

# Mapowanie genów

10 maja 2004

## Co daje identyfikacja genów?

- poszerzenie wiedzy podstawowej
- perspektywy przyspieszenia i zwiększenia efektywności selekcji
- precyzyjne wprowadzanie obcych genów do populacji (introgresja)

## Klonowanie pozycyjne genów kandydujących

Najbardziej preferowana metoda identyfikacji genów

Etapy:

1. **określenie pozycji genu na chromosomie względem markerów**
2. określenie genów kandydatów we wskazanym rejonie (mapy transkrypcyjne)
3. zidentyfikowanie zmienności kształtującej cechę

## Markery genetyczne

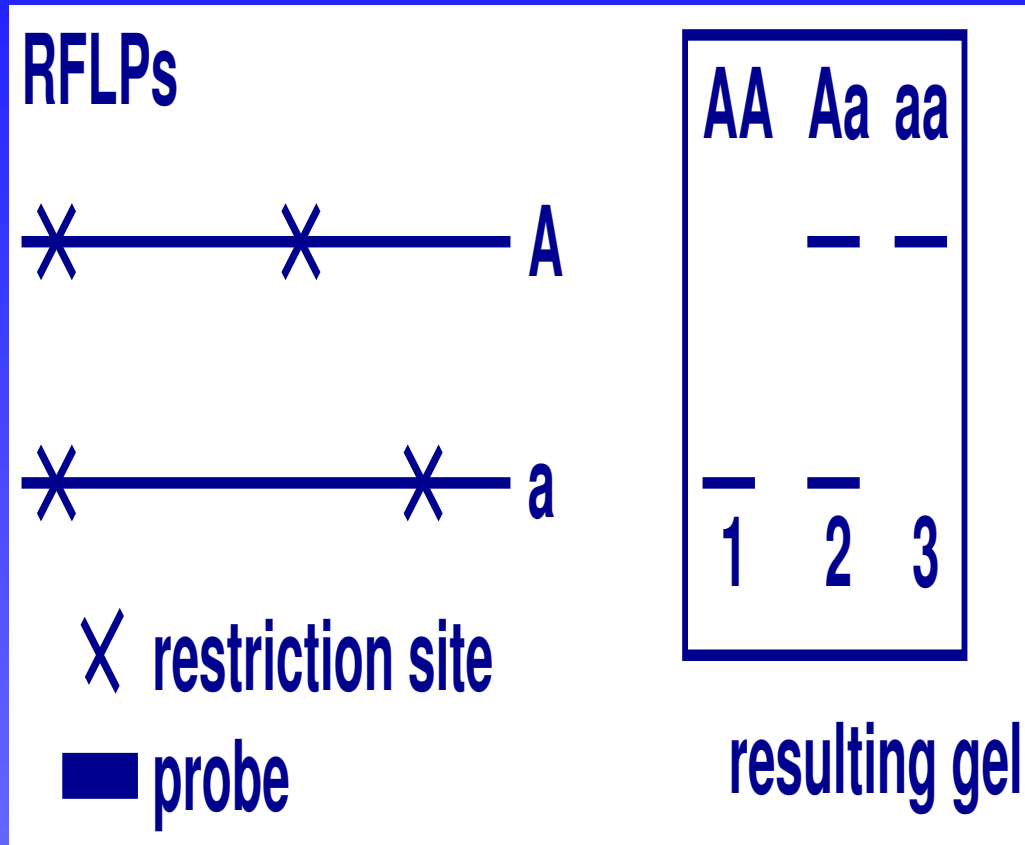
Trzy własności markera:

1. specyficzny dla danego locus
2. polimorficzny w danej populacji
3. łatwy do oznaczenia

## Typy markerów genetycznych

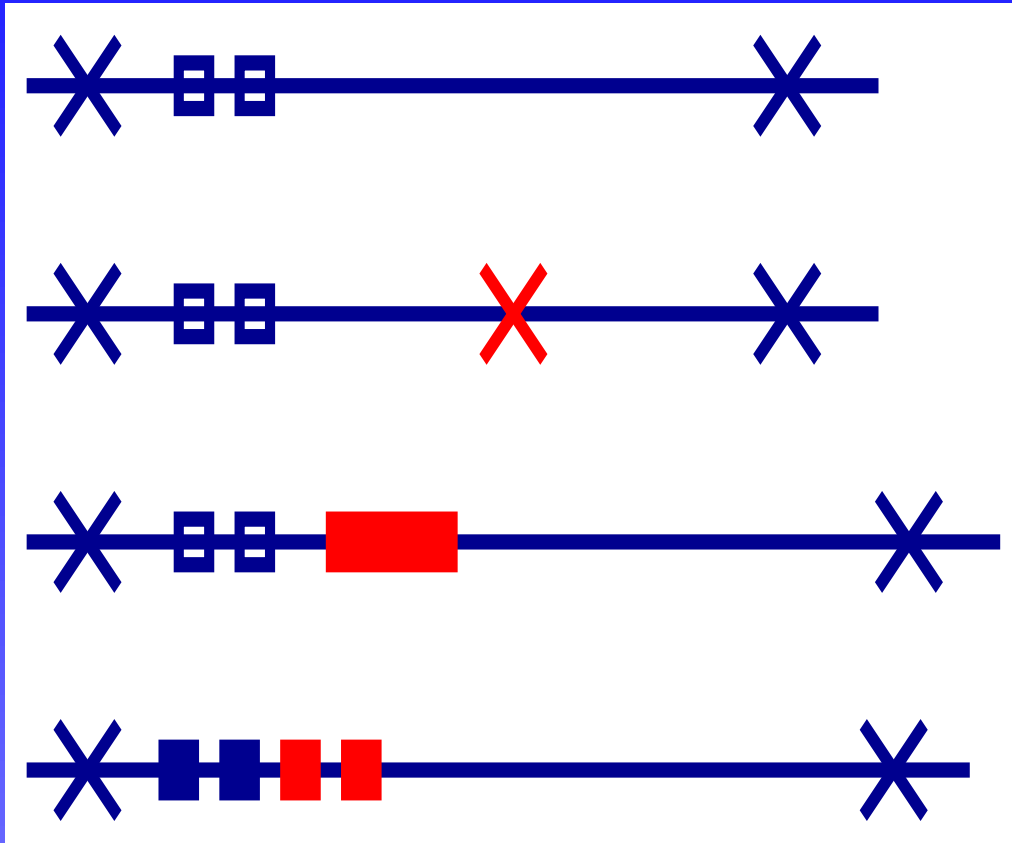
- markery fenotypowe
- grupy krwi
- polimorfizm biochemiczny
- RFLPs
- Minisatelite (VNTR)
- Mikrosatelite (SSR)
- SNPs

## RFLPs



- kodominujące
- potrzebna właściwa kombinacja enzym-sonda
- tylko 200 enzymów, liczba sond nieograniczona

