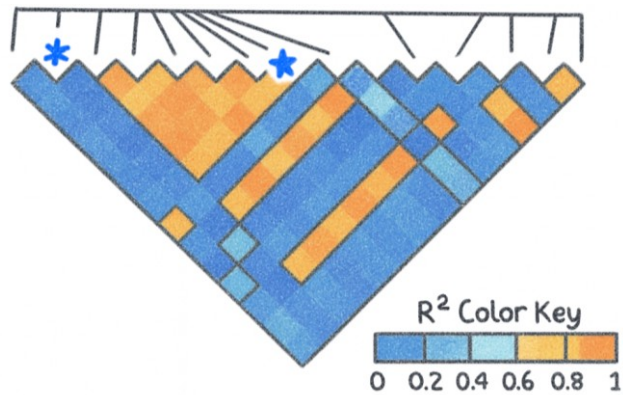


Πληθυσμιακή και Εξελικτική Γενετική

Ανισορροπία Σύνδεσης και Απλότυποι



Μαργαρίτης Τσιφιντάρης

mtsifintaris@gmail.com / mtsifint@mbg.duth.gr

Τι θα μάθουμε σήμερα?

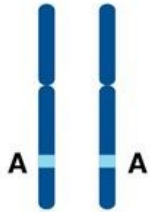
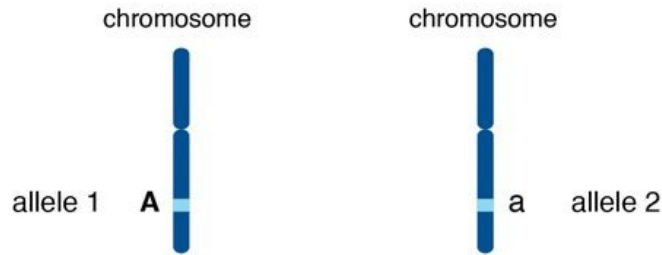
- Τι είναι η ισορροπία σύνδεσης (linkage equilibrium)
- Τι είναι η ανισορροπία σύνδεσης (**linkage disequilibrium** - LD)
- Τι είναι ο απλότυπος - παραδείγματα
- Ποια είναι τα μέτρα εκτίμησης της LD
- Πώς αναπαρίσταται η LD
- Από ποιους παράγοντες επηρεάζεται η LD
- Πώς μεταβάλλεται η LD στο χρόνο
- Πού και πώς αξιοποιείται αυτή η γνώση

Μία γενετική θέση
ή
Ένας γενετικός τόπος
(locus)

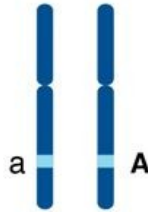
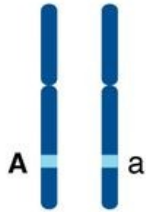


Αλληλόμορφα

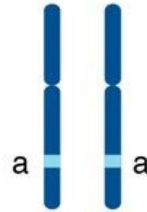
Εναλλακτικές μορφές ενός γονιδίου ή μίας αλληλουχίας (τμήματος) DNA σε ένα συγκεκριμένο γενετικό τόπο



Homozygous AA

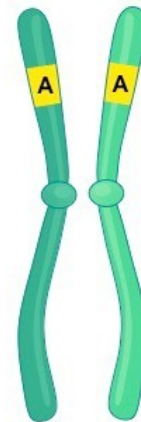


Heterozygous Aa

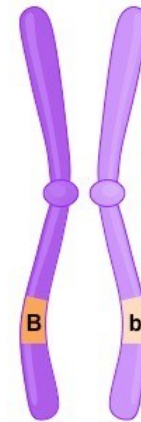


Homozygous aa

1^{ος} Νόμος του Mendel [segregation of alleles] –
Ο διαχωρισμός των αλληλομόρφων ενός γονιδίου στους γαμέτες είναι τυχαίος



Homozygous
Alleles are same



Heterozygous
Alleles are different



Hemizygous
Only one allele
(e.g. XY)

Επανάληψη...

$$p^2 + 2pq + q^2 = (p+q)^2 = 1$$

Θεώρημα Hardy-Weinberg

Θηλυκοί γαμέτες (συχνότητα)

Αρσενικοί γαμέτες (συχνότητα)

	A (p)	a (q)
A (p)	AA ($p \cdot p = p^2$)	Aa ($p \cdot q$)
a (q)	aA ($q \cdot p$)	aa ($q \cdot q = q^2$)

$$P(AA) = p^2$$

$$P(Aa) = p \cdot q \text{ (μητέρα)} + p \cdot q \text{ (πατέρα)} = 2pq$$

$$P(aa) = q^2$$

Προϋποθέσεις για την ισχύ του θεωρήματος H-W και . . .

αίτια απόκλισης από αυτό

- Διπλοειδής οργανισμός
- Σεξουαλική αναπαραγωγή
- Τυχαία σύζευξη των ατόμων (παμμιξία) (**vs** διαστρωμάτωση, ομοιοτυπική σύζευξη, συγγένεια εξ αίματος & ενδογαμία)
- Μη επικαλυπτόμενες γενιές
- Μέγεθος πληθυσμού πολύ μεγάλο (~άπειρο) [**vs** μικρό (γενετική παρέκκλιση)]
- Ίδιες συχνότητες αλληλομόρφων σε αρσενικά και θηλυκά άτομα (**vs** mtDNA, Y-chr, non-PAR1/2 στα αρσενικά)
- Δεν υπάρχει επιλογή γονοτύπου (**vs** θνησιγόνα αλληλόμορφα, πλεονέκτημα ετεροζυγώτη)
- Δεν υπάρχει εισροή αλληλομόρφων στον πληθυσμό και άρα αλλαγή συχνοτήτων των αλληλομόρφων (**vs** μετανάστευση ή μεταλλαγή)

• • • • Ο Νόμος Hardy-Weinberg

Ο νόμος Hardy-Weinberg ισχύει υπό τις εξής προϋποθέσεις:

- Ο πληθυσμός είναι μεγάλος και οι συζεύξεις είναι τυχαίες, όσον αφορά στον υπό μελέτη γενετικό τόπο.
- Οι συχνότητες των αλληλομόρφων παραμένουν σταθερές με την πάροδο του χρόνου γιατί:

Ο ρυθμός μεταπληξίγνενσης δεν είναι σημαντικός.

Τα άτομα όλων των γονότυπων είναι εξίσου ικανά να αναπαραχθούν και να μεταβιβάσουν τα γονίδιά τους στους απογόνους τους, δηλαδή δεν υπάρχει επιλογή εις βάρος κάποιου συγκεκριμένου γονοτύπου.

Δεν υπήρχε σημαντική είσοδος ατόμων από έναν άλλον πληθυσμό με συχνότητες αλληλομόρφων πολύ διαφορετικές από αυτές του ενδογενούς πληθυσμού.

Hardy-Weinberg Equilibrium for a Three-Allele System

$$p + q + r = 1$$

p, q, and r are the three alleles in the population (system).

$$(p + q + r)^2 = 1$$

"Cross" the individual alleles together by squaring the trinomial.



$$p^2 + q^2 + r^2 + 2pq + 2pr + 2qr = 1$$

@wormthoughts

$(p_1 + p_2 + p_3 + \dots + p_n)^2 = 1$, εάν σε έναν γενετικό τόπο έχουν περιγραφεί n διαφορετικά αλληλόμορφα

Hardy–Weinberg Principle

- X-linked genes are a special case because males have only one X-chromosome
- Genotype frequencies among females: $HH = p^2$; $Hh = 2pq$; $hh = q^2$
- Genotype frequencies among males are the **same** as allele frequencies:
 $H = p$,
 $h = q$

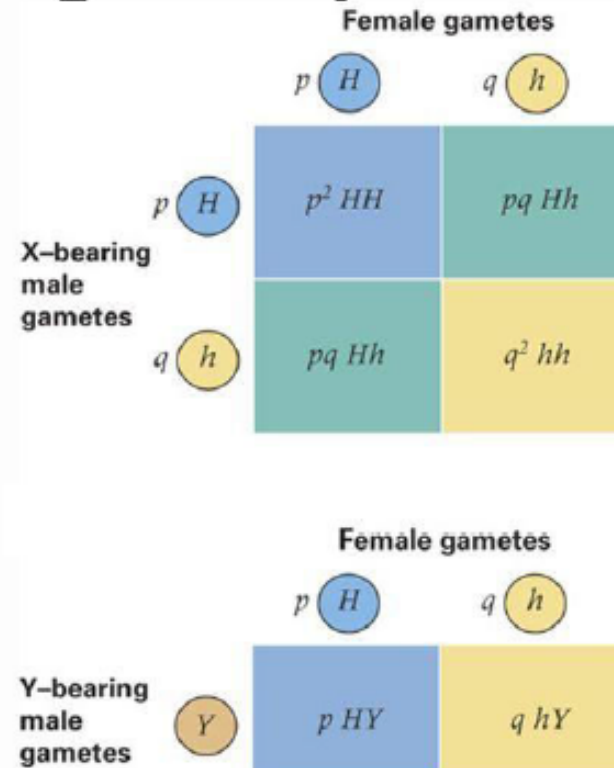
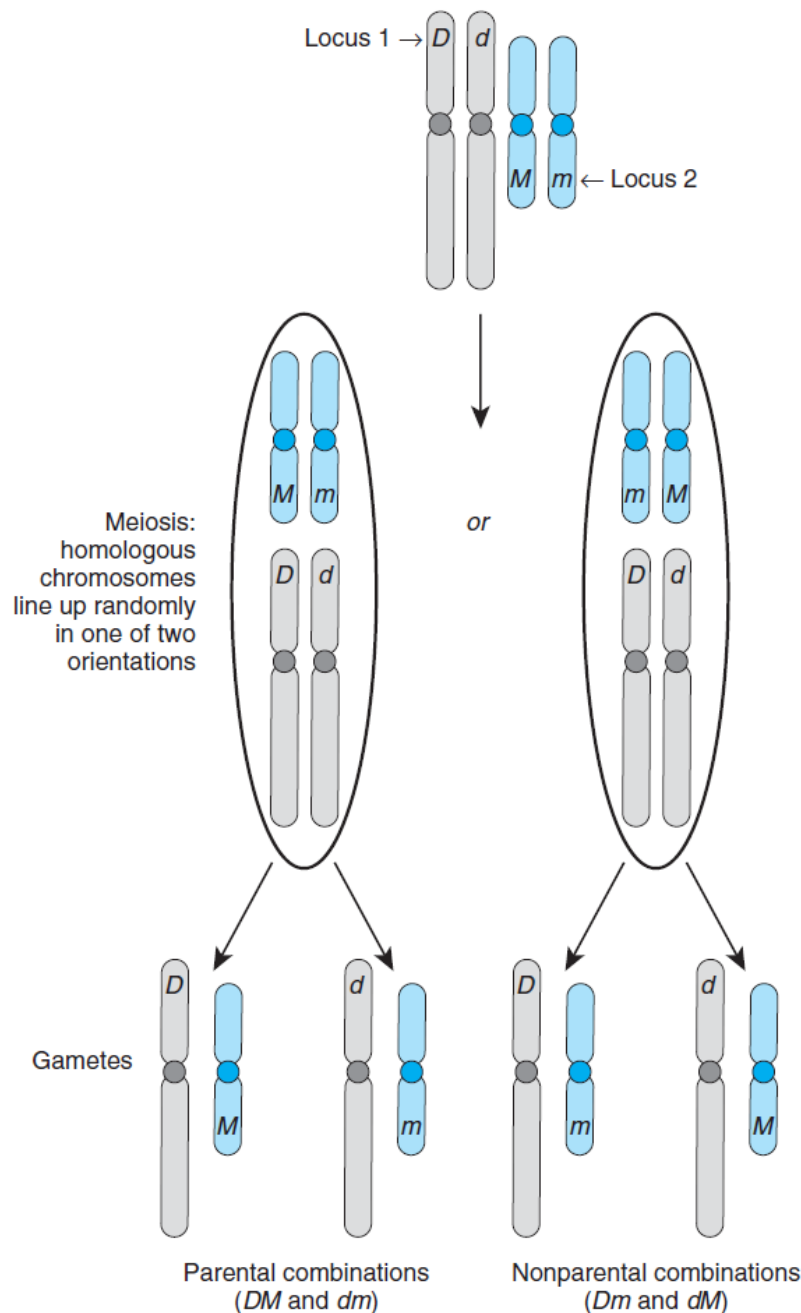


Fig. 14.15



2^{ος} Νόμος του ανεξάρτητου συνδυασμού του Mendel

- Τα αλληλόμορφα διαφορετικών γονιδίων διαχωρίζονται ανεξάρτητα



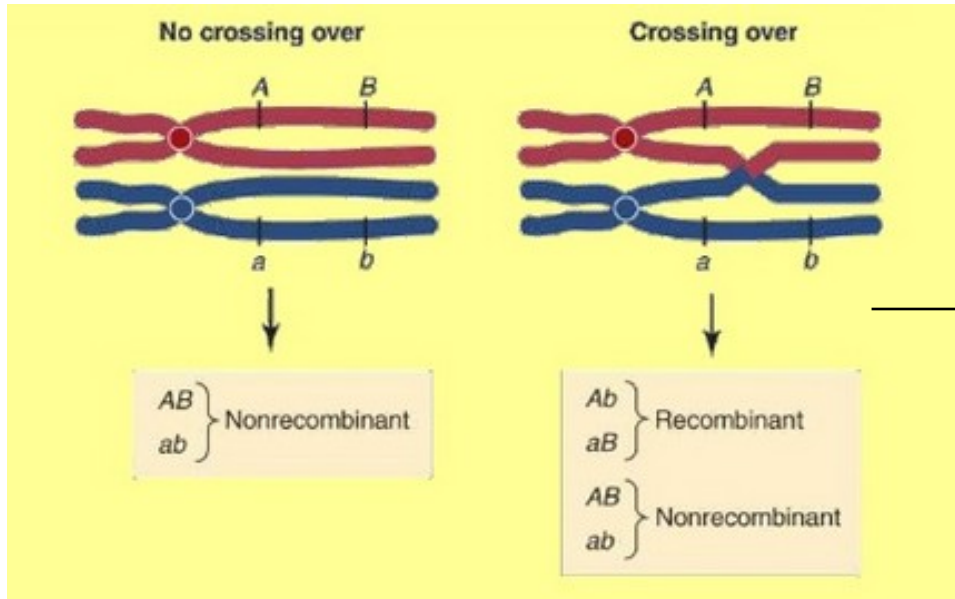
Τα αλληλόμορφα D/d και M/m
χαρτογραφούνται σε **διαφορετικά**
χρωμοσώματα

Figure 10-2 ■ Independent assortment of alleles at two loci, 1 and 2, when they are located on different chromosomes. Assume that alleles D and M were inherited from one parent, d and m from the other.

