταφέρει τόσο γρήγορα να συναρμολογήσει / ολοκληρώσει το ανθρώπινο γονιδίωμα, χωρίς την ενσωμάτωση δεδομένων από τη δημιουργία των φυσικών και γενετικών χαρτών από τη δημόσια προσπάθεια.





# Χάρτες αλληλουχίας (8/10)

## Εικόνα 11: Ολική Γονιδιωματική Στρατηγική



# Χάρτες αλληλουχίας (9/10)

## Ολική Γονιδιωματική Στρατηγική

Εικόνα 12: Η ολοκληρωμένη αλληλουχία ενός τμήματος DNA προσδιορίζεται μέσω της συναρμολόγησης επικαλυπτόμενων αναγνώσεων αλληλουχίας.

### (α) Αναγνώσεις αλληλουχίας

- Ανάγνωση 1 CACATACACATGG
- Ανάγνωση 2 TCAATGGGGCTAA
- Ανάγνωση 3 AGCACGGACTTGTACATACACATG
- Ανάγνωση 4 ACACATGGAAATA
- Ανάγνωση 5 GGGCTAATGATTGTCAC
- Ανάγνωση 6 TGATTGTCACATA
- Ανάγνωση 7 ATTCATGAAGCACGGA
- Ανάγνωση 8 GTCACATACACATGATCAATGGGG

Συναρμολόγηση αναγνώσεων αλληλουχίας με τη βοήθεια υπολογιστή



Συναρμολογημένη αλληλουχία

(γ) ATTCATGAAGCACGGACTTGTACATACACATGATCAATGGGGCTAATGATTGTCACATACACATGGAAATA



# Χάρτες αλληλουχίας (10/10)

## Σύγκριση Στρατηγικών

**Μειονεκτήματα** και των 2 μεθόδων είναι ότι πάντα υπάρχουν ετεροχρωματινικές συμπυκνωμένες περιοχές που είναι αδύνατο να κλωνοποιηθούν, και κάποιες κλωνοποιημένες περιοχές ανασυνδυάζονται ή χάνουν τμήματα τους

Με την πάροδο του χρόνου, αποδείχθηκε ότι τα περισσότερα ερευνητικά κέντρα καταφεύγουν πια στην WGS προσέγγιση που είναι πιο γρήγορη και πιο φθηνή. Στο μέλλον μόνο όταν οι ερευνητές ενδιαφέρονται για την αναλυτική μελέτη ολόκληρου του γονιδιώματος ενός είδους θα χρησιμοποιούνται και οι δύο τεχνικές μέσα από ένα συνδυασμό τους.



---

# Το πρόβλημα αποθήκευσης των δεδομένων



# Ρυθμοί παραγωγής δεδομένων 2004

...και οι ρυθμοί αυξάνονται

Η διαδικασία παραγωγής μεγάλου πλήθους δεδομένων για το ανθρώπινο γονιδίωμα επιτεύχθηκε, βασικά, τον τελευταίο χρόνο της ολοκλήρωσής του. Τότε δημιουργήθηκαν μηχανήματα ικανά να διαβάσουν 96 δείγματα μέσα σε 4 ώρες. Συνεπώς, κάθε μηχάνημα μπορούσε να διαβάσει 345.600 bp την ημέρα (600 bp /τριχοειδές X 96 τριχοειδή X 6 διαβάσματα την ημέρα).

Επίσης, έγινε δυνατή η προετοιμασία και παραγωγή DNA τμημάτων με ρυθμούς εργοστασίων, για να προμηθεύουν συνεχώς τα αυτόματα μηχανήματα με DNA προς εύρεση πρωτοδιάταξης.

Τέλος, δημιουργήθηκαν Ερευνητικές μονάδες με 100-300 μηχανήματα εύρεσης πρωτοδιάταξης. Θεωρητικά τέτοιες μονάδες παράγουν 103.680.000 bp την ημέρα (345.600 bp X 300 μηχανήματα). Θεωρητικά μια τέτοια μονάδα θα μπορούσε να δώσει μια κάλυψη του ανθρώπινου γονιδιώματος σε 30 ημέρες ( $3 \times 10^9 / 1,0368 \times 10^8$ ).

Βεβαίως, υπήρχαν πάντα προβλήματα που επιβράδυναν τους ρυθμούς, όπως μειωμένη παραγωγή DNA τμημάτων προς πρωτοδιάταξη, πειραματικά λάθη, βλάβες στα μηχανήματα, αποτελέσματα χαμηλής ποιότητας.



# Και οι ρυθμοί αυξάνονται

Από το 2004 και μετά ... νέα μηχανήματα με υψηλότερη ακρίβεια, χαμηλότερο κόστος και γρηγορότερα αποτελέσματα

Μέχρι το 2004 βασιζόμαστε στην ίδια τεχνική, του Sanger!