## RFLPs

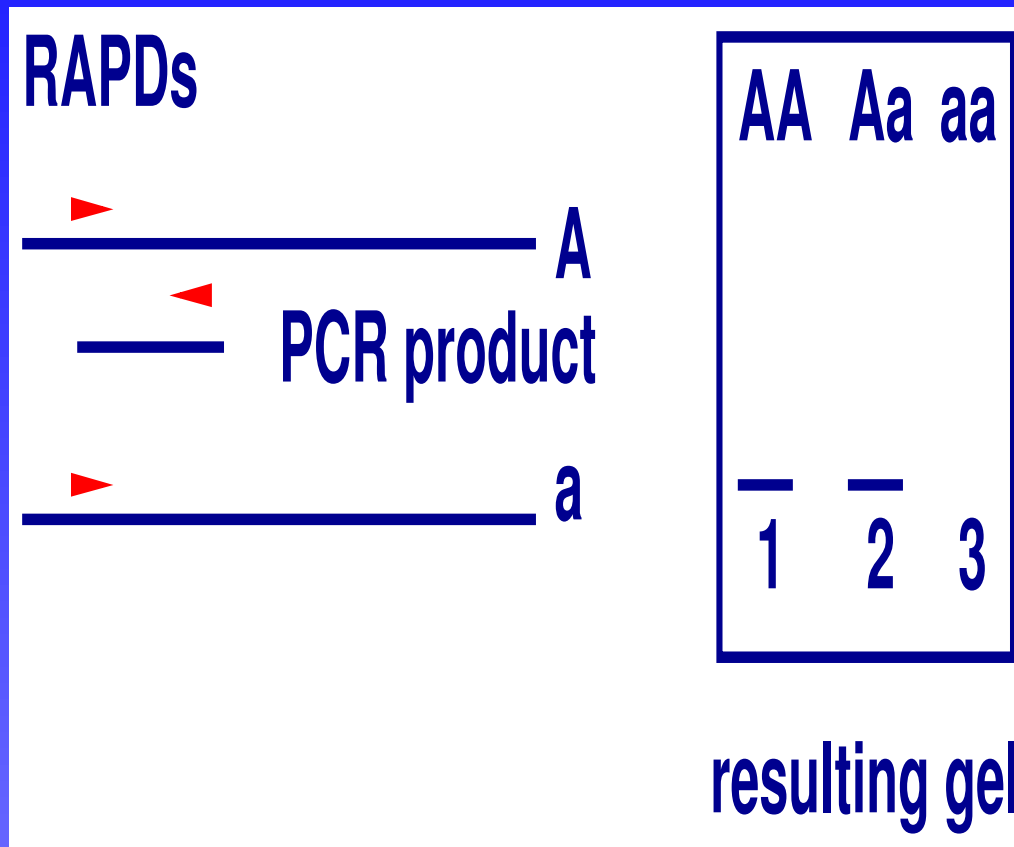


● mutacja punktowa

● insercja 35

● zmienna liczba powtórzeń

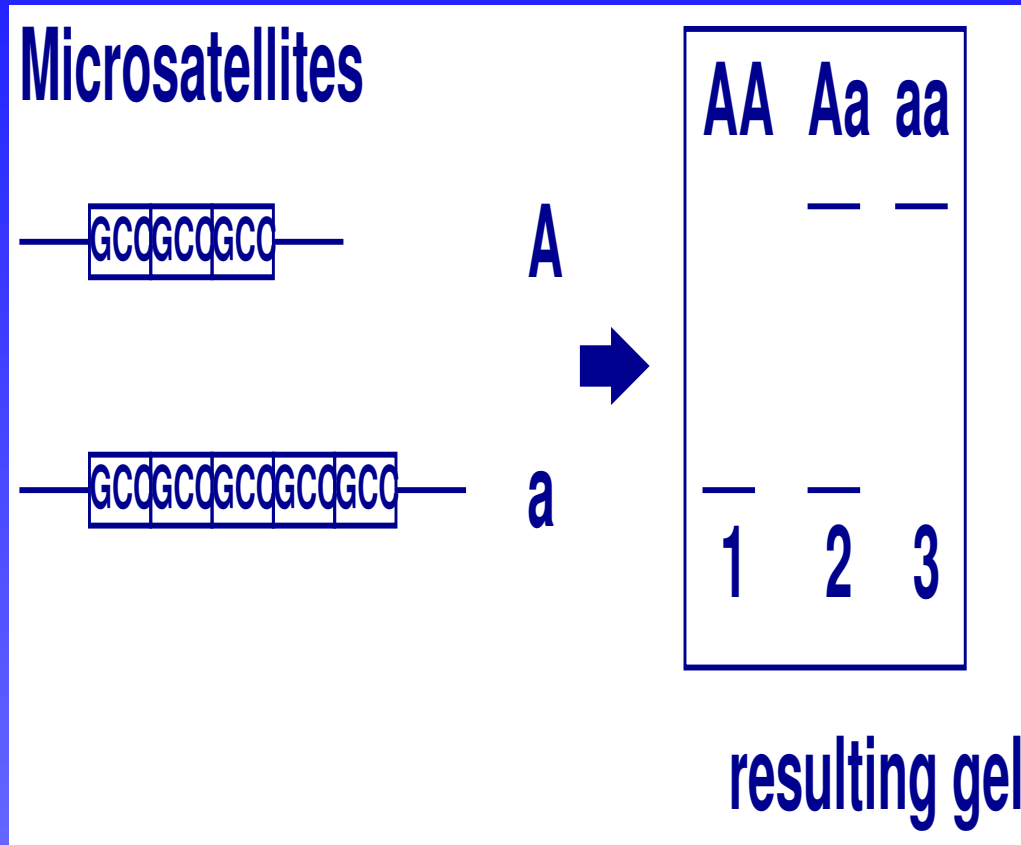
## RAPDs



- dominujące
- wiele loci jednocześnie



## Mikrosatelity



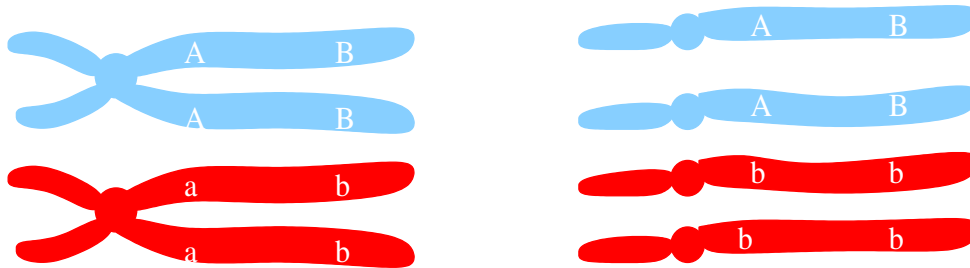
- kodominujące
- duża obfitość
- równomiernie rozłożone w genomie
- półautomatyczne wykrywanie

## Strategie mapowania

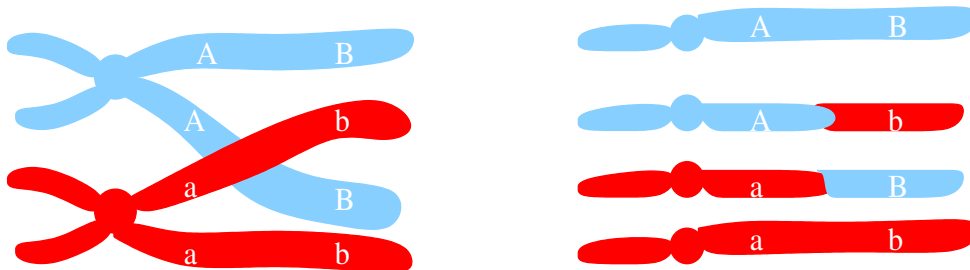
- Linkage Analysis (analiza sprzężeniowa)
- DNA Pooling / Bulk Segregant Analysis
- Genetically Directed Representational Difference Analysis
- analiza asocjacji i mapowanie genów i.p.p
- Genome Mismatch Scanning

## Tempo rekombinaciji

49 mejoz bez crossing-over



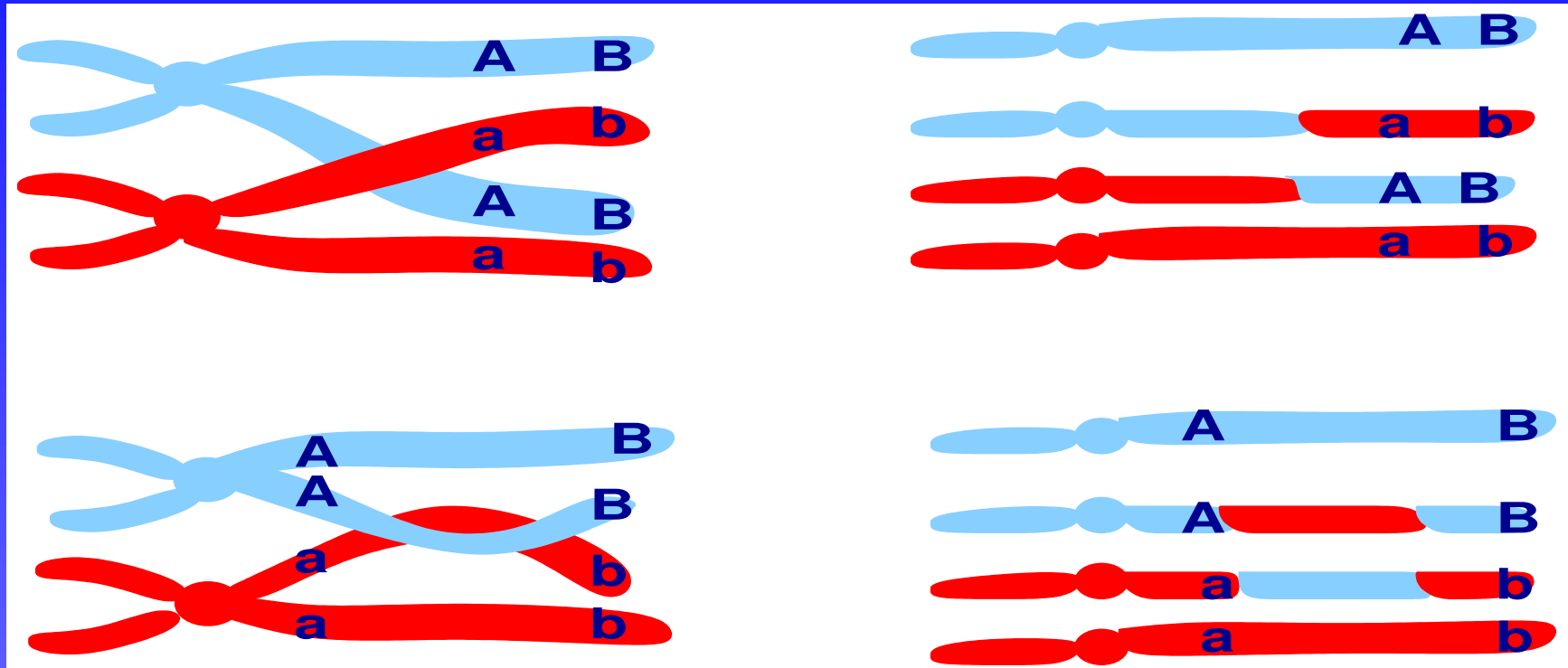
1 mejoza z crossing-over



tempo rekombinaciji

$$\frac{1+1}{49+49+49+49+1+1+1+1} = 0.01$$

## Ukryte crossing-over



## Odległość genetyczna

- mierzona częstością (prawdopodobieństwem) crossing-over
- jednostką jest Morgan
- 1 centiMorgan - prawdopodobieństwo crossing-over wynosi 0.01
- addytywna ( $r_{AC} = r_{AB} + r_{BC}$ )
- określana na podstawie rekombinacji

## Funkcje mapowe

Pozwalają 'przeliczyć' obserwowane tempo rekombinacji na odległość genetyczną

Najczęściej używane:

- Morgana
- Haldane'a
- Kosambięgo

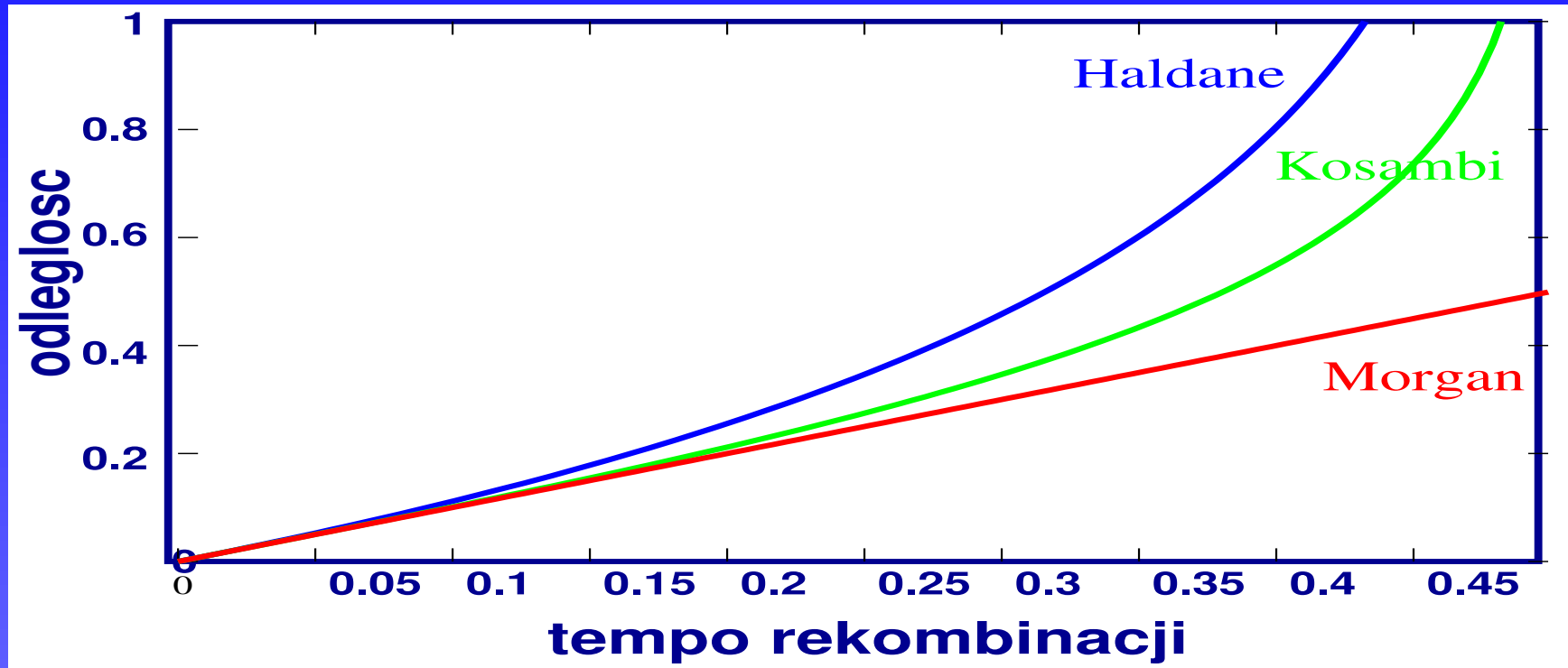
## Funkcja Kosambiego

$$x = 0.25 \times \ln \frac{1 + 2\theta}{1 - 2\theta}$$

$$\theta = 0.5 \times \frac{\exp(4x) - 1}{\exp(4x) + 1}$$

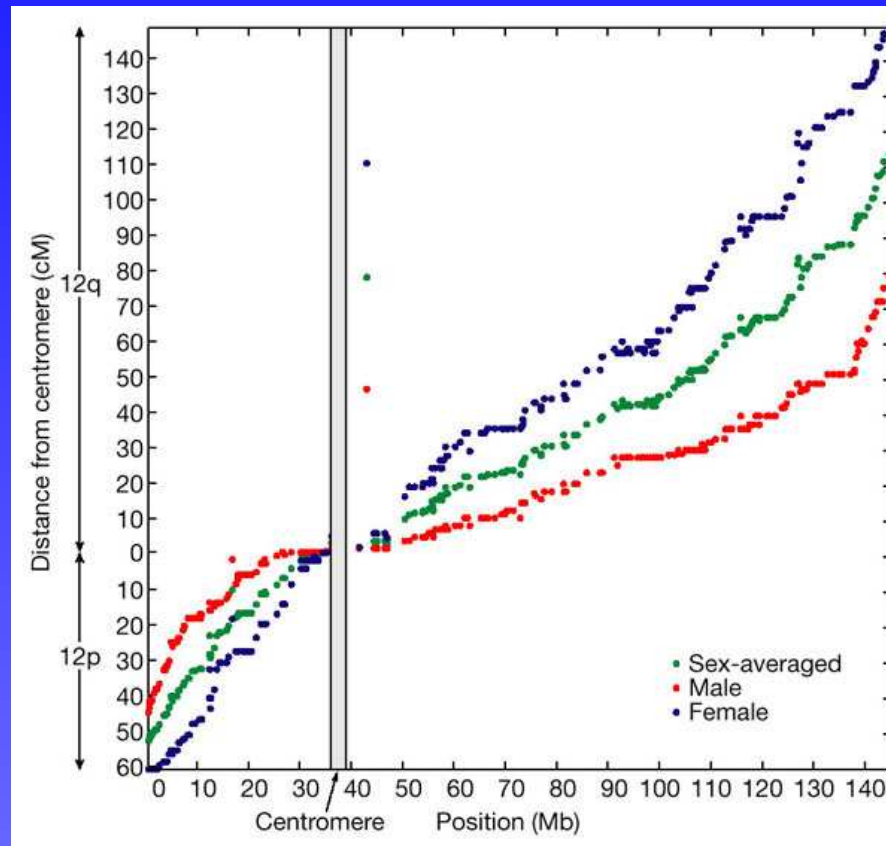
- $x$  odległość genetyczna (w Morganach)
- $\theta$  tempo rekombinacji

## Funkcje mapowe





## Odległość genetyczna i fizyczna

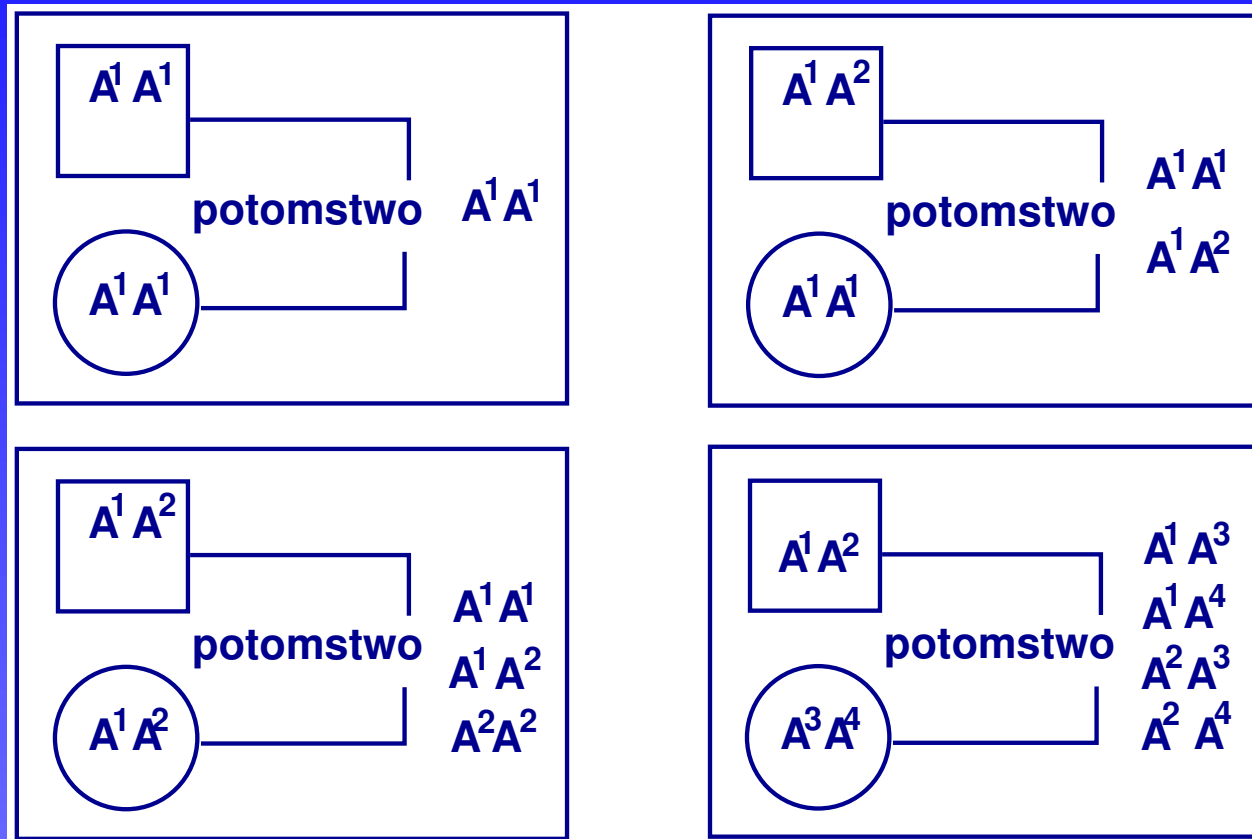


1 cM ~ 1 M p.z.