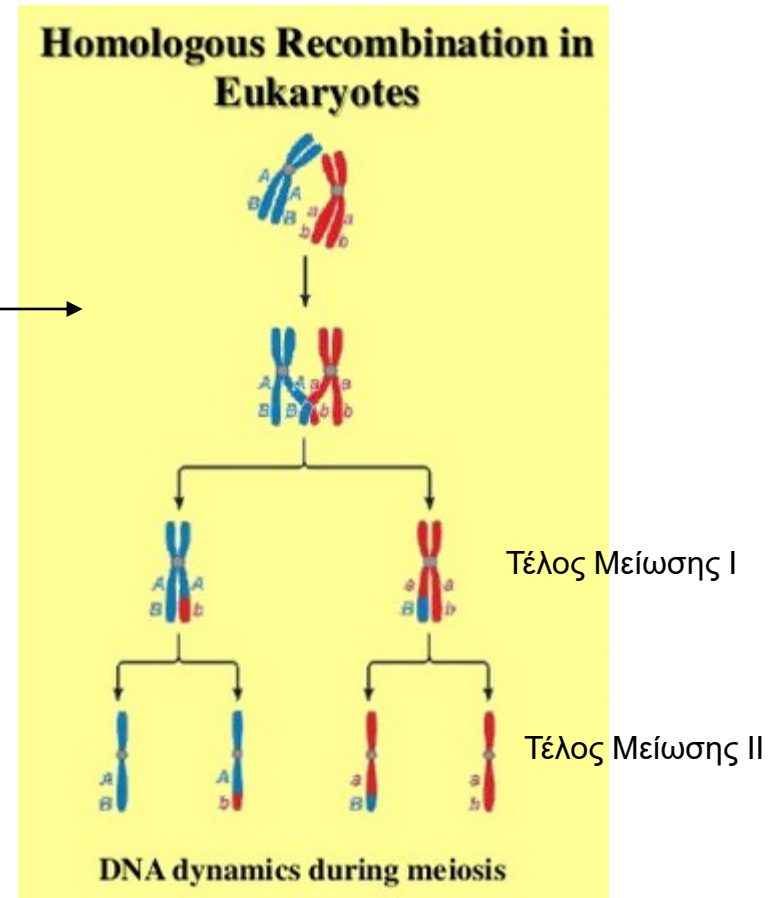
**Δύο γενετικές θέσεις
ή
Δύο γενετικοί τόποι
(loci)**

Τα αλληλόμορφα A/a και B/b χαρτογραφούνται στο **ίδιο** χρωμόσωμα

Ομόλογος ανασυνδυασμός

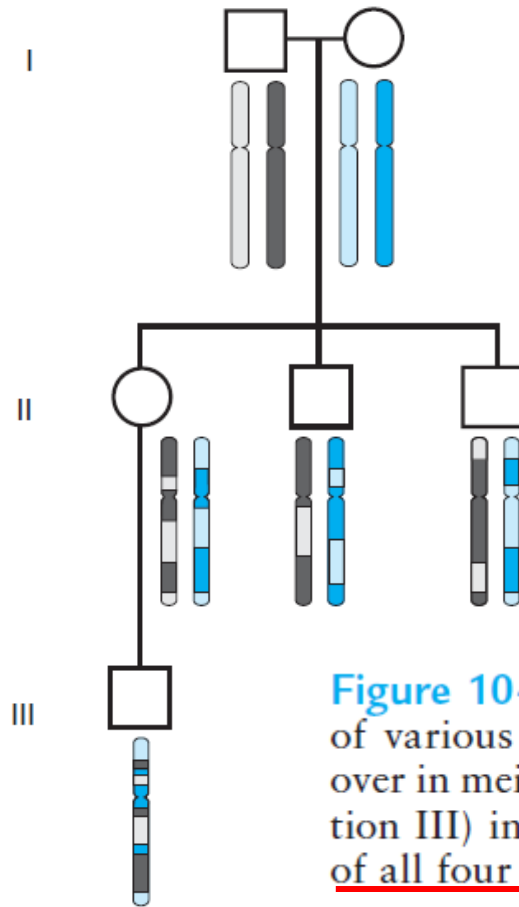


Οι γενετικές θέσεις των αλληλόμορφων A/a και B/b που χαρτογραφούνται στο **ίδιο** χρωμόσωμα μπορεί να απέχουν πολύ ή λίγο μεταξύ τους



Ομόλογος ανασυνδυασμός

Σε επίπεδο οικογένειας



Σε επίπεδο πληθυσμού

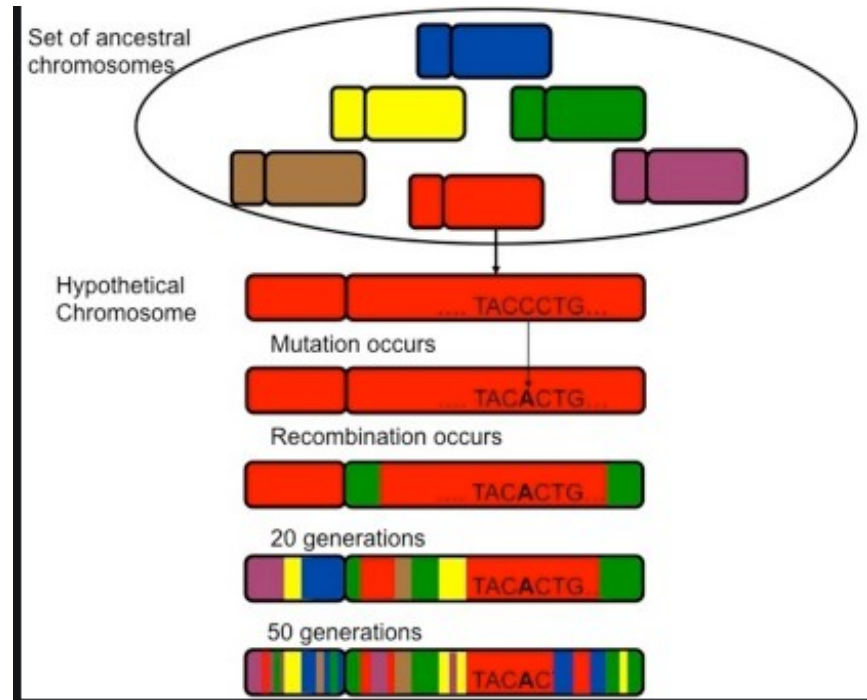


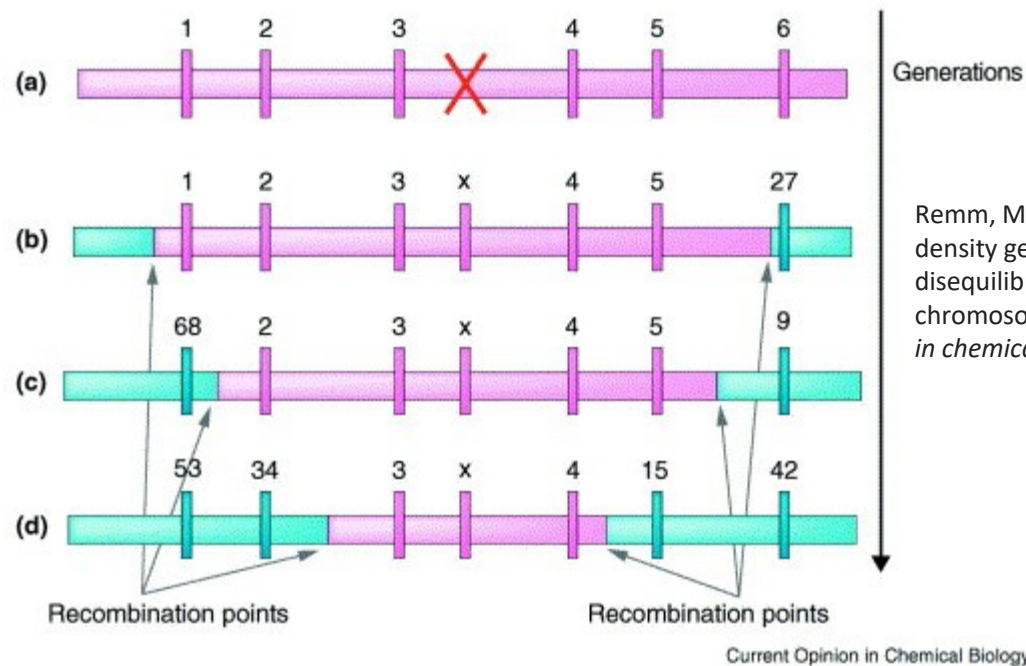
Figure 10-1 ■ The effect of recombination on the origin of various portions of a chromosome. Because of crossing over in meiosis, the copy of the chromosome the boy (generation III) inherited from his mother is a mosaic of segments of all four of his grandparents' copies of that chromosome.

Ανισορροπία σύνδεσης *Linkage disequilibrium (LD)*

Η μη τυχαία συσχέτιση αλληλομόρφων σε διαφορετικούς γενετικούς τόπους επάνω στο **ίδιο χρωμόσωμα!!!**

Απόκλιση του νόμου του ανεξάρτητου συνδυασμού των αλληλομόρφων του Mendel

Περιγραφή της ανισορροπίας σύνδεσης (LD)

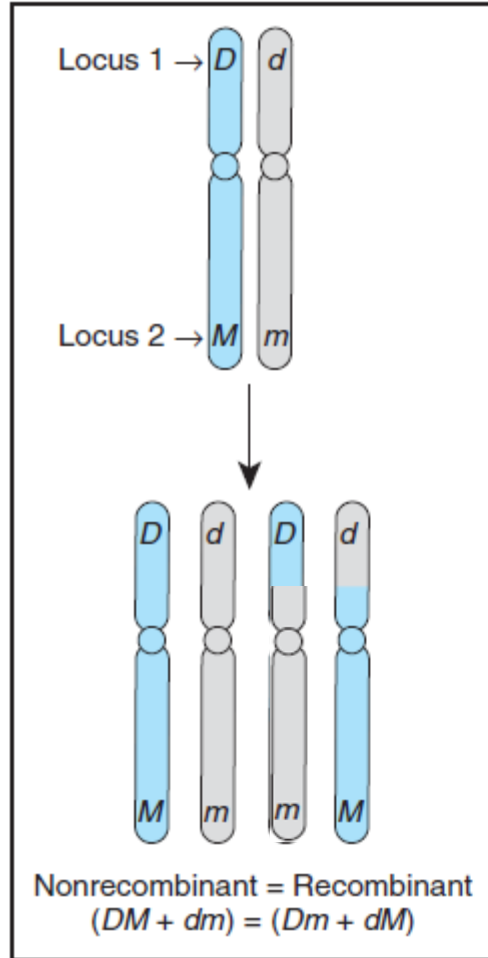


Remm, Maida, and Andres Metspalu. "High-density genotyping and linkage disequilibrium in the human genome using chromosome 22 as a model." *Current opinion in chemical biology* vol. 6,1 (2002): 24-30.

- Κατάσταση λίγο μετά από ένα συμβάν μετάλλαξης (υποδεικνύεται με κόκκινο σταυρό), το οποίο δημιούργησε ένα νέο SNP. Σε αυτό το στάδιο, το νέο αλληλόμορφο βρίσκεται σε πλήρη ανισορροπία σύνδεσης (LD) με τα υπόλοιπα αλληλόμορφα 1–6 στο ίδιο χρωμόσωμα.
- Η κατάσταση μεταβάλλεται καθώς περνούν οι γενιές. Η έκταση της LD μειώνεται λόγω του ανασυνδυασμού μεταξύ των χρωμοσωμάτων. Ο ανασυνδυασμός αναδιατάσσει τα αλληλόμορφα των δεικτών, με αποτέλεσμα να χάνουν τη συσχέτισή τους με το μεταλλαγμένο αλληλόμορφο (x).
- Η LD μεταξύ της μετάλλαξης (x) και των γειτονικών δεικτών παρατηρείται πλέον μόνο σε μια μικρή περιοχή γύρω από τη μετάλλαξη.

$$\theta = 0.5$$

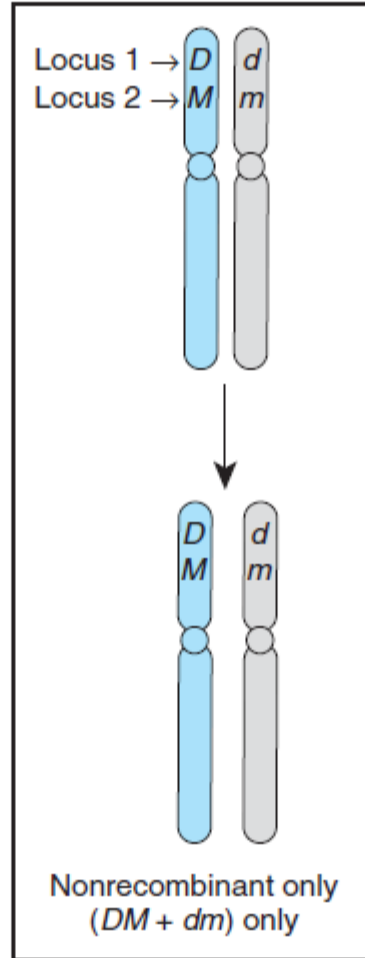
(ανεξάρτητος διαχωρισμός)
(ισορροπία σύνδεσης)