# Νέα Μηχανήματα Αλληλούχισης (1/6)

## Roche-Pyrosequencing

- 2005
- Πραγματοποιεί 1-2 εκατομμύρια αντιδράσεις
- Μήκος ανάγνωσης: 400-500 βάσεις

<http://454.com/products/technology.asp>

<https://www.youtube.com/watch?v=bFNjxKHP8Jc>



# Νέα Μηχανήματα Αλληλούχισης (2/6)

## Applied Biosystems SOLiD

- 2006
- Πραγματοποιεί 100 εκατομμύρια αντιδράσεις
- Μήκος ανάγνωσης: 35 βάσεις

<http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems/solid-sequencing-chemistry.html>

<http://www.lifetechnologies.com/gr/en/home/life-science/sequencing/next-generation-sequencing/solid-next-generation-sequencing.html.html>

<https://www.youtube.com/watch?v=nlvyF8bFDwM>





# Νέα Μηχανήματα Αλληλούχισης (3/6)

## Illumina

- 2007
- Πραγματοποιεί 80 εκατομμύρια αντιδράσεις
- Μήκος ανάγνωσης: 45 βάσεις

[http://res.illumina.com/documents/products/techspotlights/techspotlight\\_sequencing.pdf](http://res.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf)

<http://www.illumina.com/systems/miseq.ilmn>

<https://www.youtube.com/watch?v=womKfikWlxM>



# Νέα Μηχανήματα Αλληλούχισης (4/6)

## 1000 Genomes Project

- An integrated map of genetic variation from 1092 human genomes  
The 1000 Genomes Project Consortium, 2012, Nature, 491(7422): 56–65
- Η αλληλούχιση 1000 ανθρώπινων γονιδιωμάτων από διάφορους πληθυσμούς θα παρέχει μια πιο λεπτομερή εικόνα της ανθρώπινης γενετικής ποικιλότητας. Έγινε δυνατό χάρι στα τρία μηχανήματα NGS: 454, Illumina, SOLiD. Στην επόμενη διαφάνεια φαίνονται τα τρία στάδια του προγράμματος και η παραγωγή των δεδομένων

<http://www.1000genomes.org/>



# Νέα Μηχανήματα Αλληλούχισης (5/6)

Πιλοτικό

Πρόγραμμα 1

Χαμηλή κάλυψη

179 δείγματα

@ 3,5 X

**2,7 Tbp σύνολο**

202 Gbp 454

1.8 Tbp Illumina

640 Gbp AB SOLiD

Πιλοτικό

Πρόγραμμα 2

**Deep trios** (CEU & YRI)

6 δείγματα

@ 41X

**1,1 Tbp σύνολο**

87 Gbp 454

773 Gbp Illumina

270 Gbp AB SOLiD

Πιλοτικό

Πρόγραμμα 3

**Κάλυψη εξωνίων**

697 δείγματα

@ 50X

2,2 Mbp των στόχων

8140 στόχοι



# Νέα Μηχανήματα Αλληλούχισης (6/6)

Μέσα στο 2012 έχει αλληλουχηθεί το γονιδίωμα 2500 ανθρώπων.  
... δηλαδή παραγωγή ποσότητας 2 HG ( $6 \times 10^9$ ) κάθε μέρα  
Στο τέλος του προγράμματος 1000 genomes μέσα σε τρία χρόνια,  
θα υπάρχουν  $3 \times 10^{12}$ bp, δηλαδή 60 φορές περισσότερα δεδομένα  
από ότι έχουν κατατεθεί σε δημόσιες βάσεις τα τελευταία 25 χρόνια.

Εκτιμήσεις...παλιές ...

2005  $10^1$  HG

2015  $10^3$  HG (3 Terabytes)

2025  $10^6$  HG (3.000 Terabytes)

2035  $10^9$  HG

Που θα αποθηκευθούν τα δεδομένα?

Πόσο σωστά θα αναλυθούν τα δεδομένα?



# Συναρμολόγηση και αποθήκευση δεδομένων (1/7)

- Το πρόγραμμα 1000 Genomes θα παράγει συνολικά 50 terrabytes (50.000.000.000.000) δεδομένων.
- Ένας καλός υπολογιστής (που έχει και τον αντίστοιχα μεγάλο σκληρό δίσκο) θα χρειαζόταν > 4,6 μέρες μόνο, για να κατεβάσει όλα τα δεδομένα του 1000 Genome Project.



# Συναρμολόγηση και αποθήκευση δεδομένων (2/7)

Το Sequence Read Archive (SRA) αποθηκεύει raw δεδομένα από τεχνολογίες αλληλούχησης επόμενης γενιάς, συμπεριλαμβανομένου: 454, Ion Torrent, Illumina, SOLiD, Helicos και Complete Genomics. Επιπλέον αποθηκεύει πληροφορίες στοίχισης.