## Skąd uzyskać tempo rekombinacji?

- bezpośrednio genotypowanie plemników i oocytów (tylko markery genetyczne)
- genotypowanie rodziców i potomstwa
  - ★ jedyna możliwość dla cech fenotypowych
  - ★ wymaga podwójnych heterozygot

## Informatywność markera



## Miary informatywności markera

- Heterozygotyczność

$$H = 1 - \sum p_i^2$$

- Polymorphism Information Content (PIC)

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n p_i^2 p_j^2$$

## Test lod score

- Test statystyczny weryfikujący hipotezę o niezależnej segregacji genów z różnych loci
- Hipotezy:
  - ★ H0: brak sprzężenia ( $\theta = 0.5$ )
  - ★ H1: sprzężenie loci ( $\theta < 0.5$ )
- Wiarogodność (Likelihood)
  - ★ miara siły dowodów na rzecz hipotezy
  - ★ prawdopodobieństwo danych obliczone przy danej hipotezie

- Iloraz wiarygodności (likelihood ratio)

$$LR = \frac{L(dane | \theta)}{L(dane | \theta = 0.5)}$$

- Lod score

$$Z = \log_{10}(LR)$$

- Interpretacja
  - ★  $Z > 3$  geny są sprzężone
  - ★  $Z < -2$  brak sprzężenia
  - ★  $-2 < Z < 3$  brak informacji

## Test lod score

**A1 A2 B1 B2** × **A1 A1 B1 B1**

potomstwo	gamety		N
	ojcowska	matczyna	
<b>A1 A1 B1 B1</b>	<b>A1 B1</b>	<b>A1 B1</b>	<b>9</b>
<b>A1 A2 B1 B1</b>	<b>A2 B1</b>	<b>A1 B1</b>	<b>1</b>
<b>A1 A1 B1 B2</b>	<b>A1 B2</b>	<b>A1 B1</b>	<b>1</b>
<b>A1 A2 B1 B2</b>	<b>A2 B2</b>	<b>A1 B1</b>	<b>9</b>

## Test lod score

- najprawdopodobniej u ojca występuje sprzężenie A1-B1 oraz A2-B2
- oszacowane tempo rekombinacji  $\hat{\theta} = 2/20 = 0.1$
- wiarygodność przy  $\theta = 0.1$

$$L(0.1) = \binom{20}{2} (1 - 0.1)^{18} * 0.1^2$$

- wiarygodność przy braku sprzężenia ( $\theta = 0.5$ )

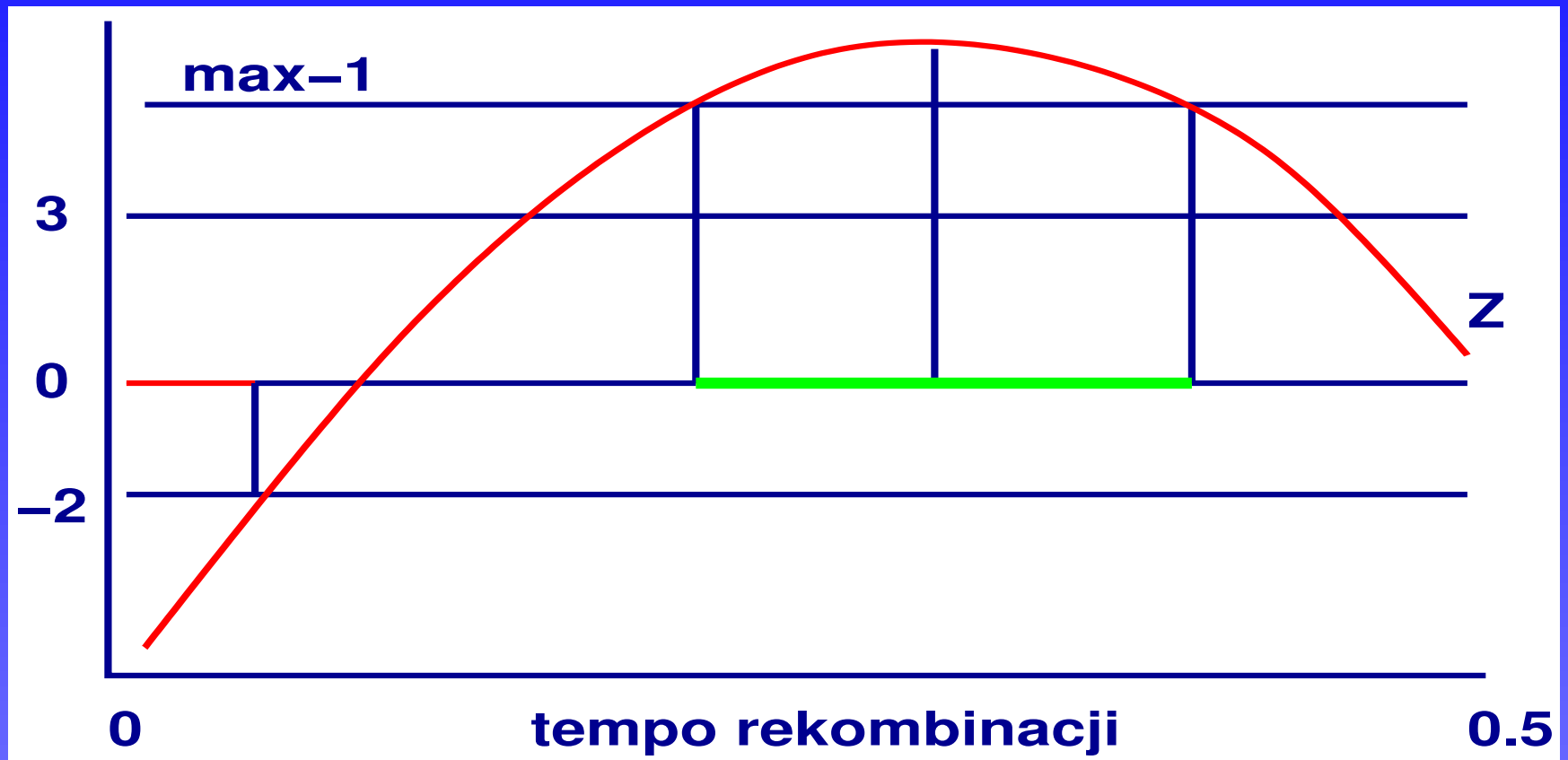


$$L(0.5) = \binom{20}{2} (1 - 0.5)^{18} * 0.5^2$$

- lod score

$$Z(0.1) = \log_{10} \frac{L(0.1)}{L(0.5)} \simeq 3.2$$

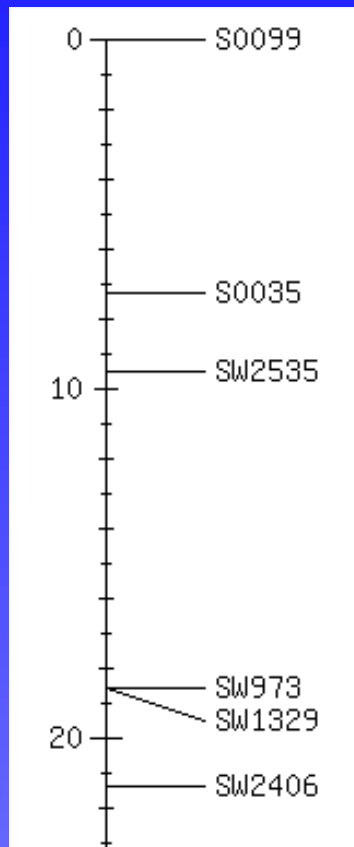
## Test lod score



## Programy komputerowe

- LINKAGE
- FASTLINK
- VITESSE
- CRIMAP - pozwala oszacować uszeregowanie genów!

## Mapy sprzężeniowe

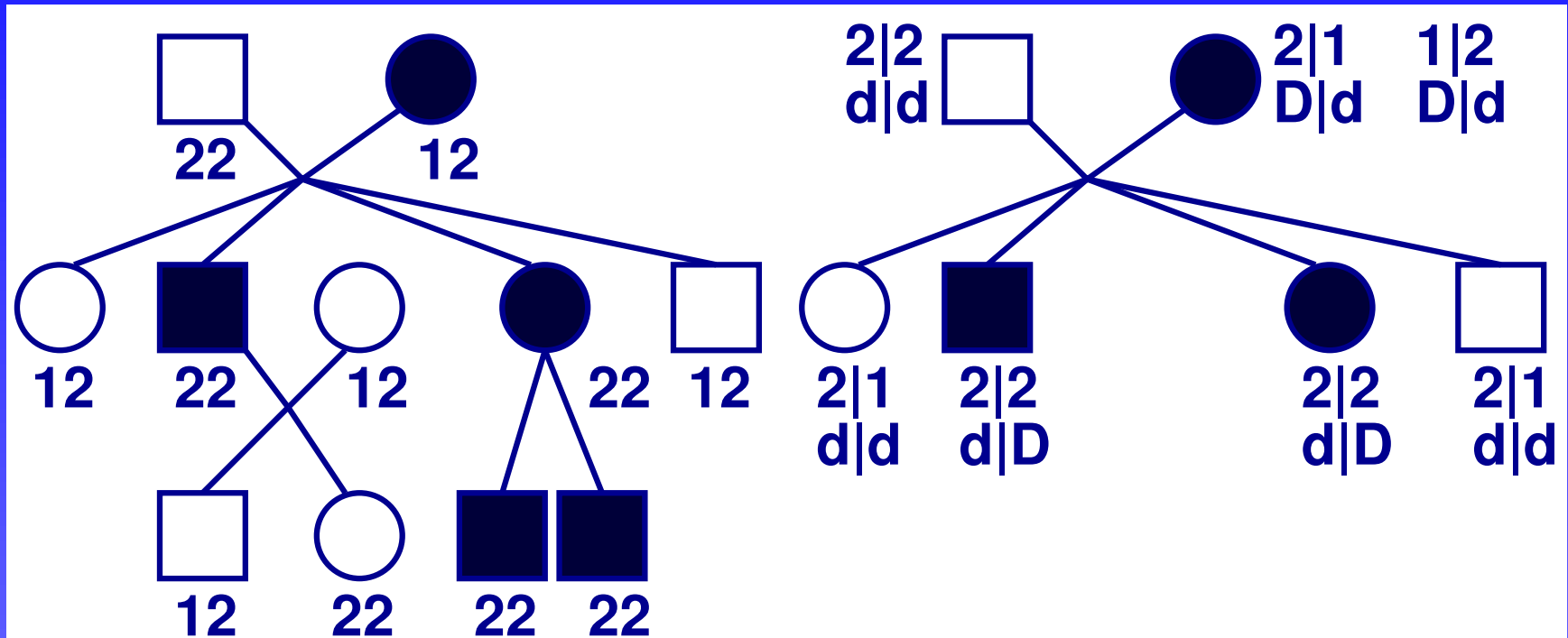


Fragment mapy sprzężeniowej chromosomu 6 świni.