A

$$\theta \sim 0$$

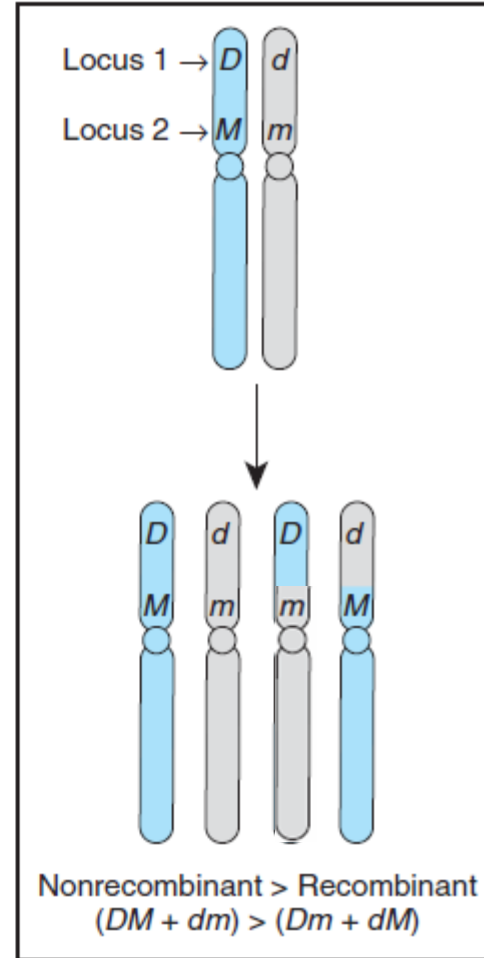
(τέλεια ανισορροπία σύνδεσης)



B

$$0.01 \leq \theta < 0.5$$

(μερική/ατελής
ανισορροπία σύνδεσης)



C

Τα αλληλόμορφα D/d και M/m χαρτογραφούνται στο **ίδιο** χρωμόσωμα

$$\theta = \frac{\text{ανασυνδυασμένοι}}{\text{σύνολο}}$$

Figure 10-4 ■ Assortment of alleles at two loci, 1 and 2, when they are located on the same chromosome. A, The loci are far apart and at least one crossover between them is likely to occur in every meiosis. B, The loci are so close together that crossing over between them is very unlikely. C, The loci are close together on the same chromosome but far enough apart that crossing over occurs in the interval between the two loci only in some meioses and not in others.

Όσο πιο μικρό είναι το ποσοστό ανασυνδυασμού (θ), τόσο πιο κοντά βρίσκονται οι δύο γενετικοί τόποι (εδώ τα D/d , M/m) $\rightarrow$ όπου $0 < \theta < 0.5$

Υπενθύμιση

Νόμος του ανεξάρτητου συνδυασμού του Mendel

Τα αλληλόμορφα D/d και M/m χαρτογραφούνται σε **διαφορετικά** χρωμοσώματα

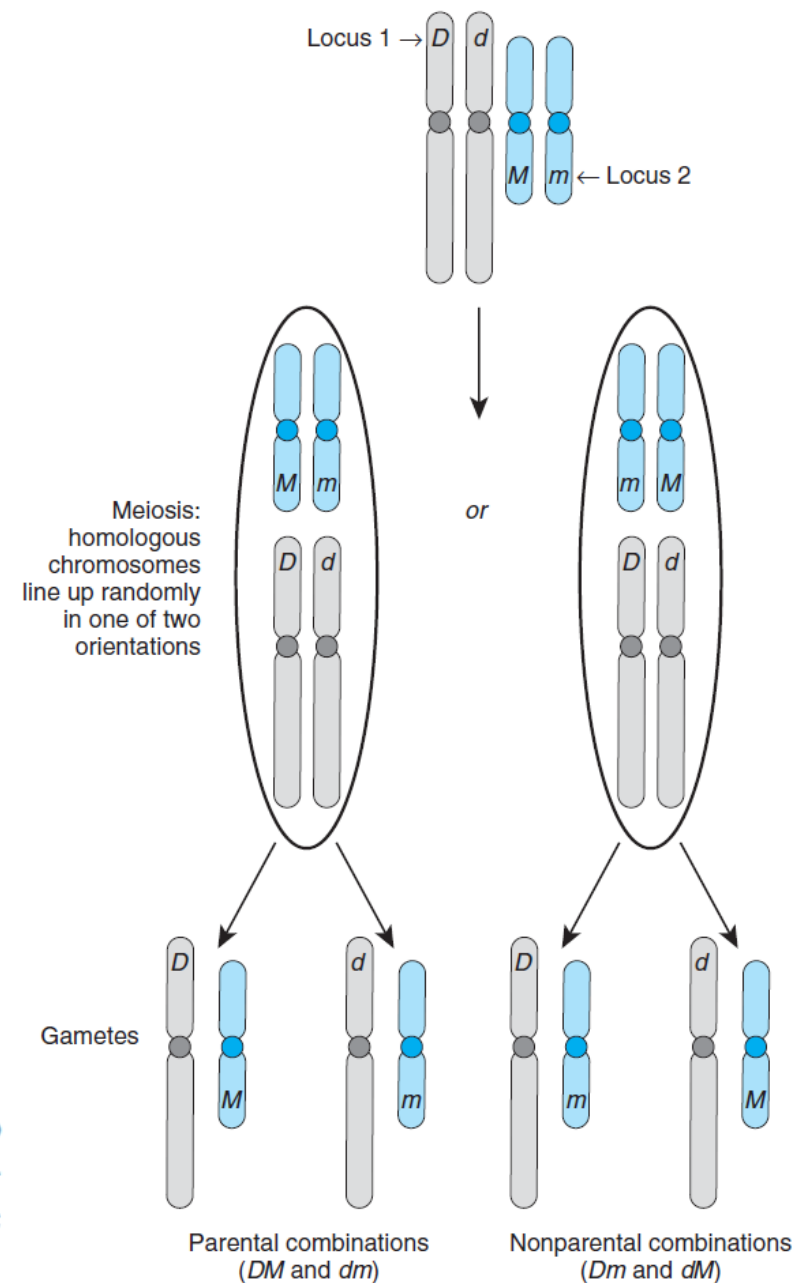


Figure 10-2 ■ Independent assortment of alleles at two loci, 1 and 2, when they are located on different chromosomes. Assume that alleles D and M were inherited from one parent, d and m from the other.

Ανασυνδυασμός

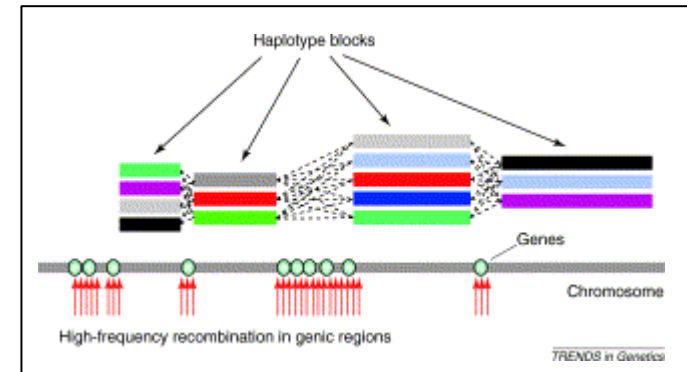
- θ (ή c ή r) = κλάσμα ανασυνδυασμού (recombination fraction)
- θ = το ποσοστό ανασυνδυασμένων γαμετών που παράγονται από έναν διπλό ετεροζυγώτη
- Τιμές θ από 0 ως 0.5
 - $\theta=0$ πλήρης σύνδεση
 - $\theta=0.5$ ανεξάρτητος διαχωρισμός
 - $\theta=0.01$ δηλ. 1/100 γαμέτες είναι ανασυνδυασμένοι
(απόσταση 1cM ανάμεσα στους δύο γενετικούς τόπους)
- Τόποι που βρίσκονται σε απόσταση $>50\text{cM}$ διαχωρίζονται ανεξάρτητα

Επίδραση του ανασυνδυασμού στη μη τυχαία συσχέτιση των αλληλομόρφων

- Ο ανασυνδυασμός ανάμεσα σε δύο γενετικούς τόπους τείνει να τυχαιοποιεί τους συνδυασμούς των αλληλομόρφων από γενιά σε γενιά.
- Ο ρυθμός αυτής της διαδικασίας εξαρτάται από τον ρυθμό ανασυνδυασμού ανάμεσα σε αυτούς τους τόπους.

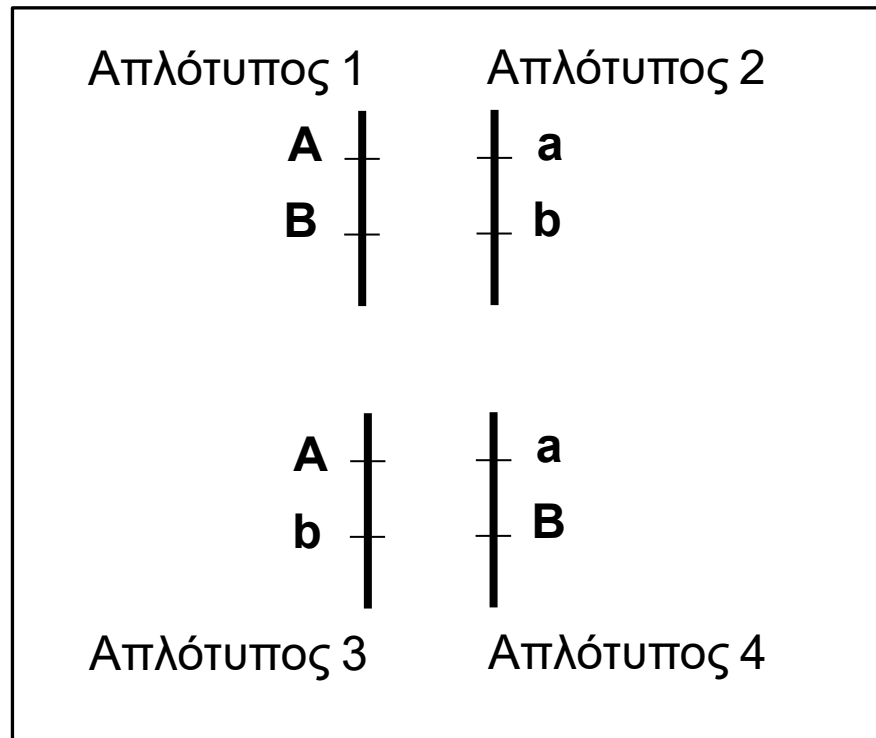
Δεν γνωρίζουμε ακριβώς τη βιολογική βάση του γιατί ορισμένες περιοχές του γονιδιώματος εμφανίζουν μεγάλη συχνότητα ανασυνδυασμού (60% hot-spots μήκους 1-2kb και υποτελομερικές περιοχές), ενώ άλλες εξαιρετικά χαμηλή (πχ κεντρικά τμήματα χρωμοσωμάτων και ετεροχρωματινικές περιοχές).

- Τα θήλεα άτομα 39.1 επιχιασμοί ανά μείωση
- Τα άρρενα άτομα 25.2 επιχιασμοί ανά μείωση
- Το ίδιο άτομο δεν έχει τον ίδιο αριθμό επιχιασμών ανά μείωση!



Απλότυποι

Γονότυπος: AaBb, Δυνατοί απλότυποι



Γονότυποι και Απλότυποι

Ένας γενετικός τόπος:

Homozygous wild type	AA
Homozygous alternative	aa
Heterozygous	Aa

Δύο γενετικοί τόποι:

<u>Locus 1</u>	<u>Locus 2</u>	<u>Απλότυποι:</u>
AA	BB	AB
		Ab
aa	bb	aB
		ab
Aa	Bb	

Πολλαπλοί γενετικοί τόποι:

<u>Locus 1</u>	<u>Locus 2</u>	<u>Locus 3</u>	<u>Απλότυποι:</u>
AA	BB	CC	ABC
			ABc
aa	bb	cc	AbC
			Abc
Aa	Bb	Cc	aBC
			aBc
			abC
			abc

Απλότυπος (haplotype)

Ένα σύνολο αλληλομόρφων των γονιδίων ή γενετικών τόπων σε έναν οργανισμό που κληρονομούνται μαζί από ένα γονέα

- Π.χ. ισχύει για συνδεδεμένα γονίδια/γενετικούς τόπους που βρίσκονται κοντά, πάνω στο ίδιο χρωμόσωμα
- Π.χ. οι πολυμορφισμοί μιτοχονδριακού και χλωροπλαστικού DNA, καθώς και του DNA του Υ χρωμοσώματος (μη ψευδοαυτοσωμικές περιοχές PAR1 & PAR2) που κληρονομούνται από ένα γονέα χωρίς ανασυνδυασμό

Παραδείγματα άλλων χρήσεων του όρου:

- Ένα σύνολο από παραλλαγές (πολυμορφισμοί), π.χ. SNPs, σε ένα χρωμόσωμα που τείνουν να εμφανίζονται μαζί (είναι στατιστικά σχετιζόμενες)
 - π.χ. έρευνα ασθενειών στον άνθρωπο μέσα από το πρόγραμμα HarMap
- Ένα σύνολο συγκεκριμένων παραλλαγών μέσα σε ένα τμήμα DNA σε ένα άτομο

Απλότυποι πολυμορφικών θέσεων

- Συνδυασμοί αλληλομόρφων διαφορετικών παραλλαγών σε ένα χρωμόσωμα (μικροδορυφόροι, SNPs, κτλ)

3 διαλληλικοί (biallelic) πολυμορφισμοί $\longrightarrow 2^3 = 8$ δυνατοί απλότυποι