<http://www.ncbi.nlm.nih.gov/Traces/home/>

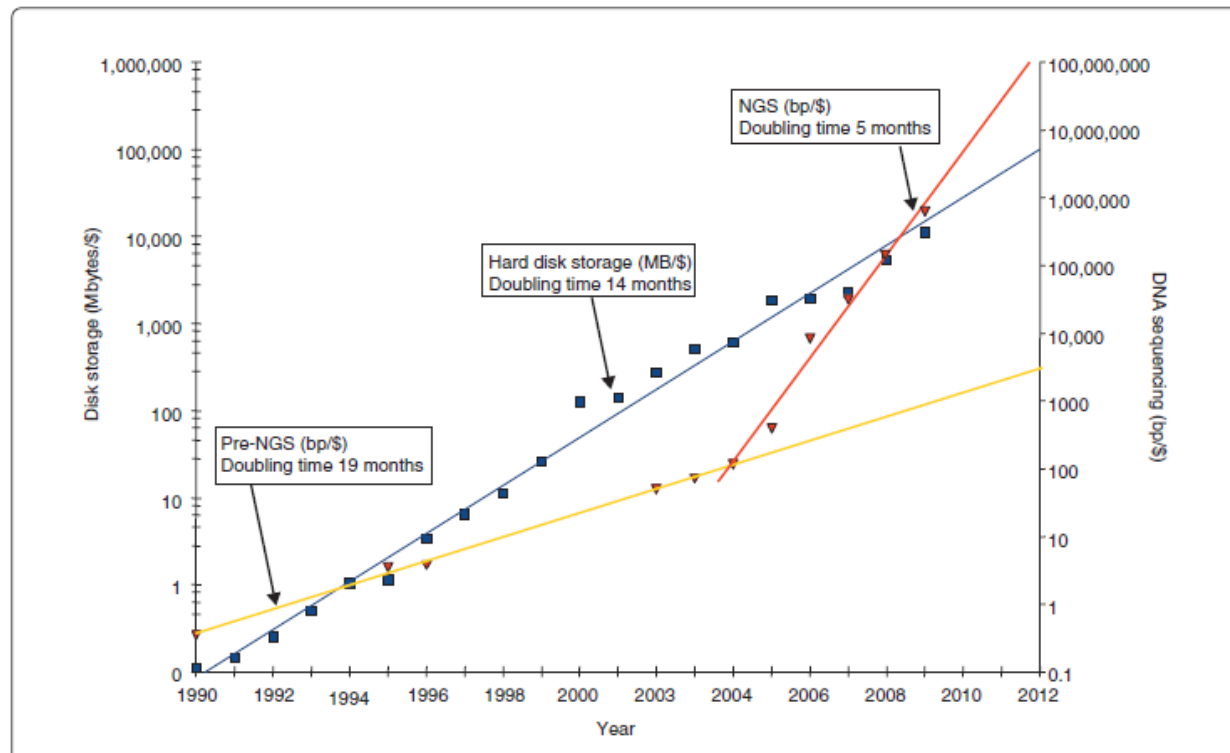


# Συναρμολόγηση και αποθήκευση δεδομένων (3/7)

## Διαχρονικό κόστος αλληλούχισης vs αποθήκευσης δεδομένων

Stein, The case for cloud computing in genome informatics  
*Genome Biology* 2010, **11**:207

**Εικόνα 13:** Από το 2004 (όταν εμφανίστηκαν τα NGS μηχανήματα) και μετά το κόστος παραγωγής DNA δεδομένων μειώνεται ανά έξι μήνες (από 1,5 χρόνο που ήταν πιο πριν). Μετά το 2010 κοστίζει λιγότερα να παράγεις DNA δεδομένα παρά να τα αποθηκεύσεις.



# Συναρμολόγηση και αποθήκευση δεδομένων (4/7)

Πόσα δεδομένα θα παράγονται; 700 Mb, 600 Gb, 1 Tb, 1 Pb;;;

Τι μήκος θα είναι ένα διάβασμα; 50bp, 150 bp, 10Kbp, 100Kbp;;;

Πόσο θα είναι το κόστος; 300M\$ 100K\$, 1000\$, 1\$;;;

Πόσος χρόνος θα χρειάζεται; 15 χρόνια, 1 εβδ., 1 h, 1 min;;;

Το BGI παράγει 6 Tb δεδομένα την ημέρα....