## Mapy sprzężeniowe (1999)

Gatunek	liczba markerów	przeciętna odległość między markerami
Bydło	1240	2,5
Świnia	1042	2,2
Owca	519	6,4
Koń	162	14,2
Pies	276	10,0

## Mapowanie choroby monogenicznej



## Mapowanie choroby monogenicznej

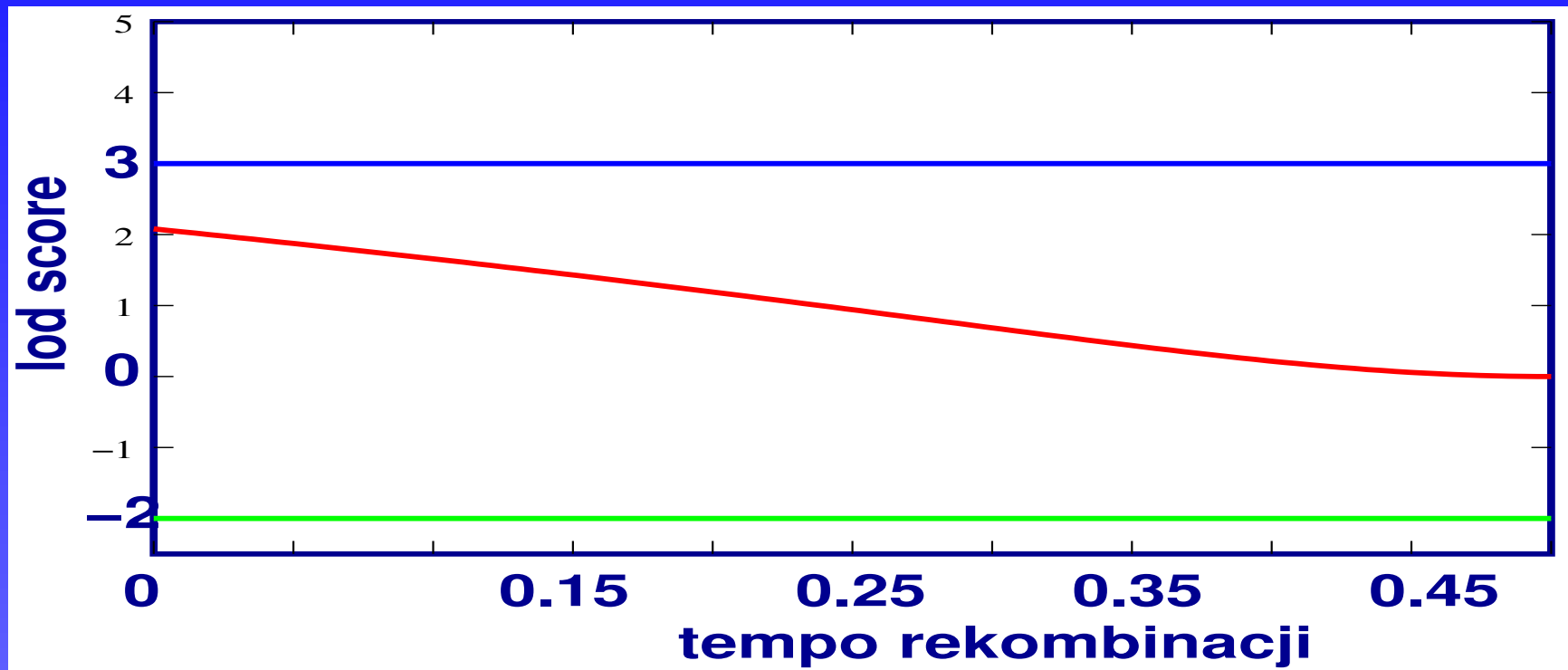
$$L^I(\theta) = (1 - \theta)^4$$

$$L^{II}(\theta) = \theta^4$$

$$L(\theta) = \frac{1}{2}[(1 - \theta)^4 + \theta^4]$$

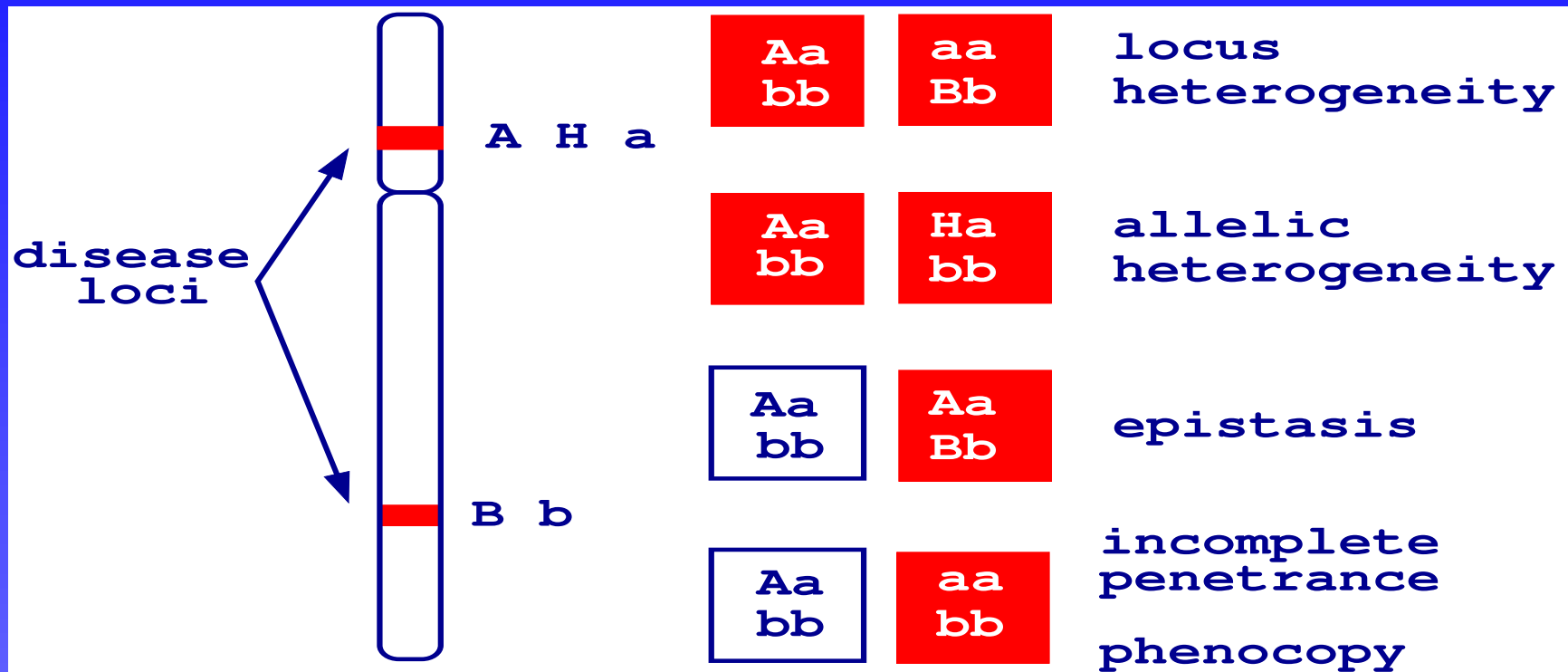
$$Z(0.25) = \log_{10} \frac{0.5 \left[ (1 - 0.25)^4 + 0.25^4 \right]}{0.5 \left[ (1 - 0.5)^4 + 0.5^4 \right]} \simeq 0.79$$