7 διαλληλικοί πολυμορφισμοί $\longrightarrow 2^7 = 128$ δυνατοί απλότυποι

- Αν υπάρχουν λιγότεροι απλότυποι από ότι αναμένεται σύμφωνα με το νόμο του ανεξάρτητου συνδυασμού, τότε οι παραλλαγές που μελετάμε είναι σε **ανισορροπία σύνδεσης** [Linkage Disequilibrium ή **LD**]

Παράδειγμα με δείκτες

Για 3 δείκτες υπάρχουν $2^3 = 8$ δυνατοί
απλότυποι – **ΘΕΩΡΗΤΙΚΑ !**

Πόσοι απλότυποι θα παρατηρηθούν
σε ένα δείγμα; - **ΠΡΑΚΤΙΚΑ ?**

Εξαρτάται από:

- την ανισορροπία σύνδεσης ανάμεσα στις θέσεις
- τον υπό μελέτη πληθυσμό
- το μέγεθος του πληθυσμιακού δείγματος
- τις συχνότητες των αλληλομόρφων στον δεδομένο πληθυσμό



1.	+	+	+
2.	+	+	-
3.	+	-	+
4.	-	+	+
5.	+	-	-
6.	-	+	-
7.	-	-	+
8.	-	-	-

Ας μελετήσουμε δύο γενετικούς τόπους A και B με αλληλόμορφα

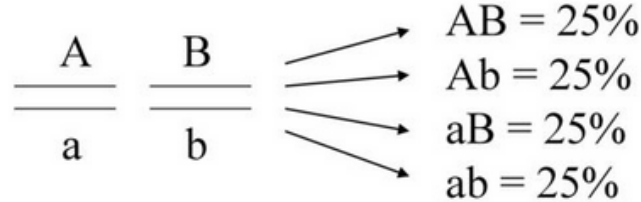
Αλληλόμορφα Συχνότητα

A =	A_1	p_1	$p_1 + p_2 = 1$
a =	A_2	p_2	
B =	B_1	q_1	$q_1 + q_2 = 1$
b =	B_2	q_2	

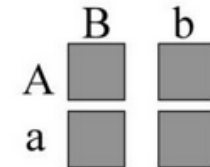
Υπάρχουν 4 πιθανοί γαμέτες

	Γαμέτες	Συχνότητα απλοτύπου	
AB =	$A_1 B_1$	x_1	
Ab =	$A_1 B_2$	x_2	
aB =	$A_2 B_1$	x_3	$x_1 + x_2 + x_3 + x_4 = 1$
ab =	$A_2 B_2$	x_4	

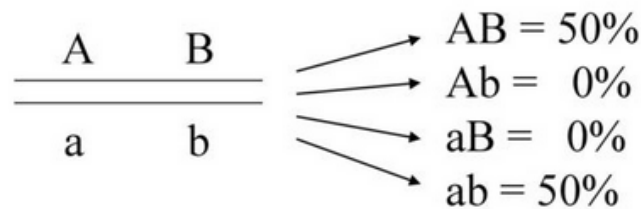
Ισορροπία Σύνδεσης
(linkage equilibrium)



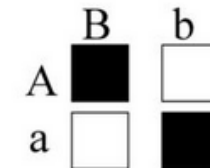
$$f(Ab) = f(A) \times f(b)$$



Ανισορροπία Σύνδεσης
(linkage disequilibrium - LD)
(πλήρης)



$$f(Ab) \neq f(A) \times f(b)$$



Linkage Equilibrium

Τυχαία συσχέτιση αλληλομόρφων στους γαμέτες

Τόπος Β

Τόπος Α

	B_1 q_1	B_2 q_2
A_1 p_1	A_1B_1 $x_1 = p_1 q_1$	A_1B_2 $x_2 = p_1 q_2$
A_2 p_2	A_2B_1 $x_3 = p_2 q_1$	A_2B_2 $x_4 = p_2 q_2$

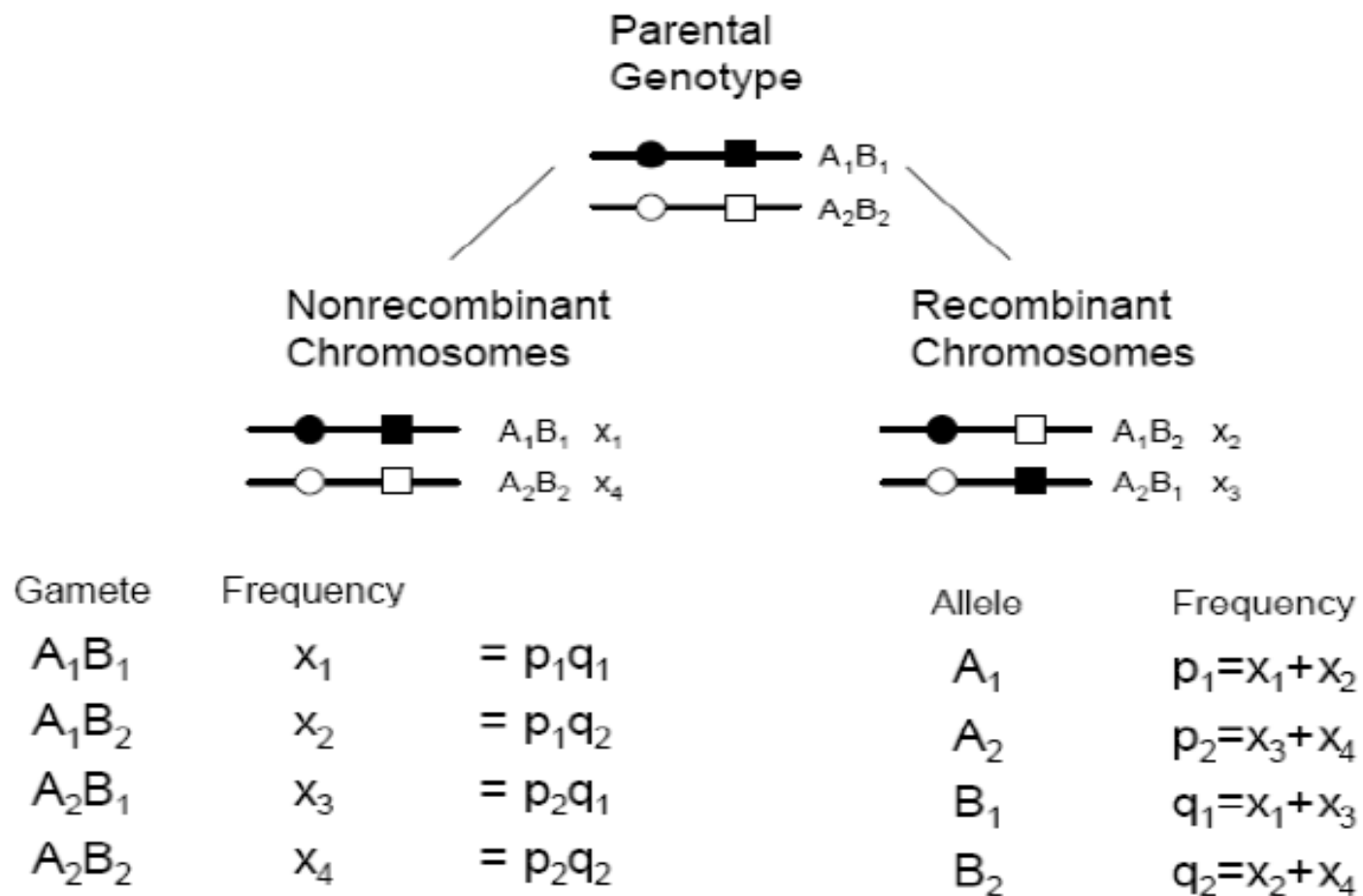
$$x_1 + x_2 + x_3 + x_4 = 1$$

Ισορροπία Σύνδεσης

	SNP#1	SNP#2	Allele frequencies	Expected haplotype frequencies
possible haplotypes (combinations)	C	A	SNP#1 { C=0.3 T=0.7	CA = 0.3 x 0.1 = 0.03
	C	G		CG = 0.3 x 0.9 = 0.27
	T	A	SNP#2 { A=0.1 G=0.9	TA = 0.7 x 0.1 = 0.07
	T	G		TG = 0.7 x 0.9 = 0.63

Στο δείγμα μας παρατηρούμε όλους τους δυνατούς απλότυπους, με ποικίλες συχνότητες

Linkage Equilibrium



Πλήρης Ανισορροπία Σύνδεσης

Παράδειγμα για 2 SNPs:

SNP1: A 50% C 50%

SNP2: A 50% G 50%

SNP1	SNP2
A	A
A	G
C	A
C	G



Αναμενόμενη Συχνότητα

$$p(AA) = 0,5 \times 0,5 = 25\%$$

$$p(AG) = 0,5 \times 0,5 = 25\%$$

$$p(CA) = 0,5 \times 0,5 = 25\%$$

$$p(CG) = 0,5 \times 0,5 = 25\%$$

Στο δείγμα μας μπορεί να **παρατηρούμε** μόνο δύο απλότυπους

A A C C **A** T A T C ... C **G** A T T ... $\rightarrow p(AG)=0.5$ ή 50%

A A C C **C** T A T C ... C **A** A T T ... $\rightarrow p(CA)=0.5$ ή 50%

Παράμετρος ανισορροπίας σύνδεσης, D

Lewontin and Kojima (1960) Evolution 14:450-472.

Γαμέτης	Συχνότητα	
A_1B_1	x_1	$= p_1q_1 + D$
A_1B_2	x_2	$= p_1q_2 - D$
A_2B_1	x_3	$= p_2q_1 - D$
A_2B_2	x_4	$= p_2q_2 + D$

$$D = x_1 - p_1q_1$$

= αυτό που παρατηρούμε

= αυτό που περιμένουμε σε ισορροπία σύνδεσης

Η τιμή D είναι μια παράμετρος της απόκλισης της συχνότητας του κάθε απλότυπου από το σενάριο της ισορροπίας σύνδεσης

Παράμετρος ανισορροπίας σύνδεσης, D

Αντικαθιστώντας:

$$D = x_1 - p_1 q_1$$

|

$$D = x_1 x_4 - x_2 x_3$$

$$p_1 = x_1 + x_2$$

$$p_2 = x_3 + x_4$$

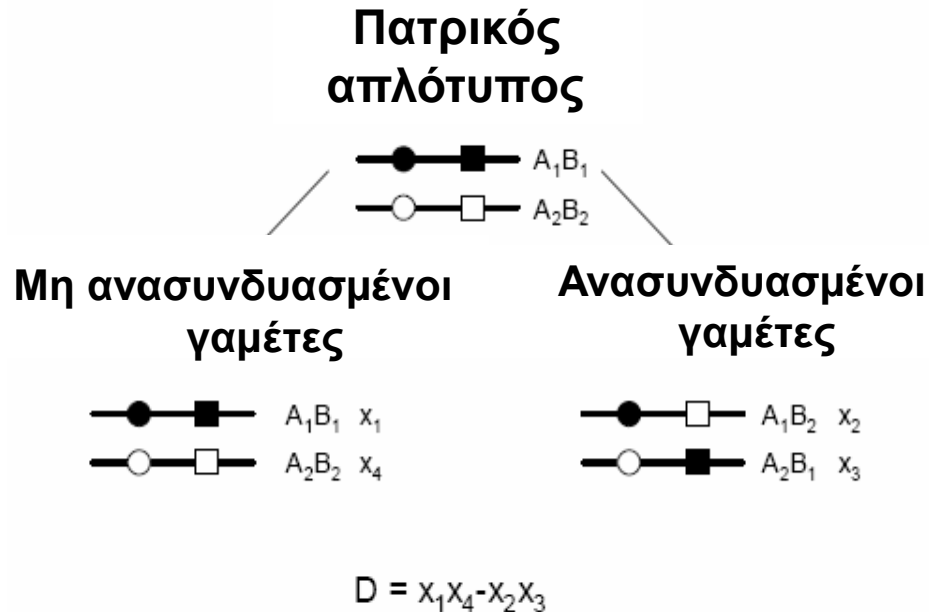
$$q_1 = x_1 + x_3$$

$$q_2 = x_2 + x_4$$

$$x_4 = 1 - x_1 - x_2 - x_3$$

Παράμετρος ανισορροπίας σύνδεσης, D

D: Πόσο αποκλίνουν οι παρατηρούμενες συχνότητες των απλοτύπων από αυτές που θα περιμέναμε αν οι δύο γενετικοί τόποι (A και B) ήταν εντελώς ανεξάρτητοι (δηλαδή σε ισορροπία σύνδεσης).