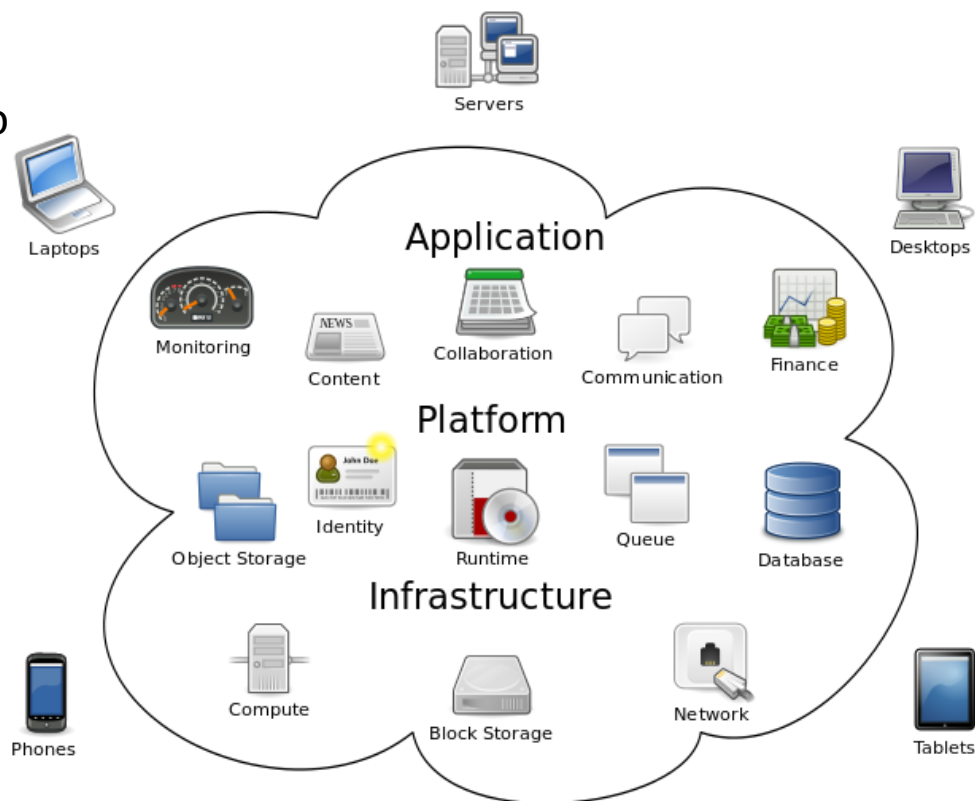




# Συναρμολόγηση και αποθήκευση δεδομένων (5/7)

Δεν είναι λοιπόν παράξενο που οι ερευνητές ανησυχούν περισσότερο για το πού (σε ποιους υπολογιστές/βάσεις δεδομένων) θα κατατεθούν αυτά τα αυξανόμενα δεδομένα στο μέλλον και το πώς θα αναλυθούν στη συνέχεια, παρά για το πώς θα παραχθούν τα δεδομένα. Και η απάντηση είναι ...στα σύννεφα!!!

Ηδη μεγάλες εταιρείες του διαδικτύου όπως η Amazon, Google, Microsoft και άλλες, συνεργάζονται με μεγάλα γονιδωματικά κέντρα προσπαθώντας να βρουν λύση και αυτό που προτείνουν είναι ... σύννεφα υπολογιστών (**cloud computing**)!