## Mapowanie choroby monogenicznej





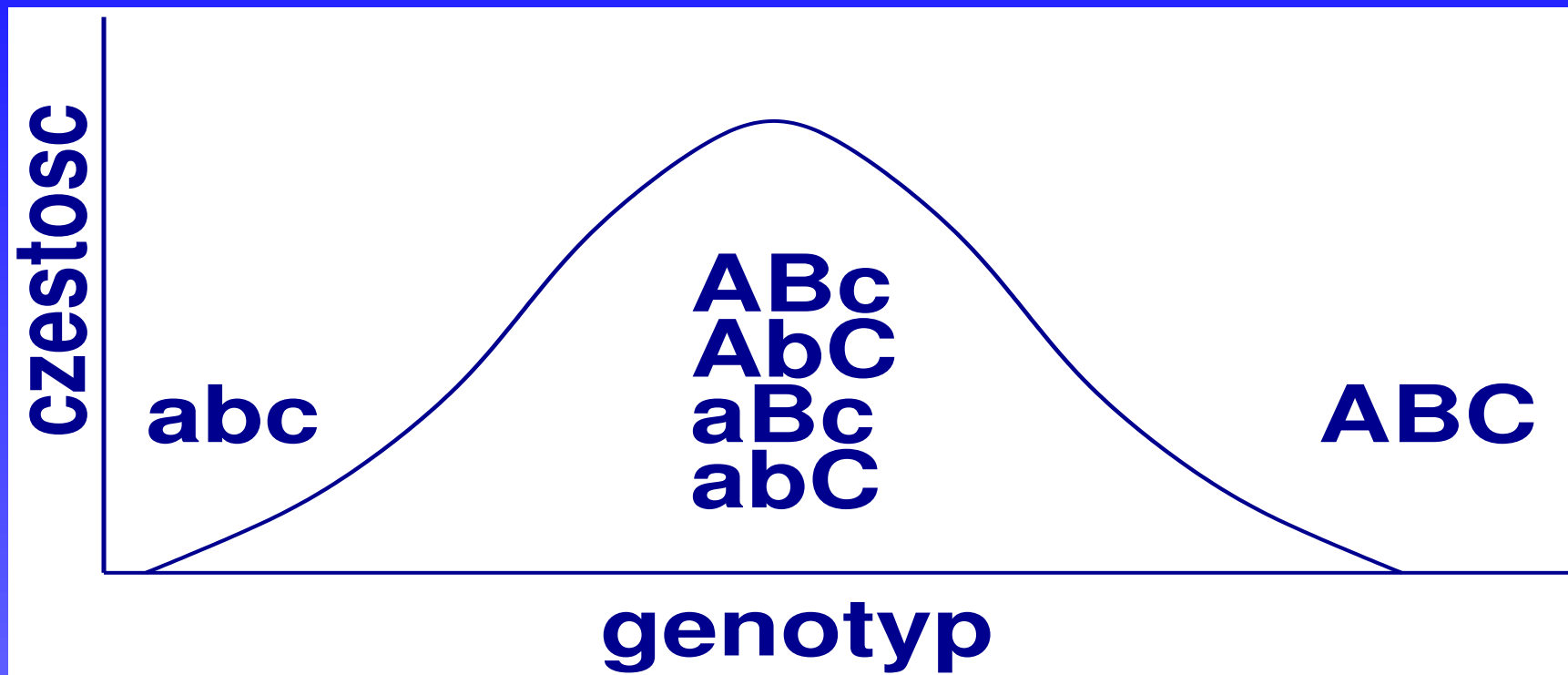
## Cechy złożone



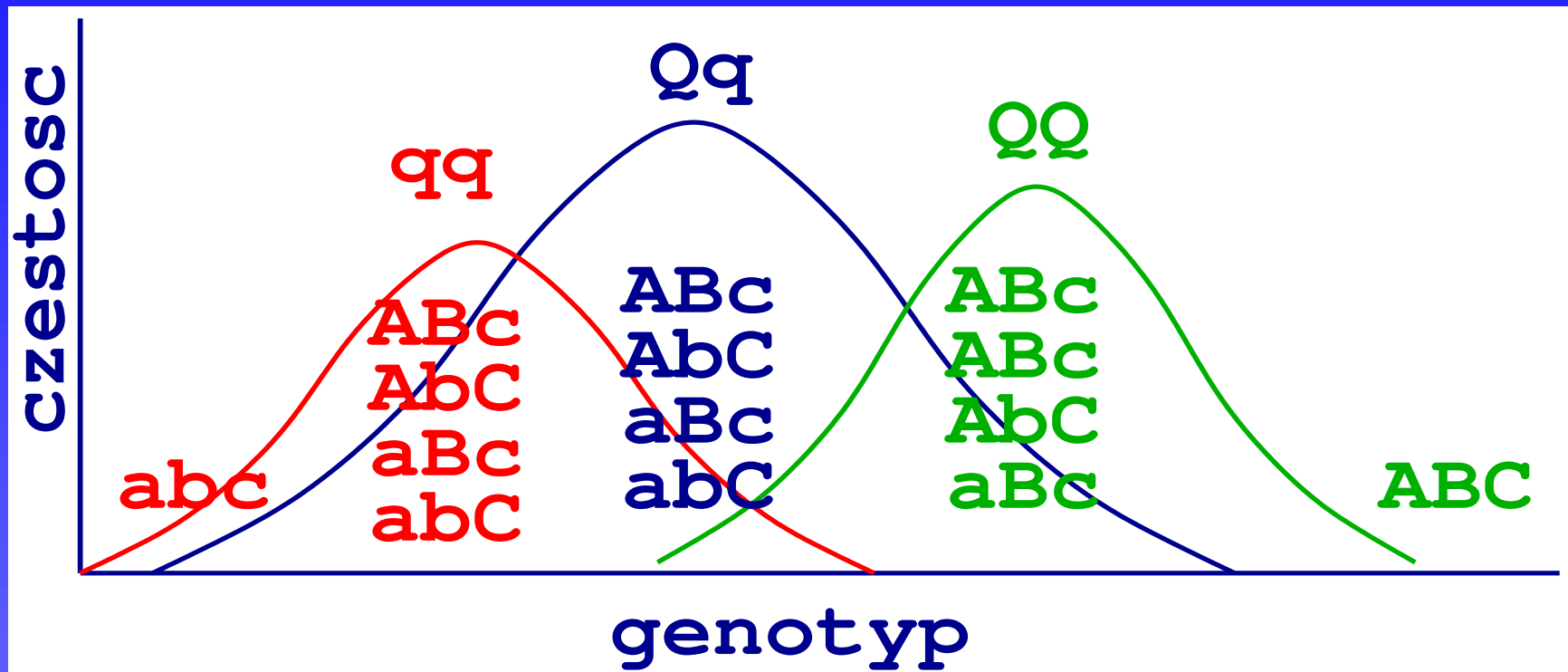
## Geny cech ilościowych

- **QTL (Quantitative Trait Locus)** to locus genu cechy ilościowej.
- Gen główny (major gene) - to QTL o bardzo dużym efekcie.

## Model poligeniczny



## Model mieszany - QTL + poligeny



# Mapowanie QTL

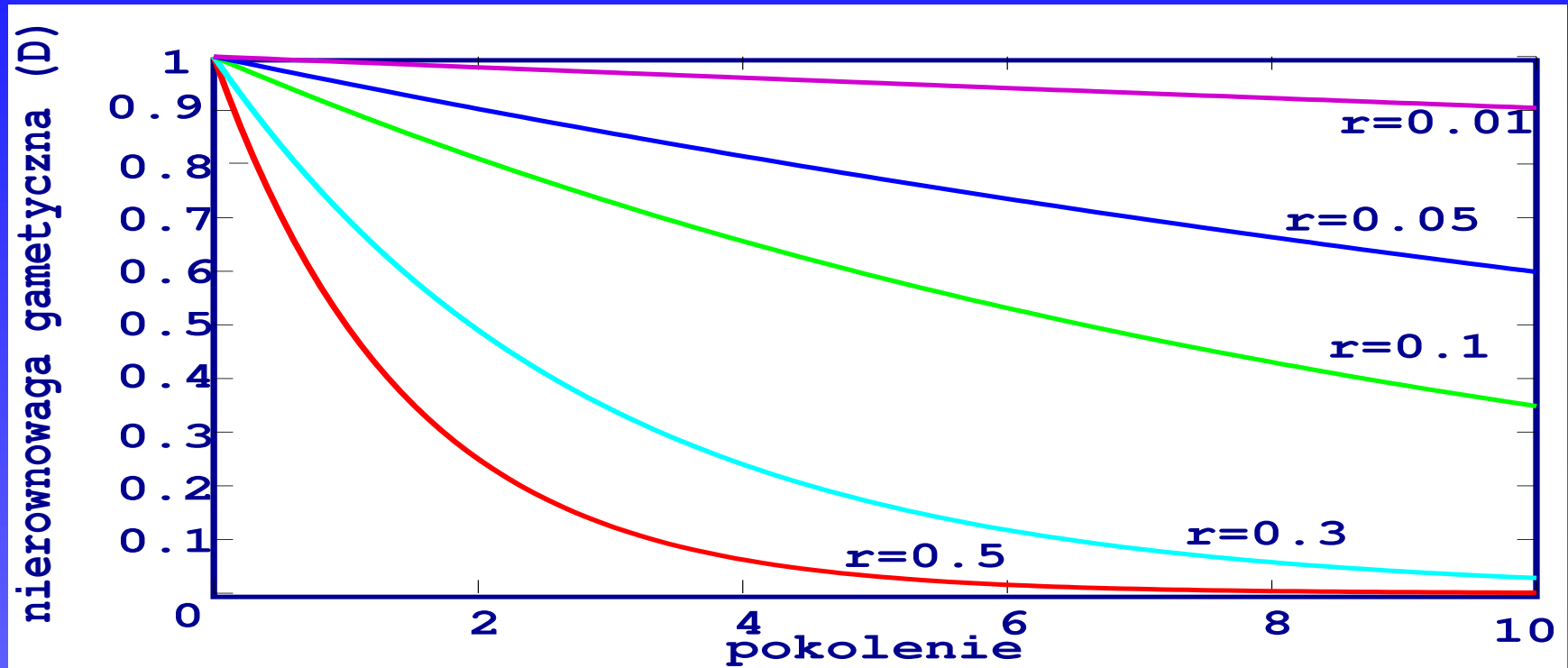
## Główna strategia

1. (utworzenie populacji w nierównowadze gametycznej)
2. śledzenie (przy pomocy markerów) przekazywania fragmentów chromosomowych od rodziców do potomstwa
3. porównanie wydajności zwierząt, które odziedziczyły różne fragmenty chromosomowe