D = το γινόμενο των συχνοτήτων των μη ανασυνδυασμένων γαμετών μείον το γινόμενο των συχνοτήτων των ανασυνδυασμένων γαμετών

Estimate of D in case of Linkage Equilibrium

If allele frequencies of p_1 and q_1 are both 0.5 and equilibrium occurs

$$x_1 = p_1q_1 = 0.5 \times 0.5 = 0.25$$

$$x_4 = p_2q_2 = 0.5 \times 0.5 = 0.25$$

$$x_2 = p_1q_2 = 0.5 \times 0.5 = 0.25$$

$$x_3 = p_2q_1 = 0.5 \times 0.5 = 0.25$$

Estimate of D in case of Linkage Equilibrium

If allele frequencies of p1 and q1 are both 0.5 and equilibrium occurs

$$D = x_1 * x_4 - x_2 * x_3$$

$$D = (0.25) (0.25) - (0.25) (0.25) = 0$$

Παράδειγμα ανισορροπίας σύνδεσης (linkage disequilibrium)

- Οι συχνότητες των αλληλομόρφων είναι: $p_1 = q_1 = 0.5$ & $p_2 = q_2 = 0.5$
- Στον πληθυσμό παρατηρούνται μόνο οι απλότυποι A_1B_1 και A_2B_2

$$x_1 = p_1q_1 + D = 0.25 + D = 0.5$$

$$x_4 = p_2q_2 + D = 0.25 + D = 0.5$$

$$x_2 = p_1q_2 - D = 0.25 - D = 0$$

$$x_3 = p_2q_1 - D = 0.25 - D = 0$$

} Δεν
παρατηρούνται
στον πληθυσμό!

$$D = (0.5)(0.5) - (0)(0) = 0.25$$

Ιδιότητες του D

Δύο ανεξάρτητοι τόποι έχουν $D = 0$

- Όταν υπάρχει ανισορροπία σύνδεσης έχουμε $D \neq 0$

Το πρόσημο είναι αυθαίρετο

- Λαμβάνει τιμές από -0,25 ως 0,25

Δύσκολη η σύγκριση ανάμεσα σε περισσότερα ζεύγη δεικτών

Η τιμή του εξαρτάται από τις συχνότητες των αλληλομόρφων

- Απαιτείται στάθμιση
- Μπορούμε να διαιρέσουμε το D με τη μέγιστη τιμή που μπορεί να λάβει για τις συγκεκριμένες συχνότητες αλληλομόρφων

Σταθμισμένος δείκτης D'

$$D' = \frac{D}{D_{\max}}$$

Αν $D > 0$, $D_{\max} = \min(p_1 q_2, p_2 q_1)$

Αν $D < 0$, $D_{\max} = \min(p_1 q_1, p_2 q_2)$

$|D'| = 1$, τότε έχουμε απόλυτη ανισορροπία σύνδεσης

- τουλάχιστον ένας απλότυπος δεν παρατηρείται
- $|D'| < 1$, σημαίνει πως η απόλυτη προγονική ανισορροπία σύνδεσης έχει διαταραχθεί

Υπερεκτίμηση της ανισορροπίας σύνδεσης

- Ιδιαίτερα όταν το μέγεθος του δείγματος είναι μικρό
- Όταν ένα τουλάχιστον από τα αλληλόμορφα είναι σπάνιο
 - Μεγάλη πιθανότητα να μην δούμε ένα αλληλόμορφο στο δείγμα από τύχη

Παράμετρος ανισορροπίας σύνδεσης, r^2

Hill and Robertson (1968) Genet Res 8:269-294.

Square of the correlation coefficient.

$$r^2 = \frac{D^2}{p_1 p_2 q_1 q_2}$$

$$0 < r^2 < 1$$

Στατιστική συσχέτιση

$$r^2 = (x_1 \cdot x_4 - x_2 \cdot x_3)^2 / (p_1 \cdot p_2 \cdot q_1 \cdot q_2)$$

- Εκτιμά στατιστική συσχέτιση
- Για χαμηλές συχνότητες αλληλομόρφων, το r^2 είναι πιο αξιόπιστο από το $|D'|$.
- $r^2 = 0 \rightarrow$ Καμία ανισορροπία
- $r^2 = 1 \rightarrow$ Απόλυτη σύνδεση (LD)

Οι παρατηρήσεις σε ένα δείκτη προσφέρουν πλήρεις πληροφορίες για τον άλλο – δεν είναι απαραίτητη η τυποποίηση και των δύο

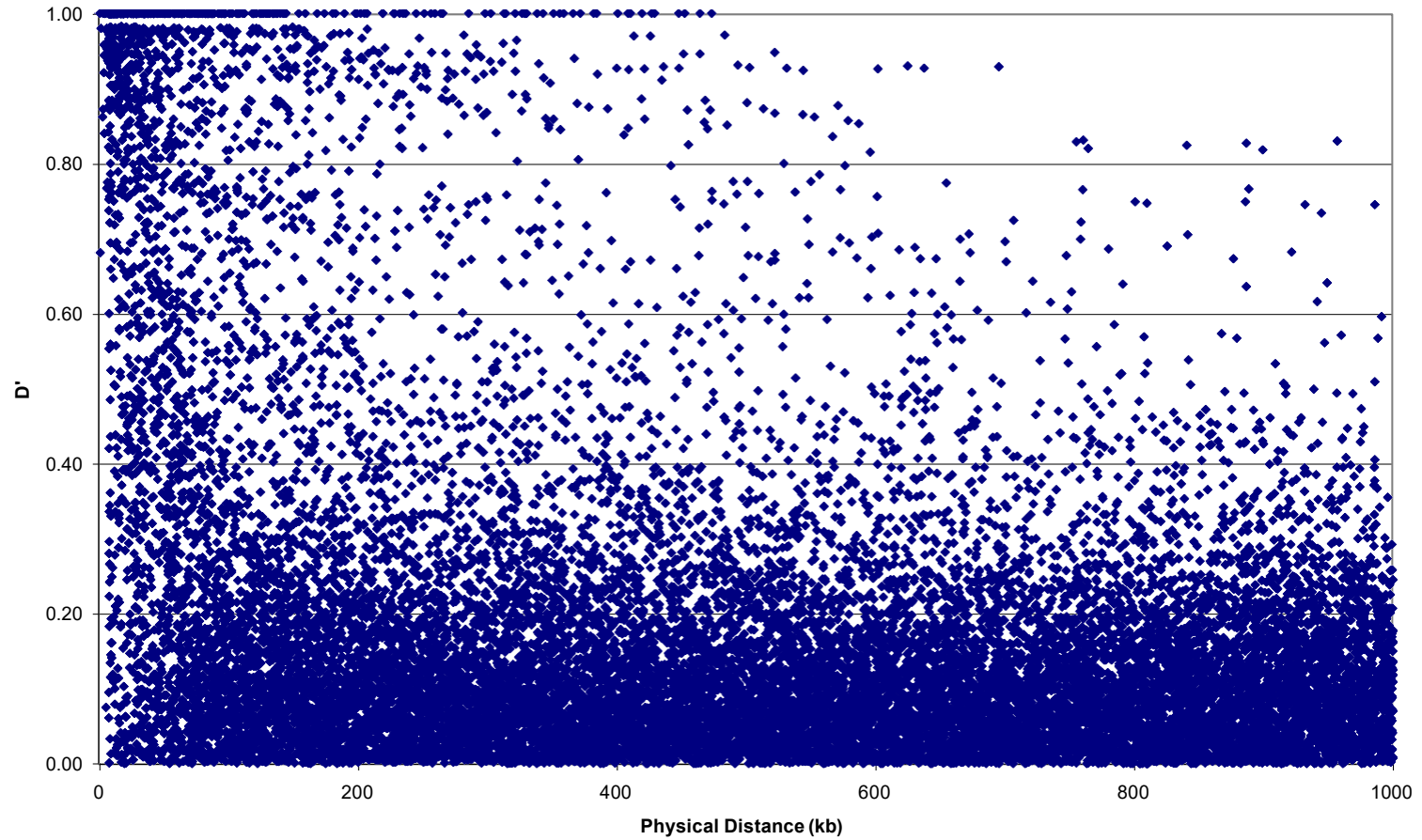
Παράδειγμα:

Αν σε έναν πληθυσμό υπάρχουν μόνο οι απλότυποι A1B1 και A2B2, τότε αν ανιχνεύσω το A1, ξέρω με σιγουριά ότι υπάρχει και το B1

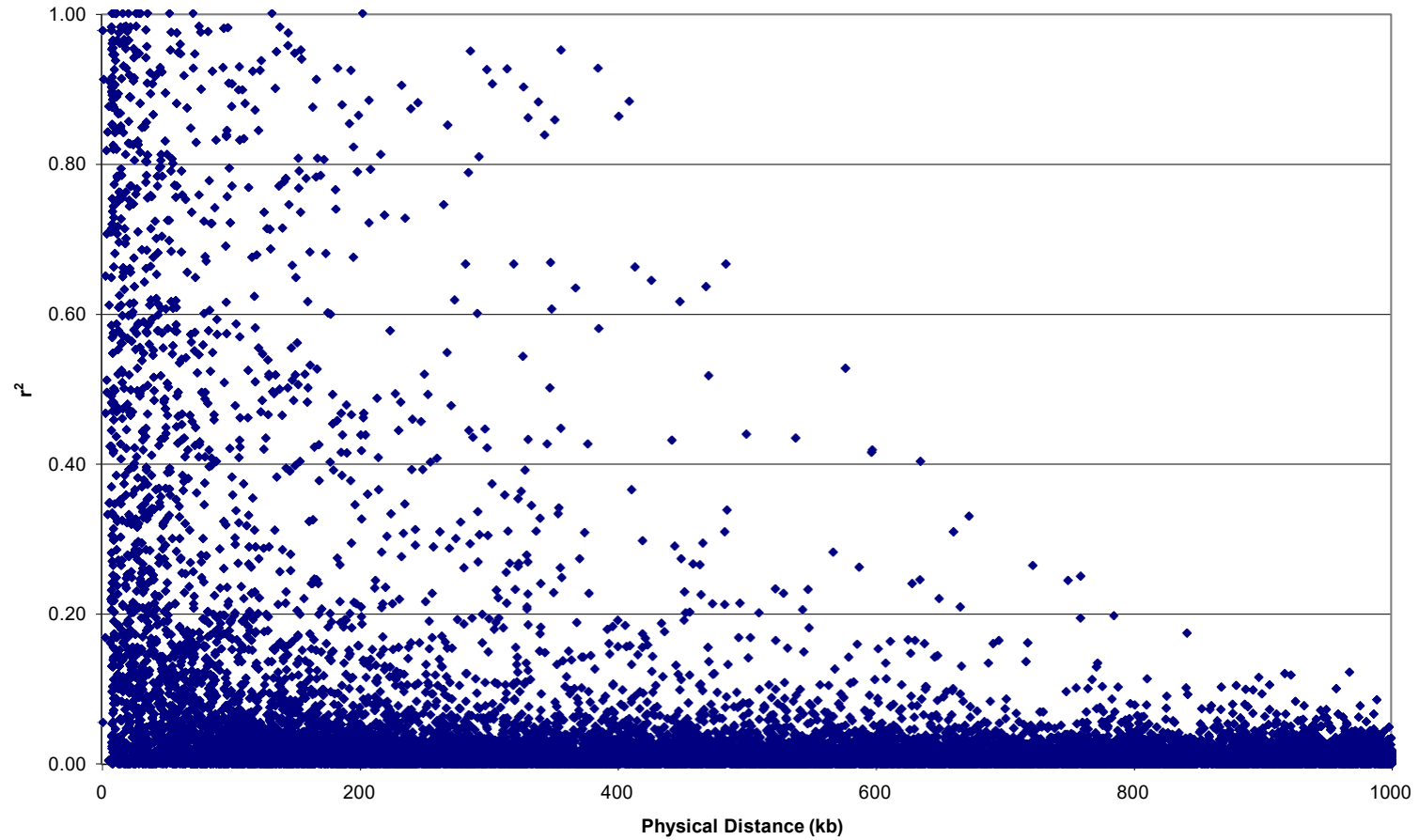
Σύγκριση

- r^2 : εξαρτάται από τις συχνότητες των αλληλομόρφων
- $|D'|$: είναι το μόνο μέτρο LD που δεν επηρεάζεται από αλληλικές συχνότητες
- $|D'|=1$, αν εμφανίζονται 2 ή 3 απλότυποι
- $r^2=1$ εμφανίζονται μόνο 2 απλότυποι – τέλεια LD
- Ενδιάμεσες τιμές του $|D'|$ είναι δύσκολο να ερμηνευτούν
- Αφορούν συγκρίσεις δεικτών ανά ζεύγη μόνο (δύσκολη η ερμηνεία για συγκληρονόμηση >2 δεικτών)