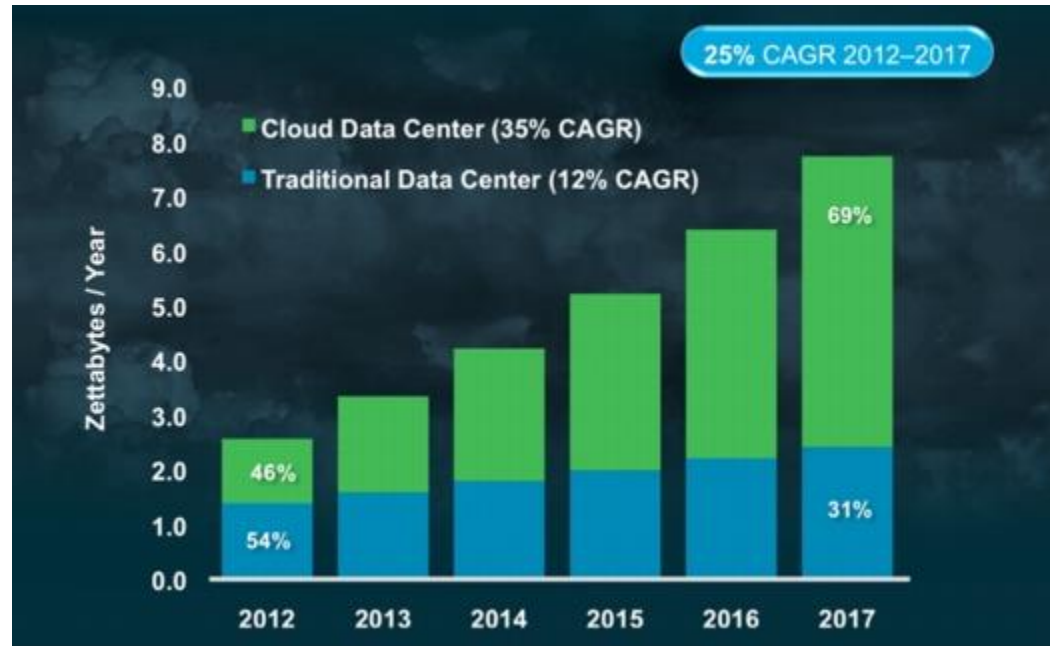


## Cloud Computing

Εικόνα 14: Cloud computing



# Συναρμολόγηση και αποθήκευση δεδομένων (6/7)

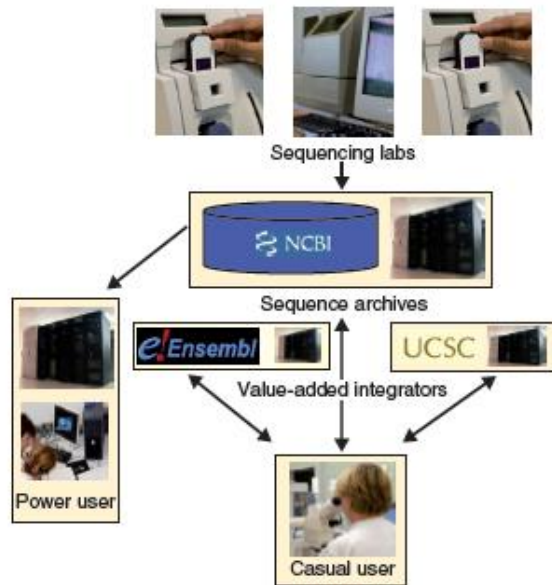


**Εικόνα 15:** Μέχρι το 2017 > τα δύο τρίτα όλων των αναλύσεων σε υπολογιστές θα γίνονται στα σύννεφα

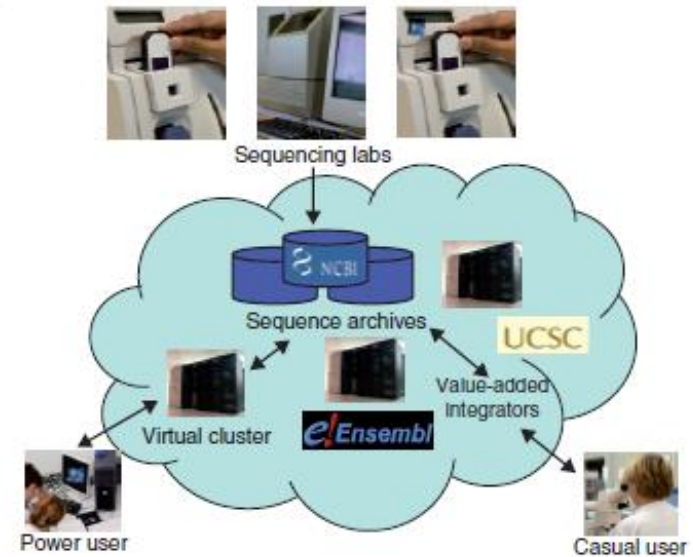
[http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud\\_Index\\_White\\_Paper.html](http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.html)



# Συναρμολόγηση και αποθήκευση δεδομένων (7/7)



**Εικόνα 16:** Κατά τον παραδοσιακό τρόπο ανάλυσης δεδομένων, είτε απλός ερευνητής, είτε μεγάλο εργαστήριο έπρεπε να κατεβούν τα δεδομένα στους προσωπικούς υπολογιστές.



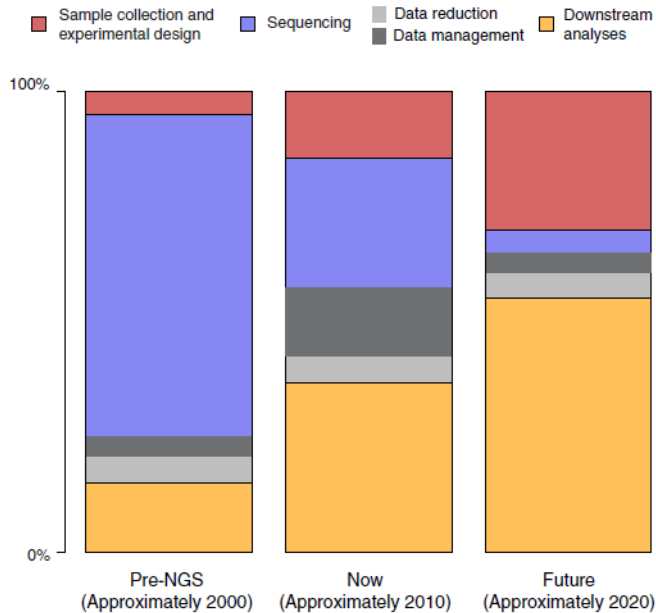
**Εικόνα 17:** Στον σύγχρονο τρόπο ανάλυσης δεδομένων, τα δεδομένα θα είναι όλα αποθηκευμένα στο διαδίκτυο και οι αναλύσεις θα γίνονται σε εικονικούς υπολογιστές.

*Stein, Genome Biology 2010, 11:207*



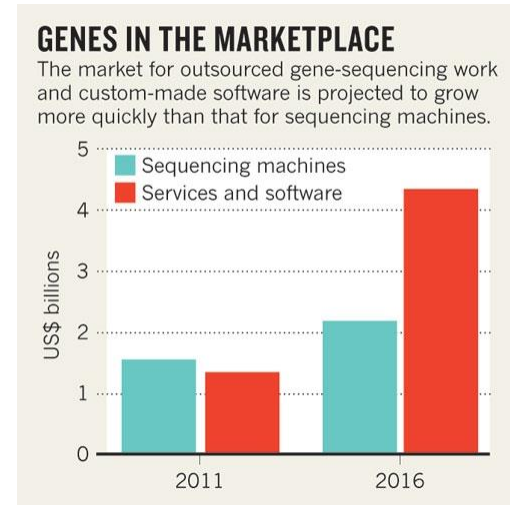
# Το μέλλον των αναλύσεων (1/3)

Sboner et al. Genome Biology 2011, 12 :125



**Εικόνα 18:** Όλο και περισσότερο σημασία και χρόνο θα έχει η ανάλυση των δεδομένων και όχι η παραγωγή τους.

Nature 2013, 495, 293



**Εικόνα 19:** Συνεπώς οι γονιδιωματικές εταιρείες δραστηριοποιούνται στην παροχή υπηρεσιών ανάλυσης δεδομένων και όχι μόνο παραγωγής τους.



# Το μέλλον των αναλύσεων (2/3)

| Primary Analysis | Secondary Analysis | Tertiary Analysis | Post Tertiary Analysis                 |
|------------------|--------------------|-------------------|--|
| Base Calling     | Alignment          | Variant Calling   | In – Depth Annotation                  |
|                  |                    |                   |  |
| Output           | Output             | Output            | Output                                 |
| Fastaq           | SAM                | VCF               | <b>Biologically meaningful results</b> |
| XSq              | BAM                | Various tables    |  |

Ακόμα είμαστε στα τρία πρώτα επίπεδα, δεν έχουμε φτάσει στο τέταρτο



# Το μέλλον των αναλύσεων (3/3)

Το μότο της εταιρείας γονιδιωματικής BGI λανσάροντας την υπηρεσία Easy Genomics που προσφέρει στο κοινό για επεξεργασία των αποτελεσμάτων NGS αλληλούχισης είναι η φράση του Albert Einstein : If you cant explain it simply you don't understand it well enough.