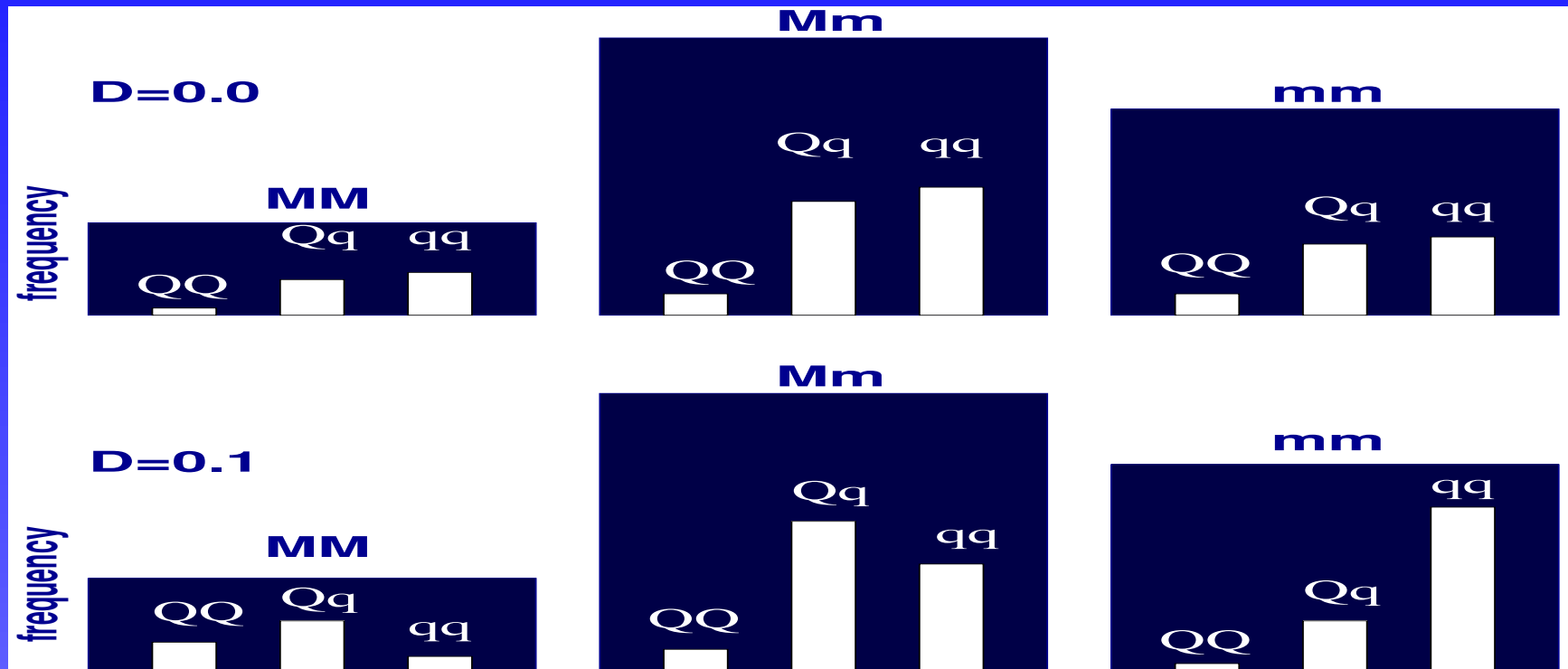
## Nierównowaga gametyczna

		A	a
		$p_A$	$p_a$
B	$q_B$	$p_A q_B + D$	$p_a q_B - D$
b	$q_b$	$p_A q_b - D$	$p_a q_b + D$

## Nierównowaga gametyczna



## Nierównowaga gametyczna





# Mapowanie QTL

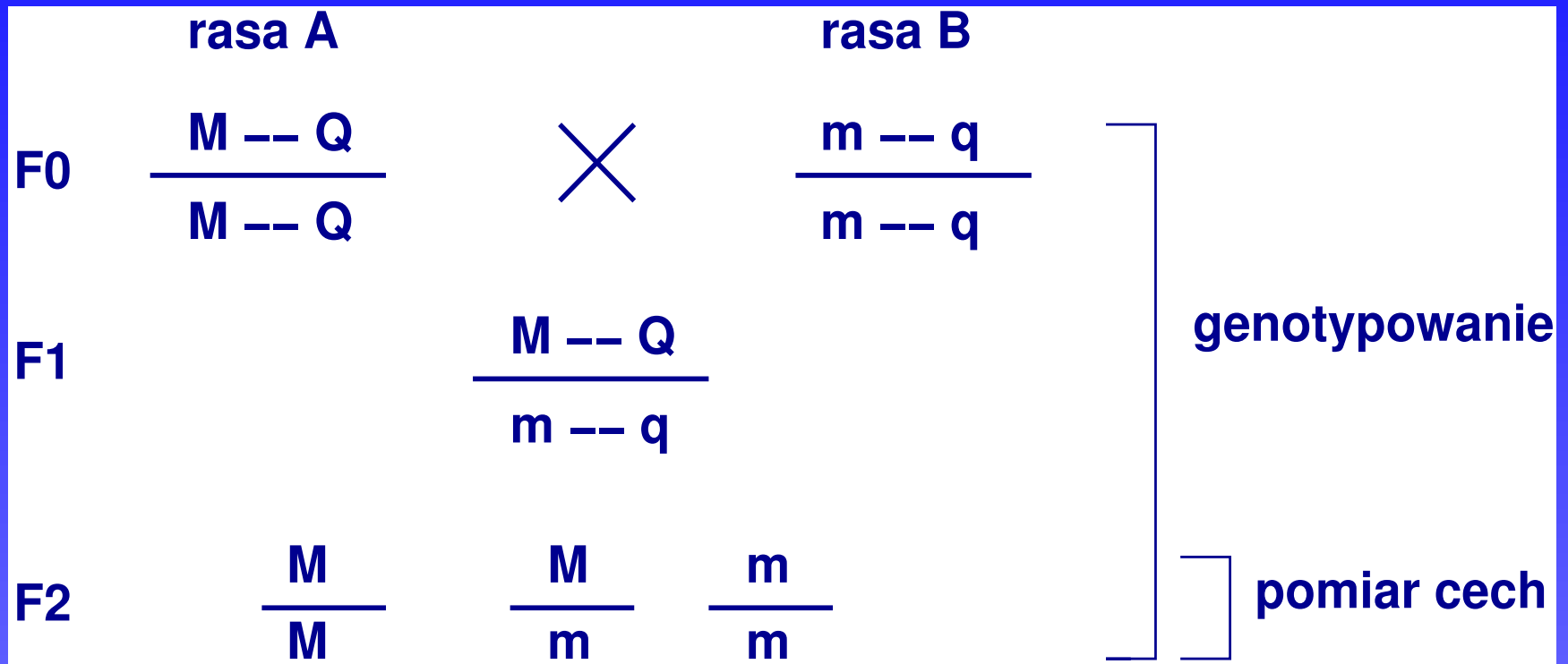
Układy doświadczalne:

- krzyżówki między liniami, rasami i gatunkami
  - ★ nierównowaga gametyczna w populacji
  - ★ duża szansa wykrycia QTL
  - ★ ale tylko tych odpowiedzialnych za różnice między rasami
- zwierzęta czystorasowe
  - ★ nierównowaga gametyczna w rodzinach
  - ★ mniejsza szansa wykrycia QTL
  - ★ QTL odpowiedzialne za zmienność osobniczą

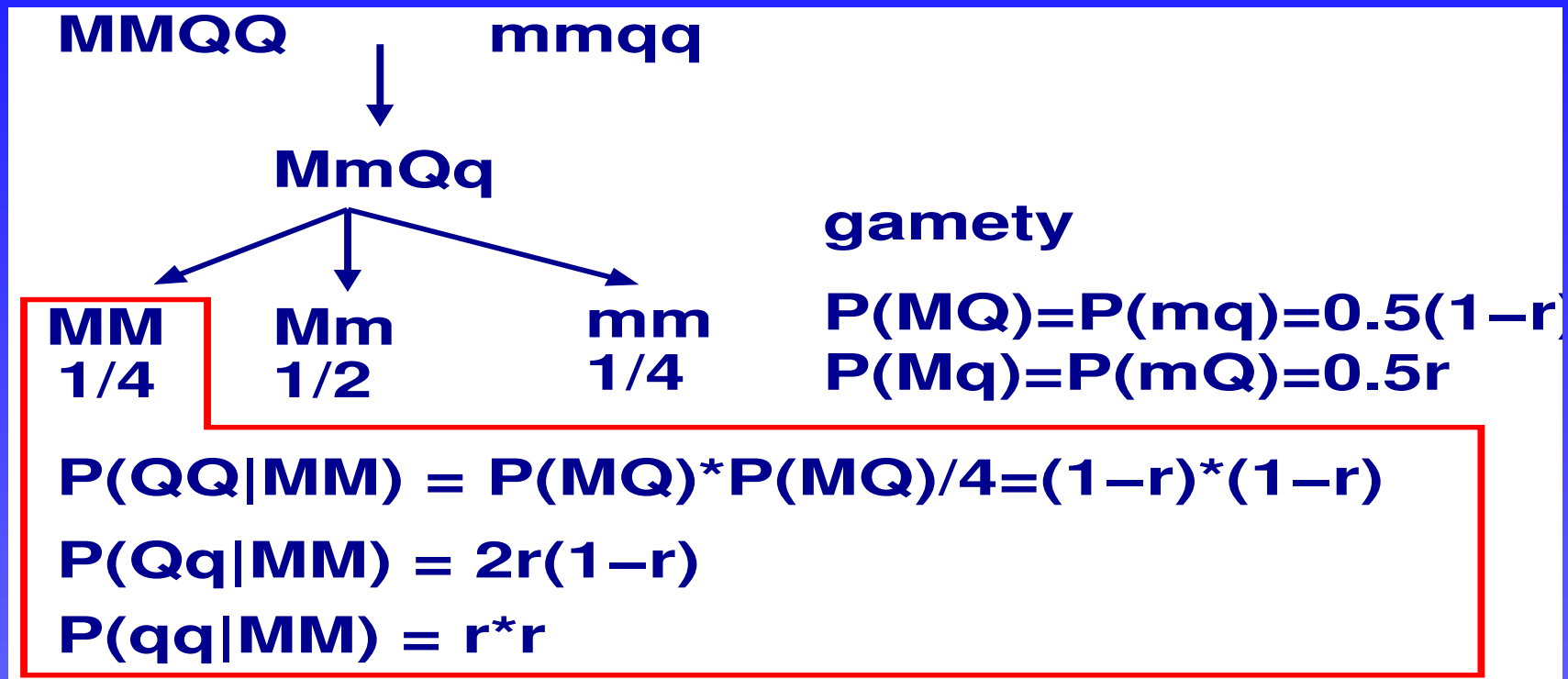
## Układ doświadczalny F2

- świnie
  - ★ dzik × świnie ras szlachetnych
  - ★ Meishan × świnie ras szlachetnych
  - ★ złotnicka pstra × wielka biała polska
- drób
  - ★ Red Jungle Fowl × White Leghorn
  - ★ Sasumadorri × White Plymouth Rock
- bydło
  - ★ bydło holsztyńsko-fryzyjskie × charolais