Raw $|D'|$ data from Chr22

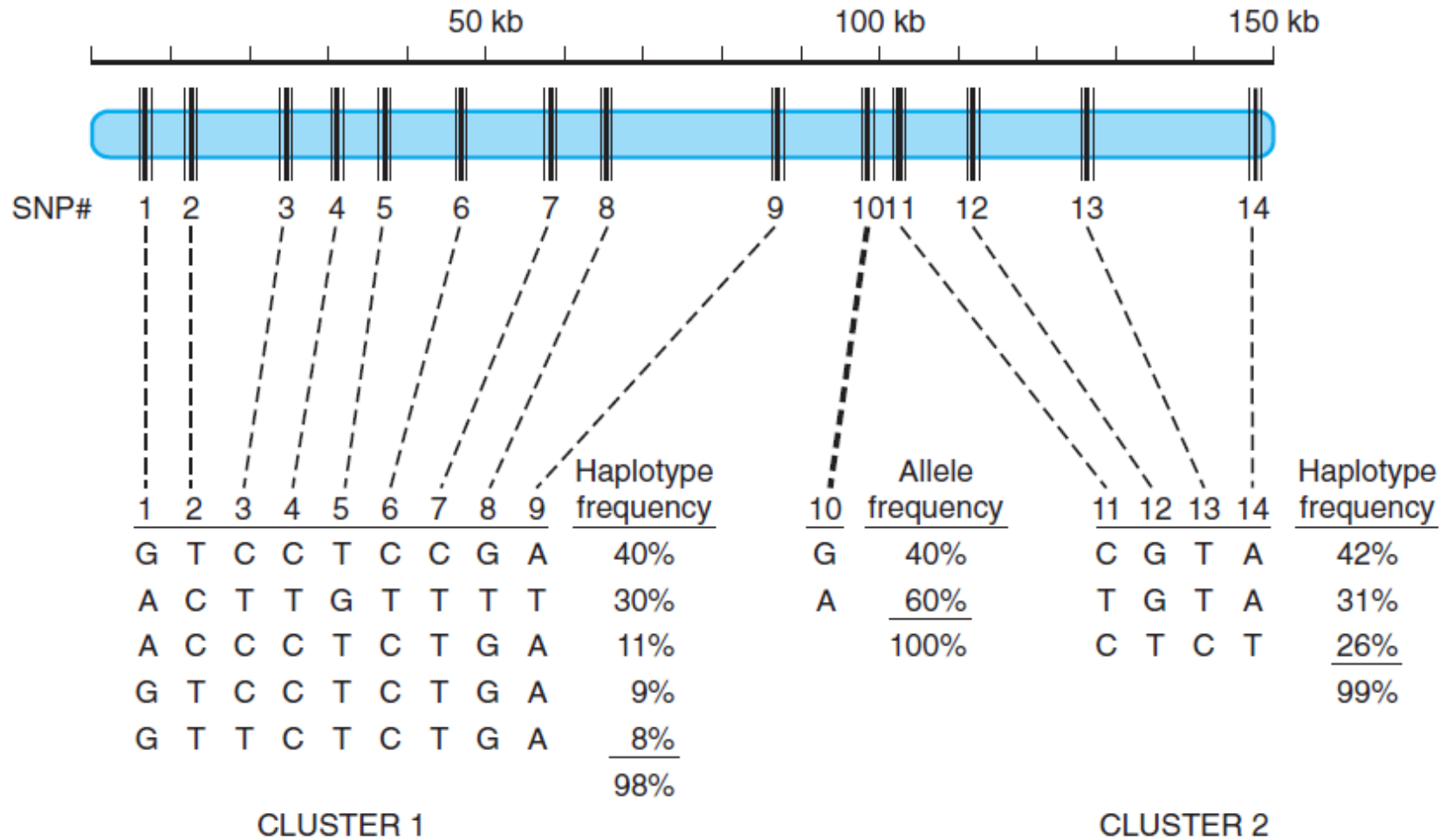


Raw r^2 data from Chr22

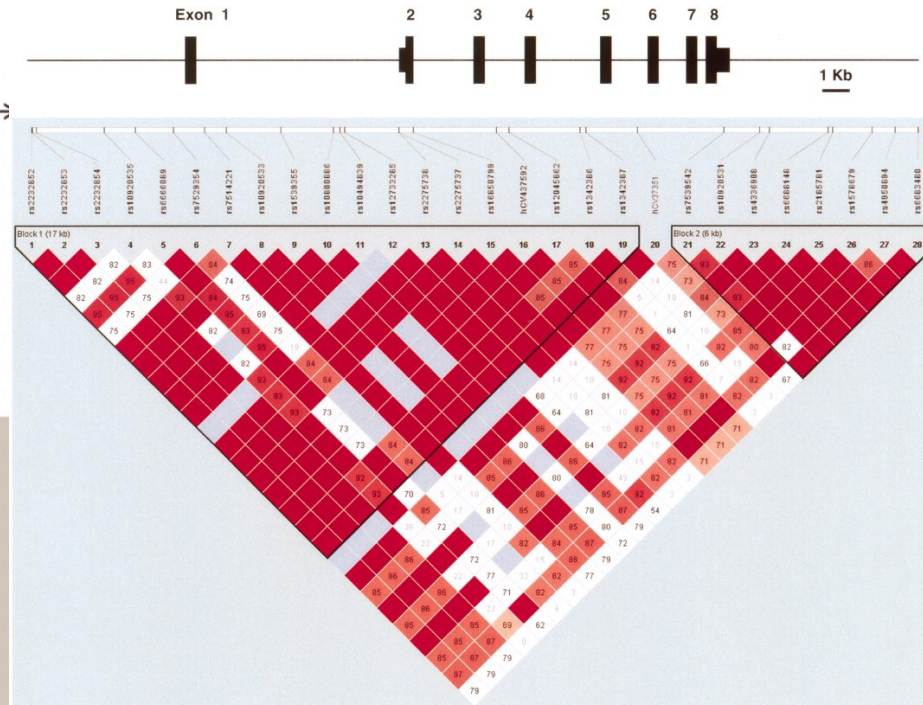
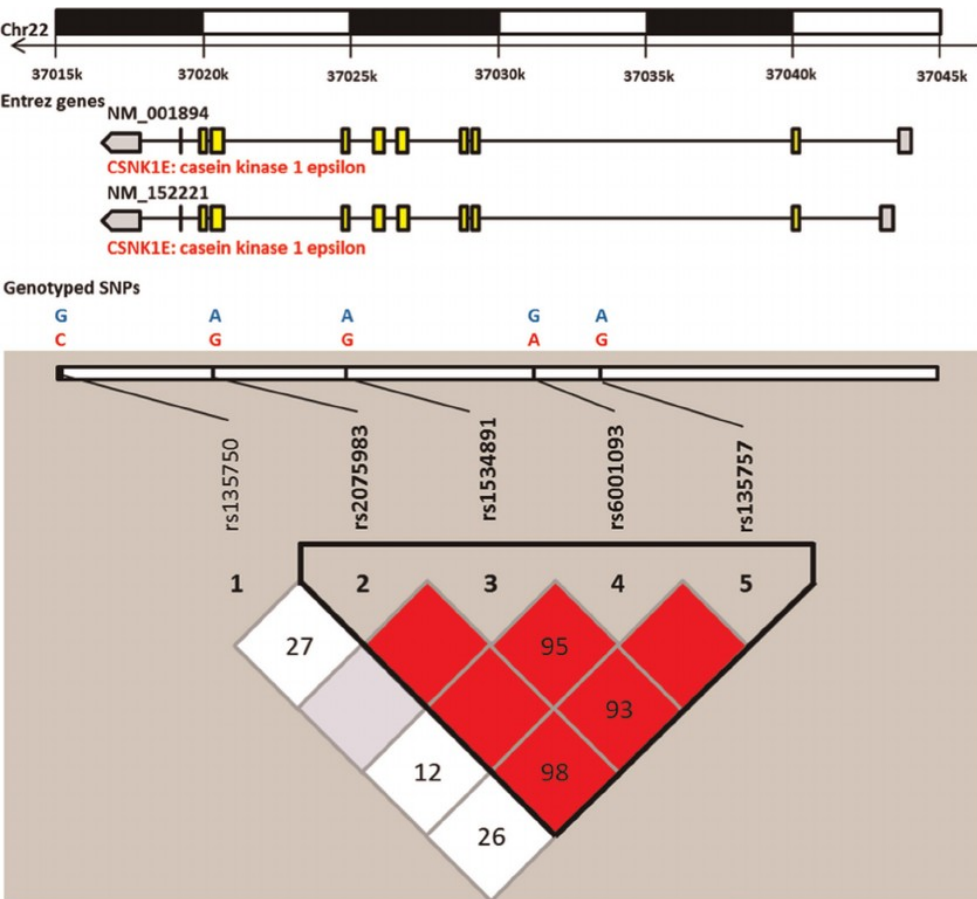


Οπτικοποίηση LD

A



Οπτικοποίηση LD

[illegible]

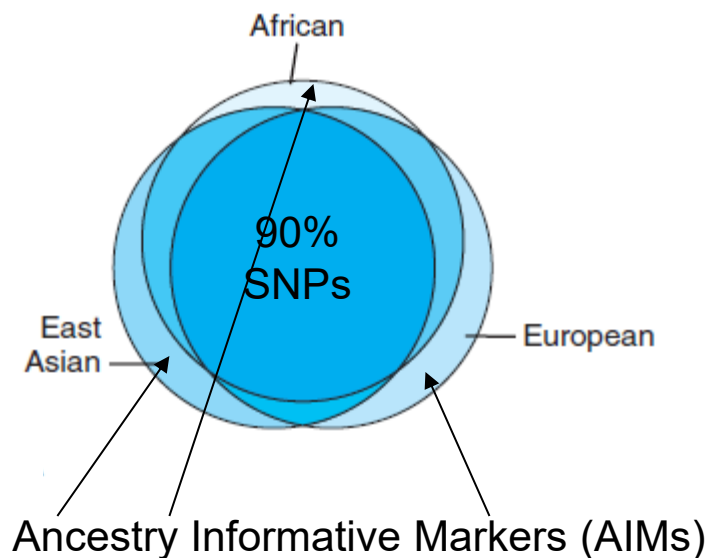
Γεωγραφική προέλευση και LD

Καθώς η LD εξαρτάται από τη γενετική απόσταση όσο και από τη χρονική περίοδο που πρωτοεμφανίστηκε μία παραλλαγή (πολυμορφισμός), πληθυσμοί με διαφορετική ιστορία (πχ Αφρικανοί vs Καυκάσιοι vs Ασιάτες) μπορεί να δείχνουν διαφορετικές τιμές D' μεταξύ των ίδιων δεικτών στο γονιδίωμα:

Africans: Avg LD block = 7.3kb

Asians: Avg LD block = 13.2kb

Caucasians: Avg LD block = 16.3kb

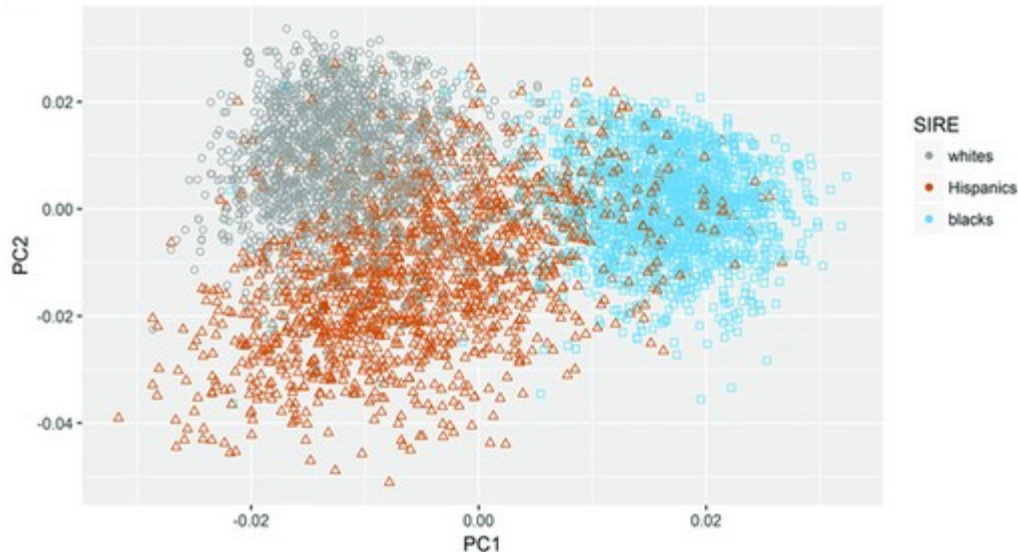


Ancestry Informative Markers – AIMs

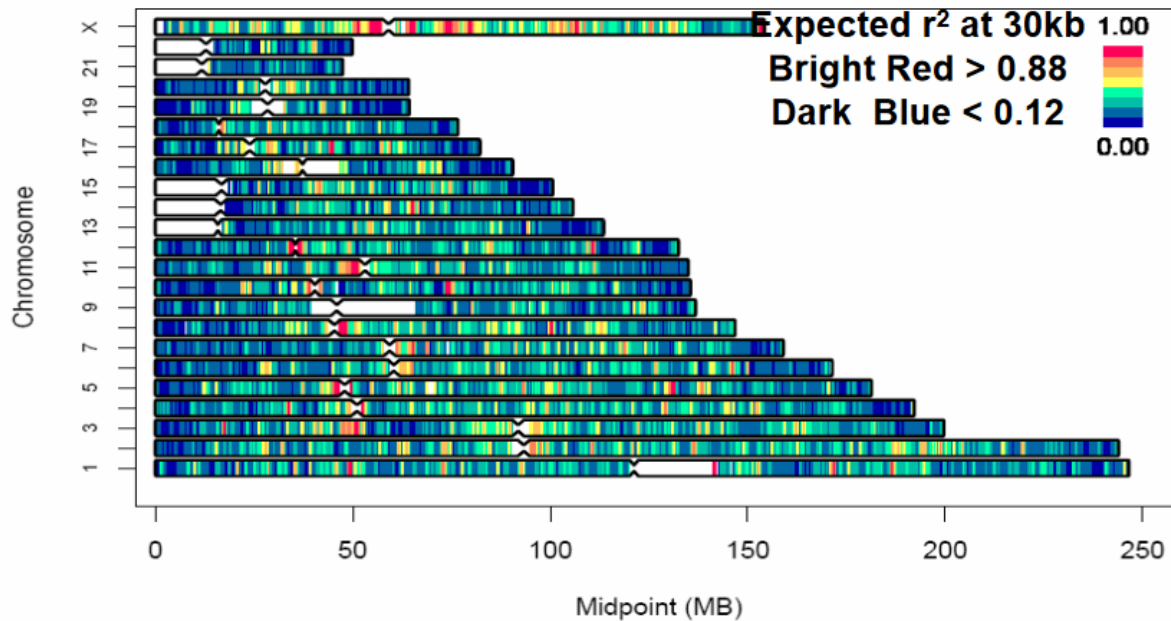
Πληροφοριακοί δείκτες της καταγωγής

Principal Component Analysis (PCA)

Ανάλυση κύριων συνιστωσών

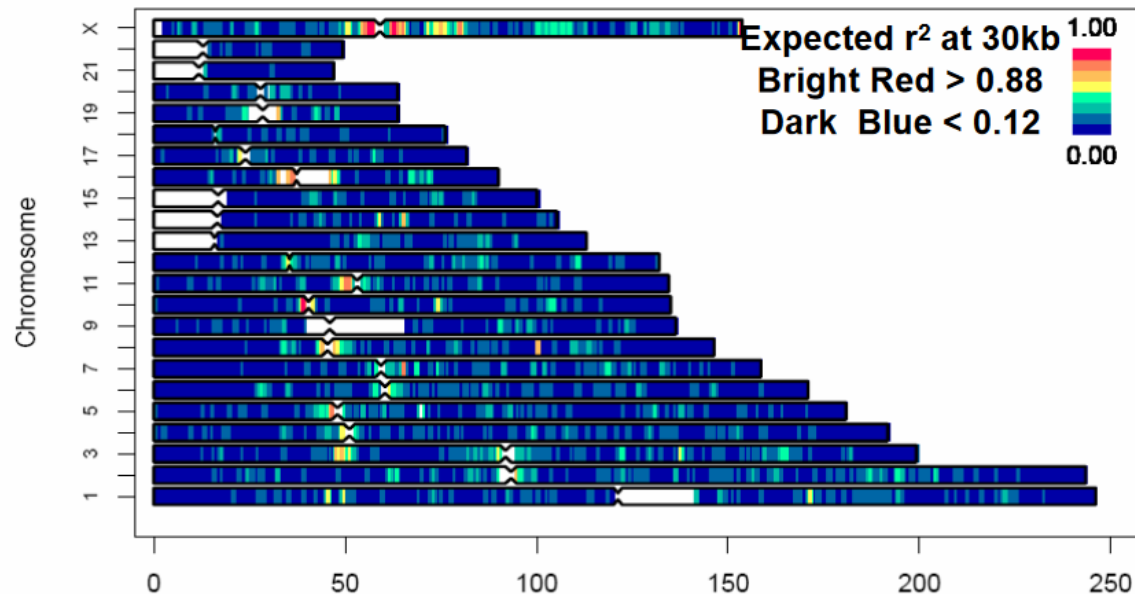


- Σημαντικές διαφορές στη συχνότητα αλληλομόρφων ορισμένων παραλλαγών τύπου SNPs παρατηρήθηκαν μεταξύ διαφορετικών πληθυσμών.
- Αποτέλεσμα γενετικής παρέκκλισης ή του φαινομένου του ιδρυτή ή διαφορές επιλογής σε συγκεκριμένες γεωγραφικές περιοχές/ συνθήκες?
- Χρησιμοποιούνται σε μελέτες καταγωγής του ανθρώπινου είδους, της μετανάστευσης και στην ιατροδικαστική (πληθυσμιακή καταγωγή του δράστη) και τις μελέτες συσχέτισης σε επίπεδο ολικού γονιδιώματος (GWAS) για την ελάττωση των ψευδώς θετικών ευρημάτων



Caucasians
(CEPH
population)

Γεωγραφική
προέλευση
και LD



Το r^2 μειώνεται!