<https://www.easygenomics.com/index>

Αντίστοιχα και η Illumina προσφέρει μια αντίστοιχη υπηρεσία την BaseSpace:

<https://basespace.illumina.com/home/index>

**Biology: The big challenges of big data**

<http://www.nature.com/nature/journal/v498/n7453/full/498255a.html>



# Προστασία Δεδομένων

- Privacy protections: The genome hacker

Nature, 8/5/2013

<http://www.nature.com/news/privacy-protections-the-genome-hacker-1.12940>

- Personal Genome Project (PGP): Αποτελεί μια πηγή ελεύθερα διαθέσιμων γενομικών και περιβαλλοντικών δεδομένων, καθώς και δεδομένων για ανθρώπινα χαρακτηριστικά. Δημιουργήθηκε από τον George Church το 2005 και έχει πρωτοπορήσει σε ηθικά, νομικά και τεχνικά ζητήματα που σχετίζονται με τη δημιουργία δημόσιων πόρων, που περιλαμβάνουν αναγνωρίσιμα δεδομένα, όπως τα ανθρώπινα γονιδιώματα.

<http://www.personalgenomes.org/organization/pgp>



# Εύρεση Γονίδιων στις Αλληλουχίες του Γονιδιώματος (1/4)

**Απαραίτητη η χρήση υπολογιστών.** Δίνουν ποσοστό ασφαλείας στις προβλέψεις. :

- Αναζήτηση με βάση «γνωστά» γονίδια. Αναζήτηση ομολογίας. Σύγκριση με αλληλουχίες cDNA ή ESTs (Expressed Sequence Tags) του ίδιου είδους με αλληλουχίες γνωστών γονιδίων άλλου είδους.
  - Αναζήτηση «στα τυφλά» χαρακτηριστικών γονιδίων. Λαμβάνονται υπόψη συγκεκριμένα χαρακτηριστικά των γονιδίων, π.χ. θέσεις έναρξης της μεταγραφής (ATG) και λήξης της μετάφρασης (π.χ. UAA), θέσεις ένωσης εξονίων/ κοπής ιντρονίων (GT/AG), περιοχές προαγωγέων (κουτιά TATA). Αναζήτηση Ανοιχτού Αναγνωστικού Πλαίσιου (open reading frames-ORFs)
  - Σύγκριση ολόκληρων γονιδιωμάτων. Εύρεση **συντηρημένων** περιοχών που θα αντιστοιχούν προφανώς σε λειτουργικές περιοχές
- Υπαρξη ιντρονίων, junk DNA, υψηλών εξελικτικών αλλαγών δυσχεραίνουν τη διαδικασία. Ποσοστό επιτυχίας 70-90%.





# Εύρεση Γονίδιων στις Αλληλουχίες του Γονιδιώματος (2/4)

## Αναγνωστικό πλαίσιο

5' atgcccaagctgaatagcgtagaggggttttcatcatttgaggacgatgtataa 3'

Στα περισσότερα είδη ένα reading frame αρχίζει με το atg (met) και τελειώνει με τα (taa, tag, tga)

Τα αναγνωστικά πλαίσια είναι τρία στη μια αλυσίδα και τρία στη συμπληρωματική

### Reading Frame 1

atg ccc aag ctg aat agc gta gag ggg ttt tca tca ttt gag gac gat gta taa  
M P K L N S V E G F S S F E D D V \*

### Reading Frame 2

tgc cca agc tga ata gcg tag agg ggt ttt cat cat ttg agg acg atg tat  
C P S \* I A \* R G F H H L R T M Y

### Reading Frame 3

gcc caa gct gaa tag cgt aga ggg gtt ttc atc att tga gga cga tgt ata  
A Q A E \* R R G V F I I \* G R C I