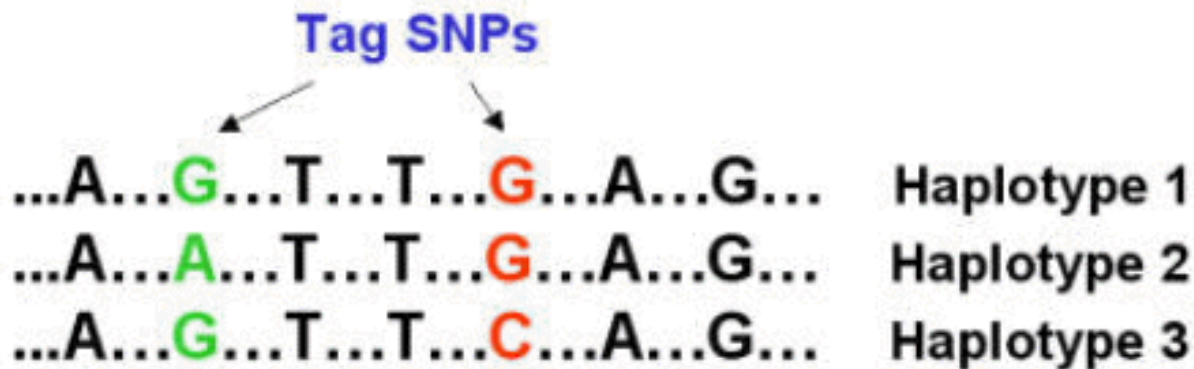
Africans
(Yorubans)

Απλοτυπικές ομάδες – Πού χρησιμεύουν (Haplotype Blocks)

- Ιδανικά θα θέλαμε να μπορούμε να συγκρίνουμε όλα τα υπάρχοντα SNPs ανάμεσα σε δύο ομάδες δειγμάτων (πχ μια μεγάλη ομάδα ασθενών και μια μεγάλη ομάδα controls)
- → Απαγορευτικό κόστος!
- Χάρη στη δομή του ανθρώπινου γονιδιώματος δεν είναι απαραίτητο να γονοτυπήσουμε όλα τα SNPs
- Μπορούμε να επικεντρωθούμε μόνο στα αντιπροσωπευτικά **tagging SNPs (tagSNPs ή tSNPs)**

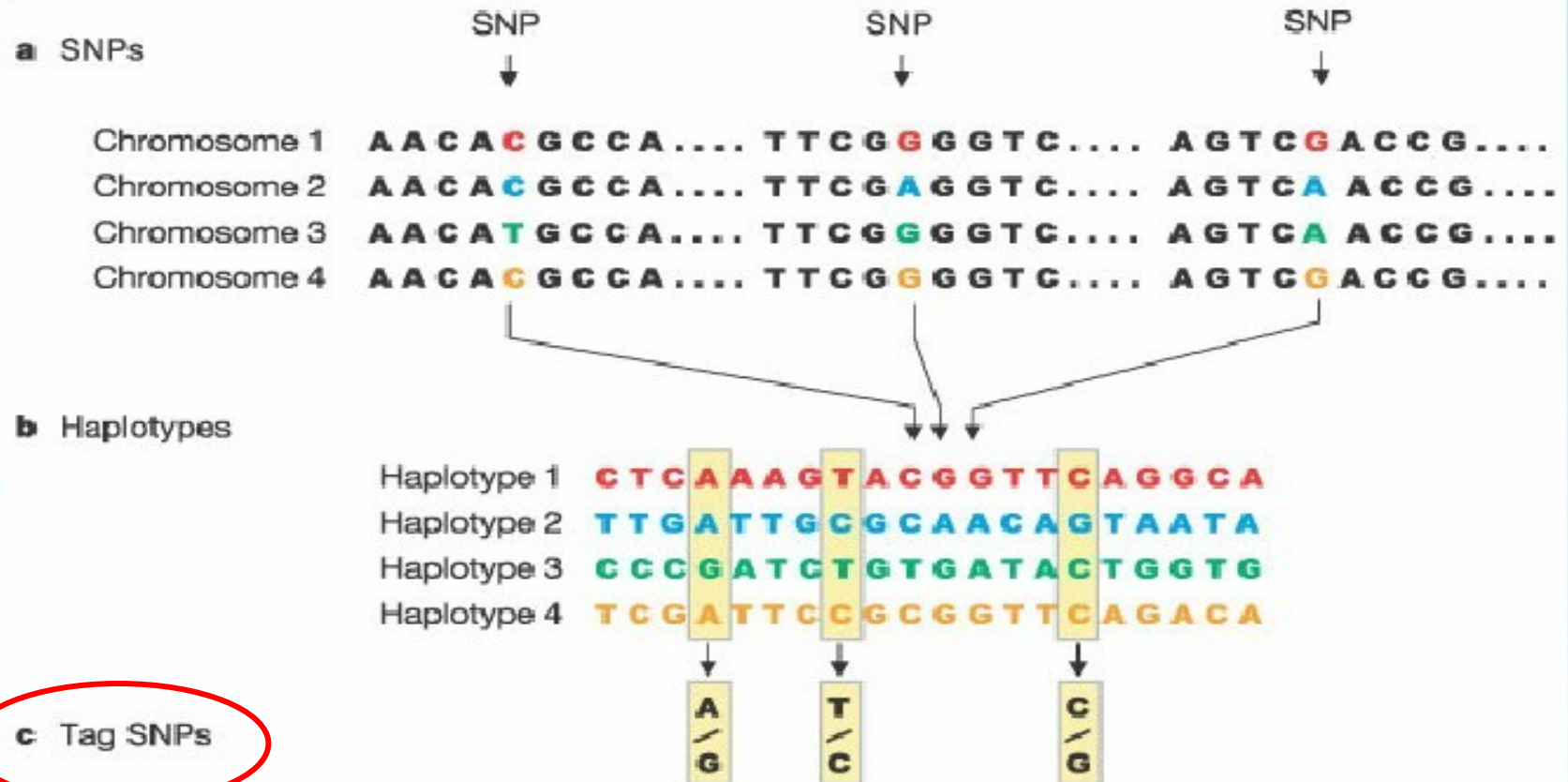
htSNPs-haplotype tagging SNPs

Μια υπο-ομάδα SNPs σε έναν απλότυπο που είναι αρκετά για να καθορίσουν όλα τα υπόλοιπα



if a chromosome has alleles G and C at the two tag SNPs position, then it has the **Haplotype 3**.

Human Sequence Variation



The International HapMap Project, Nature 2003

Συμπεράσματα...

- Πλεονασμός (redundancy) της πληροφορίας των SNPs
- Δυνατή η επιλογή tagging SNPs
- Δυνατή η ανάκτηση (πρόβλεψη) ολόκληρης της πληροφορίας για μία δεδομένη χρωμοσωμική περιοχή αναλύοντας μόνο ένα μικρό αριθμό tSNPs
- Μείωση κόστους και χρόνου
- Η ανάγκη να χρησιμοποιούμε tSNPs έχει μειωθεί δραστικά γιατί πλέον μπορούμε να δούμε "τα πάντα" με χαμηλότερο κόστος και υψηλότερη ακρίβεια (NGS).

Testing for statistical *association* between two *known* SNPs

In this example: $D=0.0-0.21 = -0.21$ and $D'=-1$ [SNP1 and SNP2 are in LD]

D_{max} = η μικρότερη τιμή των p_1q_2 και p_2q_1 όταν $D>0$

D_{max} = η μικρότερη τιμή των p_1q_1 και p_2q_2 όταν $D<0$

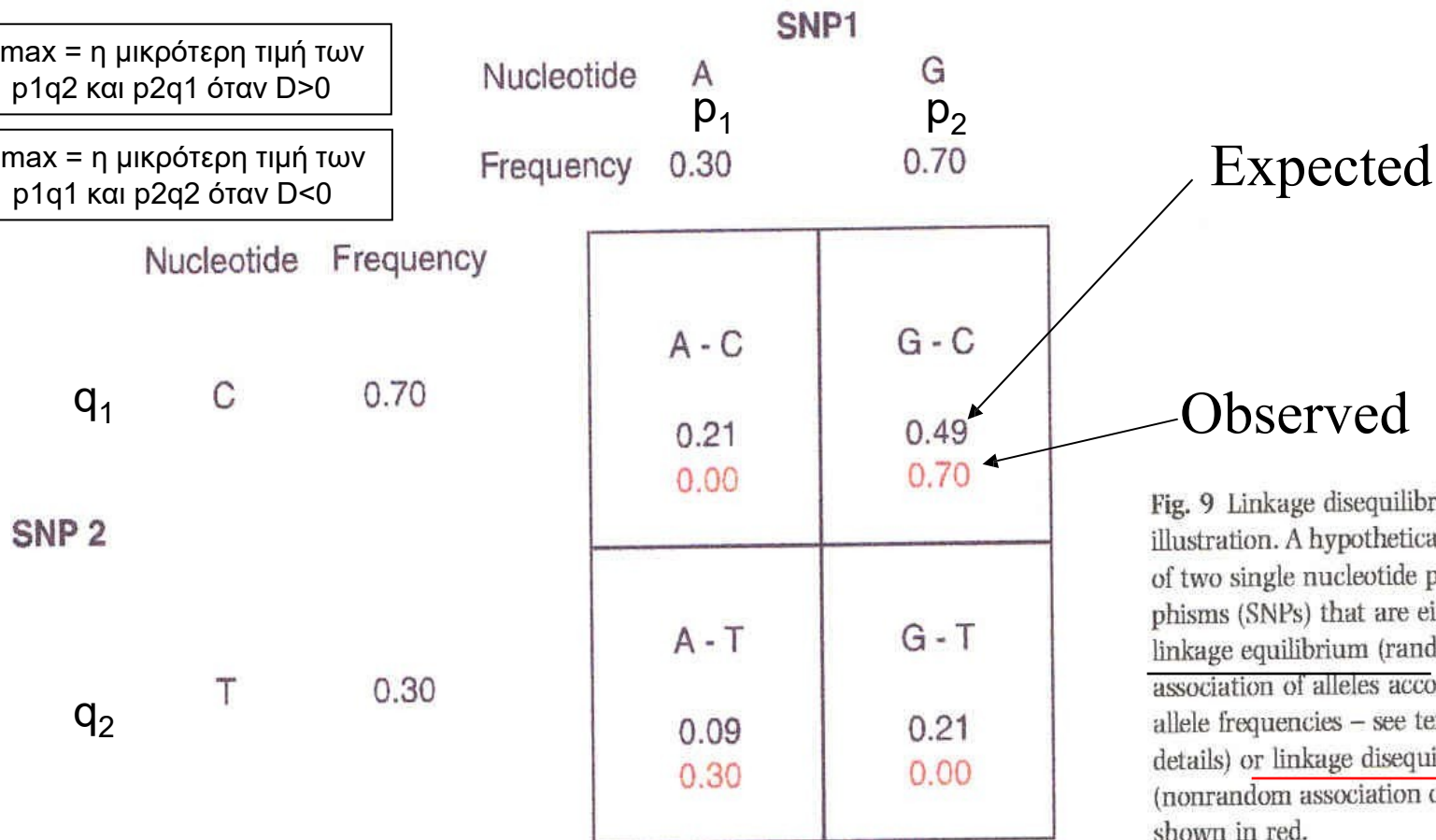


Fig. 9 Linkage disequilibrium illustration. A hypothetical example of two single nucleotide polymorphisms (SNPs) that are either in linkage equilibrium (random association of alleles according to allele frequencies – see text for details) or linkage disequilibrium (nonrandom association of alleles)– shown in red.