# Εύρεση Γονίδιων στις Αλληλουχίες του Γονιδιώματος (4/4)

Με τη χρήση υπολογιστών βρέθηκαν όλα τα ανοιχτά αναγνωστικά πλαίσια (ORFs) στη συγκεκριμένη γονιδιωματική περιοχή και απεικονίζονται με κόκκινο χρώμα.

```
GTGGATAACTTGGGTAGAAATGGCGACCCCTTCTCATCAGGAAGGTTAATCTTTAAATGATTTGAATTTAAAACGCAGACATAGGGGATACACATGCTTTGGACA/
GACTGCTTAACTCGCTTGGGACAAAGAGCTCTCTGATAACGCTCTTGGGATGTGGATTCGCCCTTTAGTAGCTGAAGAAGTAGAGGGGATACACTGCTCTATGCA/
TCCTAATCCTTATTTGGACCGGTATATTTCAAGAGAATCATTITAGAGTTAATTTCTATATTTGGCTGAACAAATGTGAGAAGGGCGCGTCAGGTTGAAATTTT/
GGTAGATTTCTCGTCTGGTAGTATTTTGTCCCTTASTGAACAGCCTGCAACAACTACAGCAGCTTTACAACCTGCCCTTATACCTCAACCTGCTAAGGTTAAAAG/
AGAACCGGAACCTGTTGCTAATACTGCAGTTAGTTCTAAGAGTTCAAAAAAGAAACTATTAATCCACAAATTTACTTTTTCACATATTGTTGAAGGCCGTTCTAA/
ATGAATGGCAGCAGAAACCTGTAGAAAAGTATTAACACAGTTAGGTGCTTCTCAACATAACCCCTTTATTTTATATGGTCCGACAGGCTTAGGTAAAGACTCACT/
AATGCAAGCAGTTGGTAATGCTTACTGCAAGCGAAGCCGAATGCAAGAGTCATGTATATGACTTCAGAAAAGTTTGTACAAGATTTGTGAGCTCATACAAA/
AGGAAGGTAGAAAGCTTTAAGAAAAATTCGCCGCTCTACTAGTCACTGCTACGTTACAGCTACACAGATCAGCATTATATATCGCCGCTACAGTCTGATCTACGTT/
GTGGATAACTTGGGTAGAAATGGCGACCCCTTCTCATGAGGAAGGTTAATCTTTAAATGATTTGAATTTAAAACGCAGACATAGGGGATACACATGCTTTGGACA/
GACTGCTTAACTCGCTTGGGACAAAGAGCTCTCTGATAACGCTCTTGGGATGTGGATTCGCCCTTTAGTAGCTGAAGAAGTAGAGGGGATACACTGCTCTATGCT/
TCCTAATCCTTATTTGGACCGGTATATTTCAAGAGAATCATTITAGAGTTAATTTCTATATTTGGCTGAACAAATGTGAGAAGGGCGCGTCAGGTTGAAATTTG/
GTAGATTTCTCGCTCTGGTAGTATTTTGTCCCTTGAAGAGCCTGCAACAACTACAGCAGCTTTACAACCTGCCCTTATACCTCAACCTGCTAAGGTTAAAAG/
AGAACCGGAACCTGTTGCTAATACTGCAGTTAGTTCTAAGAGTTCAAAAAAGAAACTATTAATCCACAAATTTACTTTTTCACATATTGTTGAAGGCCGTTCTAA/
CCAAATGGCAGCAGAAACCTGTAGAAAAGTATTAACACAGTTAGGTGCTTCTCAACATAACCCCTTTATTTTATATGGTCCGACAGGCTTAGGTAAAGACTCACT/
AATGCAAGCAGTTGGTAATGCTTACTGCAAGCGAAGCCGAATGCAAGAGTCATGTATATGACTTCAGAAAAGTTTGTACAAGATTTGTGAGCTCATACAAA/
AGCAAGCTAGAAAGCTTTAAGAAAAATTCGCCGCTCTACTAGTCACTGCTACGTTACAGCTACACAGATCAGCATTATATATCGCCGCTACAGTCTGATCTACGTT/
TAGATTTCTCGTCTGGTAGTATTTTGTCCCTTGAAGAGCCTGCAACAACTACAGCAGCTTTACAACCTGCCCTTATACCTCAACCTGCTAAGGTTAAAAG/
AGAACCGGAACCTGTTGCTAATACTGCAGTTAGTTCTAAGAGTTCAAAAAAGAAACTATTAATCCACAAATTTACTTTTTCACATATTGTTGAAGGCCGTTCTAA/
GTGGATAACTTGGGTAGAAATGGCGACCCCTTCTCATGAGGAAGGTTAATCTTTAAATGATTTGAATTTAAAACGCAGACATAGGGGATACACATGCTTTGGACA/
TGCTTAACTCGCTTGGGACAAAGAGCTCTCTGATAACGCTCTTGGGATGTGGATTCGCCCTTTAGTAGCTGAAGAAGTAGAGGGGATACACTGCTCTATGCTCCT/
CCTTATTTGGACCGGTATATTTCAAGAGAATCATTITAGAGTTAATTTCTATATTTGGCTGAACAAATGTGAGAAGGGCGGGTCCGTCAGGTTGAAATTTTGGTAG/
CGTCTGGTAGTATTTTGTCCCTTGAAGAGCCTGCAACAACTACAGCAGCTTTACAACCTGCCCTTATACCTCAACCTGCTAAGGTTAAAAGAGAACCGG/
TGTGCTAATACTGCAGTTAGTTCTAAGAGTTCAAAAAAGAAACTATTAATCCACAAATTTACTTTTTCACATATTGTTGAAGGCCGTTCTAACCAATGGCAGC/
AACCTGTAGAAAAGTATTAACACAGTTAGGTGCTTCTCAACATAACCCCTTTAATTTTATATGGTCCGACAGGCTTAGGTAAAGACTCACTTAAATGCAAGCAGTTGG/
TGCTTACTGCAAGCCGAAGCCGAATGCAAGAGTCATGTATATGACTTCAGAAAAGTTTGTACAAGATTTTGTGAGCTCAITACAAAAGGAAAGGTAGAAGAGTT/
GAAAAATTCGCCGCTAGCTGATCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT/
GTGGATAACTTGGGTAGAAATGGCGACCCCTTCTCATGAGGAAGGTTAATCTTTAAATGATTTGAATTTAAAACGCAGACATAGGGGATACACATGCTTTGGACA/
CTGCTTAACTCGCTTGGGACAAAGAGCTCTCTGATAACGCTCTTGGGATGTGGATTCGCCCTTTAGTAGCTGAAGAAGTAGAGGGGATACACTGCTCTATGCTCCT/
TCCTTATTTGGACCGGTATATTTCAAGAGAATCATTITAGAGTTAATTTCTATATTTGGCTGAACAAATGTGAGAAGGGCGGGTCCGTCAGGTTGAAATTTGGTAG/
CTCGCTCTGGTAGTATTTTGTCCCTTGAAGAGCCTGCAACAACTACAGCAGCTTTACAACCTGCCCTTGTCTTATACCTCAACCTGCTAAGGTTAAAAG/
AGAACCGGAACCTGTTGCTAATACTGCAGTTAGTTCTAAGAGTTCAAAAAAGAAACTATTAATCCACAAATTTACTTTTTCACATATTGTTGAAGGCCGTTCTAA/
AATGGCAGCAGAAACCTGTAGAAAAGTATTAACACAGTTAGGTGCTTCTCAACATAACCCCTTTAATTTTATATGGTCCGACAGGCTTAGGTAAAGACTCACTTAA/
AGCAGTTGGTAATGCTTACTGCAAGCGAAGCCGAATGCAAGAGTTGAGTCATGTATATGACTTCAGAAAAGTTTGTACAAGATTTTGTGAGCTCATACAAA/
AAAGGTAGAAAGCTTTAAGAAAAATTCGCCGAAAGGTAGAAGAGCAGCCGCGCCGCGCCGCGCAAGAAAAATTCGCCGATATATATATAGCATCGATCGATGCAT/
GTGGATAACTTGGGTAGAAATGGCGACCCCTTCTCATGAGGAAGGTTAATCTTTAAATGATTTGAATTTAAAACGCAGACATAGGGGATACACATGCTTTGGACA/
TGCTTAACTCGCTTGGGACAAAGAGCTCTCTGATAACGCTCTTGGGATGTGGATTCGCCCTTTAGTAGCTGAAGAAGTAGAGGGGATACACTGCTCTATGCTCCT/
CCTTATTTGGACCGGTATATTTCAAGAGAATCATTITAGAGTTAATTTCTATATTTGGCTGAACAAATGTGAGAAGGGCGGGTCCGTCAGGTTGAAATTTTGGTAG/
CGTCTGGTAGTATTTTGTCCCTTGAAGAGCCTGCAACAACTACAGCAGCTTTACAACCTGCCCTTATACCTCAACCTGCTAAGGTTAAAAGAGAACCGG/
TGTGCTAATACTGCAGTTAGTTCTAAGAGTTCAAAAAAGAAACTATTAATCCACAAATTTACTTTTTCACATATTGTTGAAGGCCGTTCTAACCAATGGCAGC/
AACCTGTAGAAAAGTATTAACACAGTTAGGTGCTTCTCAACATAACCCCTTTAATTTTATATGGTCCGACAGGCTTAGGTAAAGACTCACTTAAATGCAAGCAGTTGG/
TGCTTACTGCAAGCCGAAGCCGAATGCAAGAGTCATGTATATGACTTCAGAAAAGTTTGTACAAGATTTTGTGAGCTCAITACAAAAGGAAAGGTAGAAGAGTT/
GAAAAATTCGCCGCTAGCTGATCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT/
```



# Σημείωμα χρήσης έργων τρίτων

**Ιεραρχημένη στρατηγική αλληλούχησης**, [http://en.wikipedia.org/wiki/File:Whole\\_genome\\_shotgun\\_sequencing\\_versusHierarchical\\_shotgun\\_sequencing.png](http://en.wikipedia.org/wiki/File:Whole_genome_shotgun_sequencing_versusHierarchical_shotgun_sequencing.png), by Commins, J., Toft, C., Fares, M. A., CC-BY-SA-2.5 (<http://creativecommons.org/licenses/by-sa/2.5/deed.en>)

**Ολική γονιδιωματική στρατηγική**, [http://en.wikipedia.org/wiki/File:Whole\\_genome\\_shotgun\\_sequencing\\_versusHierarchical\\_shotgun\\_sequencing.png](http://en.wikipedia.org/wiki/File:Whole_genome_shotgun_sequencing_versusHierarchical_shotgun_sequencing.png), by Commins, J., Toft, C., Fares, M. A., CC-BY-SA-2.5 (<http://creativecommons.org/licenses/by-sa/2.5/deed.en>)

**Cloud computing**, [http://es.wikipedia.org/wiki/Archivo:Cloud\\_computing.svg](http://es.wikipedia.org/wiki/Archivo:Cloud_computing.svg), by Sam Johnston, CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/deed.en>)



# Σημείωμα Αναφοράς

Copyright Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Τριανταφυλλίδης  
Αλέξανδρος. «Ειδικά Θέματα Γενετικής. Δομική Γονιδιωματική». Έκδοση: 1.0.  
Θεσσαλονίκη 2015. Διαθέσιμο από τη δικτυακή διεύθυνση:  
[http://opencourses.auth.gr/eclass\\_courses](http://opencourses.auth.gr/eclass_courses).



# Σημείωμα Αδειοδότησης

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



Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

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





# Τέλος ενότητας

Επεξεργασία: Μηνούδη Στυλιανή  
Θεσσαλονίκη, Χειμερινό εξάμηνο 2014-2015

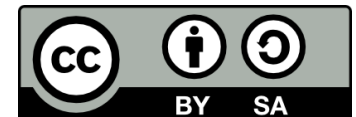


Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα χρήσης έργων τρίτων

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