From Kaessman & Paabo 2002

(The statistical test is a chi-square test)

Προσδιορισμός της στατιστικής σημαντικότητας

Determining Significance

$$X^2 = \sum_{i=1}^k \frac{(O - E)^2}{E}$$

$$df = 1$$

O: παρατηρούμενη συχνότητα (Observed)

E: αναμενόμενη συχνότητα (Expected)

Αν $p < 0.05 \rightarrow$ απορρίπτεις την H_0 (LD)

Αν $p \geq 0.05 \rightarrow$ δεν έχεις επαρκή ένδειξη για LD

Παράδειγμα – Ομάδες Αίματος

N = 2000 άτομα
(γονότυποι)

	S	s
M	474	611
N	142	773

	S	s
M	x_1 0.237	x_2 0.3055
N	x_3 0.071	x_4 0.3865

$$MS = 474/2000 = 0.237$$

$$Ms = 611/2000 = 0.3055$$

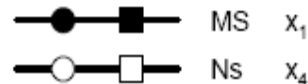
$$NS = 142/2000 = 0.071$$

$$Ns = 773/2000 = 0.3865$$

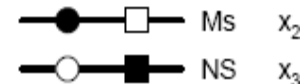
**Πατρικός
απλότυπος**



**Μη ανασυνδυασμένα
χρωμοσώματα**



**Ανασυνδυασμένα
χρωμοσώματα**



Παράδειγμα – Ομάδες Αίματος

	S	s
M	x_1 0.237	x_2 0.3055
N	x_3 0.071	x_4 0.3865

Gamete	Frequency	Allele	Frequency
MS	$x_1 = 0.2370$	M	$p_1 = x_1 + x_2 = 0.237 + 0.3055 = 0.5425$
Ms	$x_2 = 0.3055$	N	$p_2 = x_3 + x_4 = 0.071 + 0.3865 = 0.4575$
NS	$x_3 = 0.0710$	S	$q_1 = x_1 + x_3 = 0.237 + 0.071 = 0.308$
Ns	$x_4 = 0.3865$	s	$q_2 = x_2 + x_4 = 0.3055 + 0.3865 = 0.692$

Υπόθεση H_0 : Εφόσον παρατηρούμε όλους τους απλότυπους, οι τόποι βρίσκονται σε ισορροπία σύνδεσης (ασύνδετοι)

Παράδειγμα – Ομάδες Αίματος

Gamete	Frequency	Allele	Frequency
MS	$X_1 = 0.237$	M	$p_1 = 0.5425$
Ms	$X_2 = 0.3055$	N	$p_2 = 0.4575$
NS	$X_3 = 0.071$	S	$q_1 = 0.308$
Ns	$X_4 = 0.3865$	s	$q_2 = 0.692$

$$D = X_1X_4 - X_2X_3$$

$$D = (0.237)(0.3865) - (0.3055)(0.071) = 0.07$$

$$D' = \frac{D}{D_{\max}}$$

$$D' = \frac{0.07}{0.141} = 0.5$$

$$p_2q_1 = 0.141 < p_1q_2 = 0.377$$

Παράδειγμα – Ομάδες Αίματος

$$X^2 = \sum_{i=1}^k \frac{(O - E)^2}{E} \quad df = 1$$

degrees of freedom, $df = (r - 1) * (c - 1)$
 where r = number of rows
 c = number of columns

Observed

	S	s
M	474	611
N	142	773

Expected

	S	s
M	334.18	750.82
N	281.82	633.18

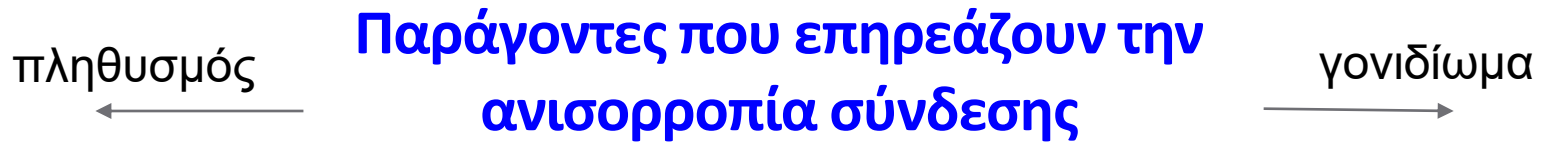
E_i is the expected frequency

$$E_i = \frac{r_{i\text{total}} * c_{i\text{total}}}{\text{grand}_{\text{total}}}$$

$$X^2 = \frac{(474 - 334.18)^2}{334.18} + \frac{(611 - 750.82)^2}{750.82} + \frac{(142 - 281.82)^2}{281.82} + \frac{(773 - 633.18)^2}{633.18} \approx 184$$

Καταλήγουμε πως η υπόθεση H_0 καταρρίπτεται: Υπάρχει σημαντικό LD μεταξύ των δύο γενετικών τόπων ($D'=0.5$) - Η παρατηρούμενη απόκλιση ανάμεσα στο πείραμα (observed) και τις θεωρητικές (expected) τιμές είναι μάλλον απίθανο (όχι όμως αδύνατο) ότι οφείλεται μόνο στην τύχη.

$$p < 0.001$$



Η ανισορροπία σύνδεσης διαφέρει ανάμεσα σε διαφορετικούς πληθυσμούς και ανάμεσα σε διαφορετικές γενετικές περιοχές

- Φυσική απόσταση ανάμεσα σε δύο γενετικούς τόπους
 - Ρυθμός ανασυνδυασμού του γονιδιώματος
 - Ρυθμός μεταλλάξεων (γονιδιακή ροή)
 - Φυσική επιλογή
-
- Μέγεθος πληθυσμιακού δείγματος
 - Τυχαία γεγονότα (γενετική παρέκκλιση)
 - Φαινόμενο του ιδρυτή (founder effect)
 - Γενετική στενωπός (bottleneck)
 - Πληθυσμιακές προσμίξεις (admixture) ή ομομιξία (απομονωμένοι πληθυσμοί)
-
- γονιδίωμα ↑
↓
πληθυσμός

Το D σχετίζεται με το ρυθμό
ανασυνδυασμού

$$D_t = (1 - \theta)^t D_0$$

Όπου θ το κλάσμα ανασυνδυασμού

Χωρίς ανασυνδυασμό

	B	b	
A	$p_A p_B + D_{AB}$	$p_A p_b - D_{AB}$	p_A
a	$p_a p_B - D_{AB}$	$p_a p_b + D_{AB}$	p_a
	p_B	p_b	

Οι συχνότητες των απλοτύπων παραμένουν σταθερές στην διάρκεια του χρόνου

Με ανασυνδυασμό

	B		b
A	$p_A p_B + (1 - \theta) D_{AB}$	$p_A p_b - (1 - \theta) D_{AB}$	p_A
a	$p_A p_b - (1 - \theta) D_{AB}$	$p_a p_b + (1 - \theta) D_{AB}$	p_a
	p_B		p_b

Η ανισορροπία σύνδεσης μειώνεται...

Θεωρητικός ρυθμός μείωσης της LD μεταξύ δύο γενετικών τόπων ως συνάρτηση του χρόνου (γενιές)

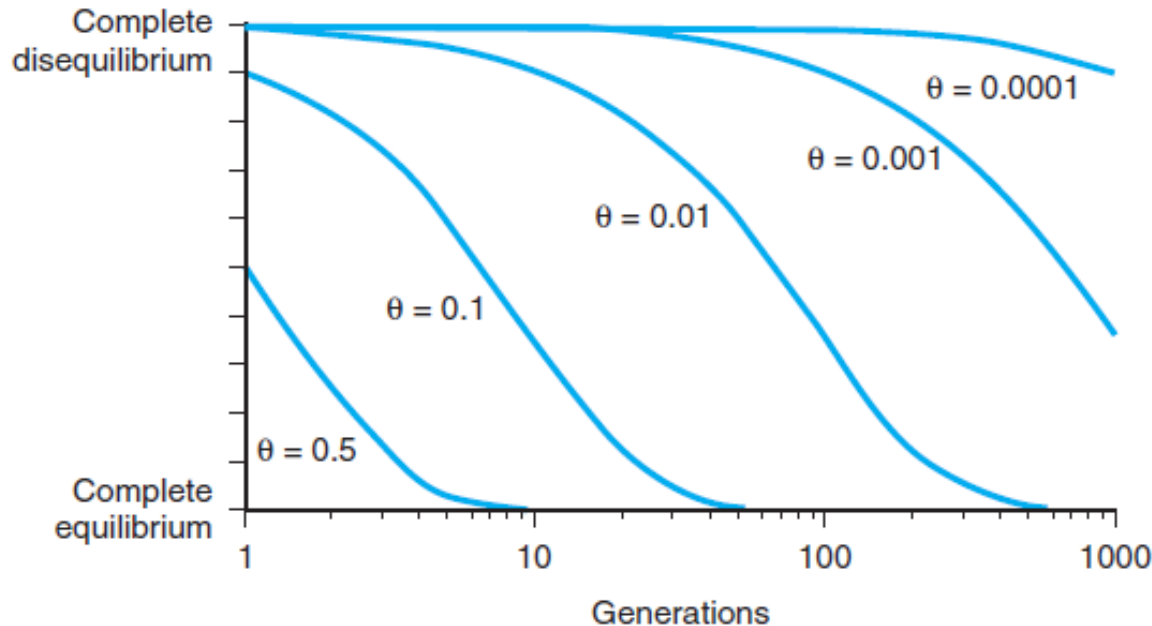


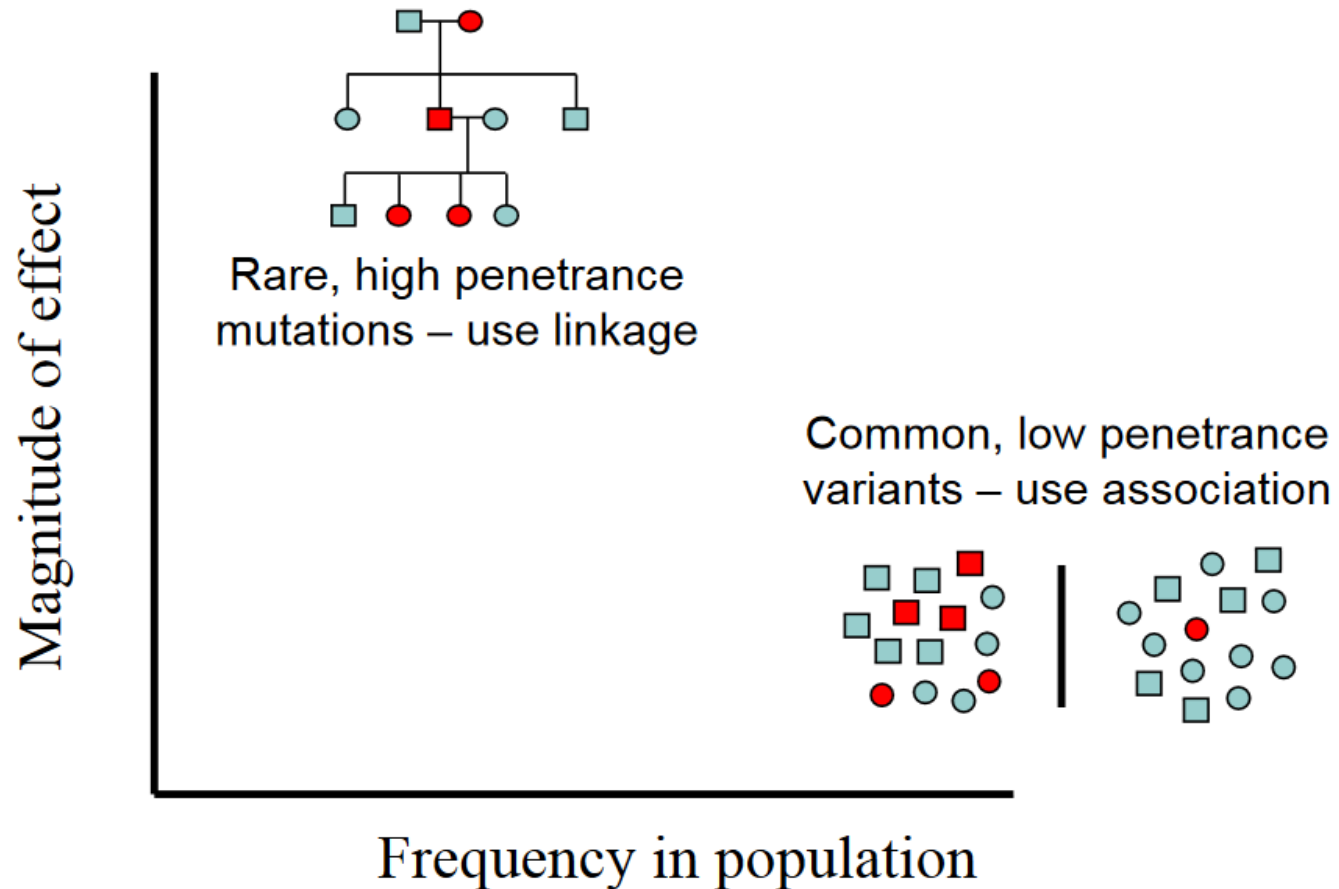
Figure 10-11 ■ Theoretical rate at which initial linkage disequilibrium between alleles at two loci decays and the alleles approach linkage equilibrium as a function of time and various values of recombination frequency, θ , between the markers. (Adapted from original figure by G. Abecasis, University of Michigan. <http://www.sph.umich.edu/csglabecasis/class/666.03.pdf>.)

Η ανισορροπία σύνδεσης είναι χρήσιμη...

- Για τη διερεύνηση της ιστορίας των πληθυσμών
- Για τη χαρτογράφηση γονιδίων (γνωρίσματα ή νοσήματα)

The Allelic Architecture of Disease

What is it and how do we discover it?



Linkage disequilibrium – understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Nature Reviews Genetics 9, 477–485(2008) | [Cite this article](#)

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima¹ and it persists because LD was initially the concern of population geneticists who were

not picky about definition was c which to study L biology and huma of population ge rapidly in the 1980 mapping became evid closely linked loci was too well establ

- Linkage disequilibrium (LD) is the nonrandom association of alleles of different loci.

There is no single best statistic that quantifies the extent of LD. Several statistics have been proposed that are useful for different purposes.

- In humans and other model organisms, LD between marker alleles and traits of interest allow fine-scale gene mapping. Many recent genome-wide association studies have successfully mapped SNPs associated with complex inherited diseases in humans.