## Układ doświadczalny F2



## Układ doświadczalny F2

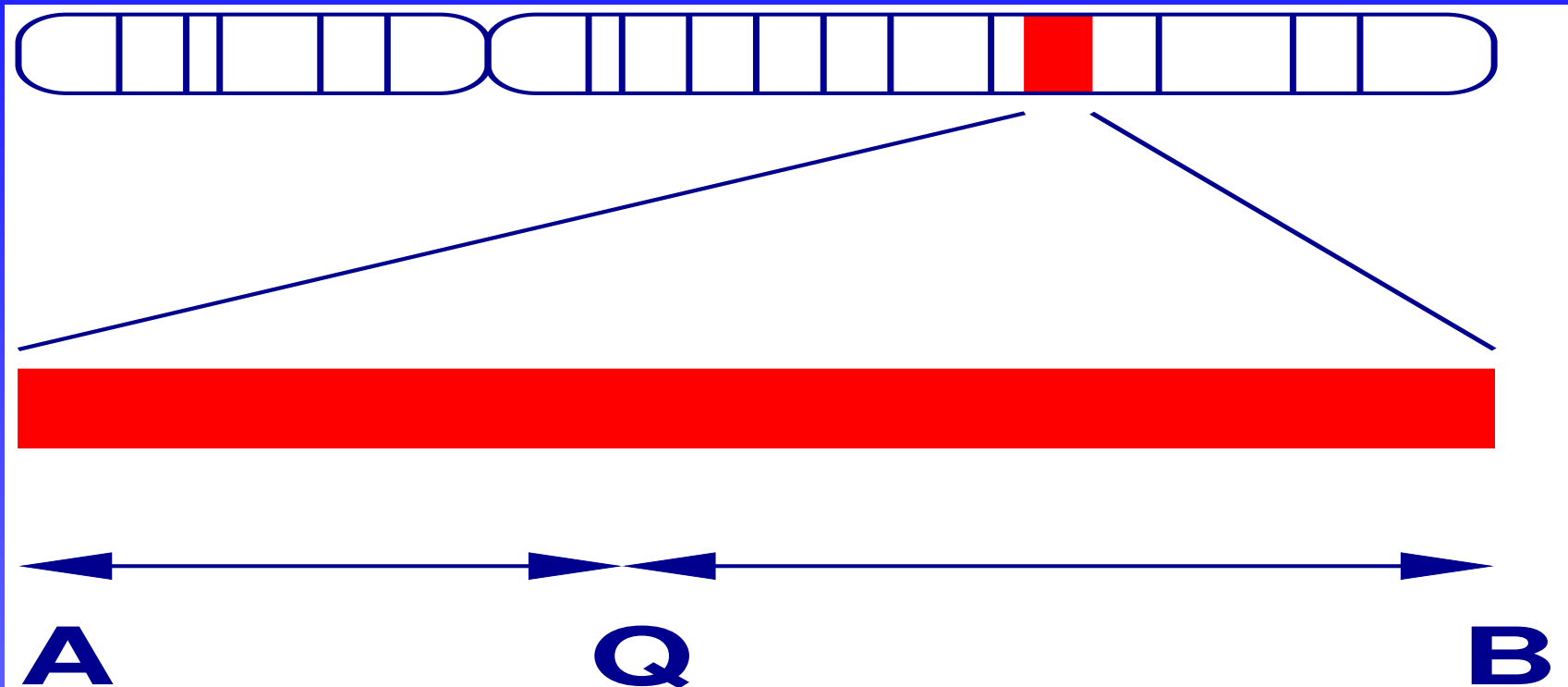


## Układ doświadczalny F2

$$\begin{aligned} L(z|MM) = & P(QQ|MM) \times f(z, \mu_{QQ}, \sigma) \\ & + P(Qq|MM) \times f(z, \mu_{Qq}, \sigma) \\ & + P(qq|MM) \times f(z, \mu_{qq}, \sigma) \end{aligned}$$

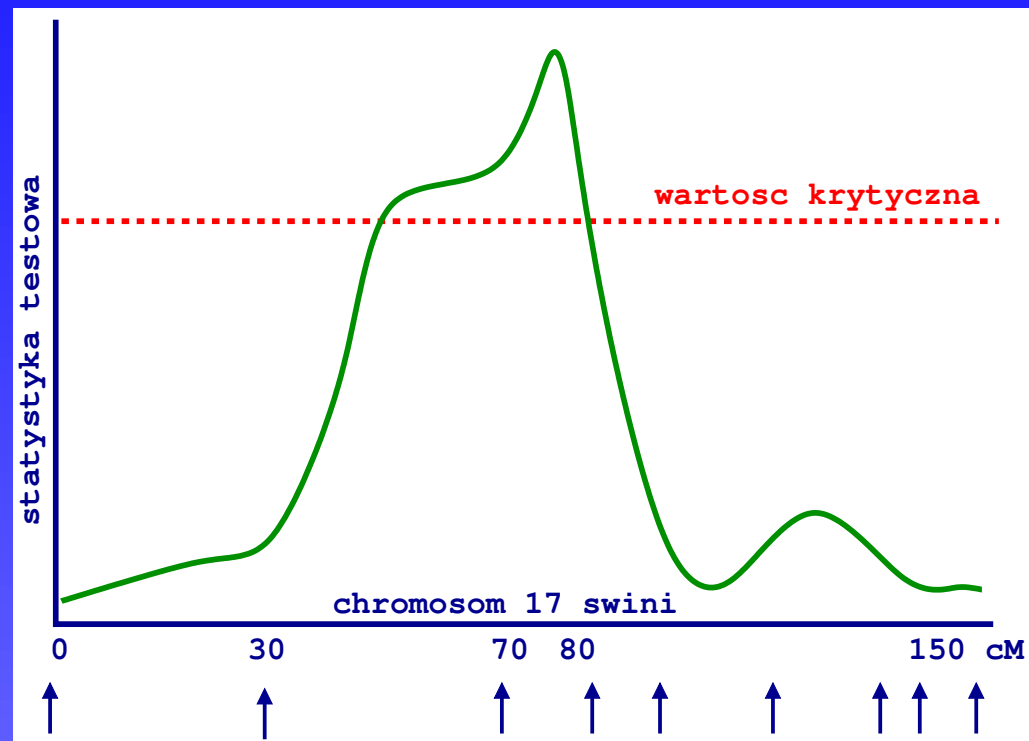
$$f(z, \mu_{QQ}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-\mu_{QQ})^2}{2\sigma^2}\right)$$

## Mapowanie przedziałowe



## Mapowanie QTL

W populacji F2, po krzyżowaniu knurów Meishan z lochami holenderskimi, badano grubość słoniny. Zlokalizowano QTL na chromosomie 17, którego efekt addytywny wynosi  $a = 2.1\text{mm}$  (De Koning 2001). Strzałki wskazują pozycje markerów.



## Mapowanie w populacjach hodowlanych

Trudniejsze niż u mieszańców:

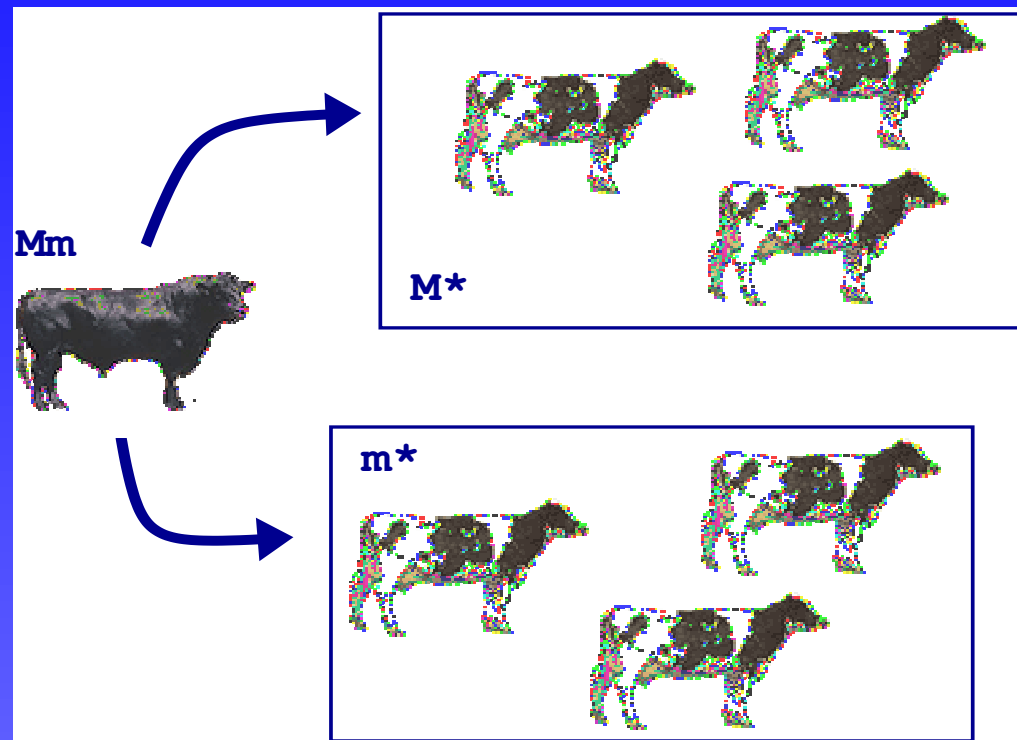
- selekcja prowadzi do utrwalenia korzystnych alleli i redukcji zmienności genetycznej
- także mniejsza zmienność markerów
- w poszczególnych rodzinach
  - ★ segregują różne QTL i allele
  - ★ różne fazy sprzężenia
- pomiar fenotypów mniej doskonały

Łatwiejsze niż u człowieka - duże grupy (pół)rodzeństwa



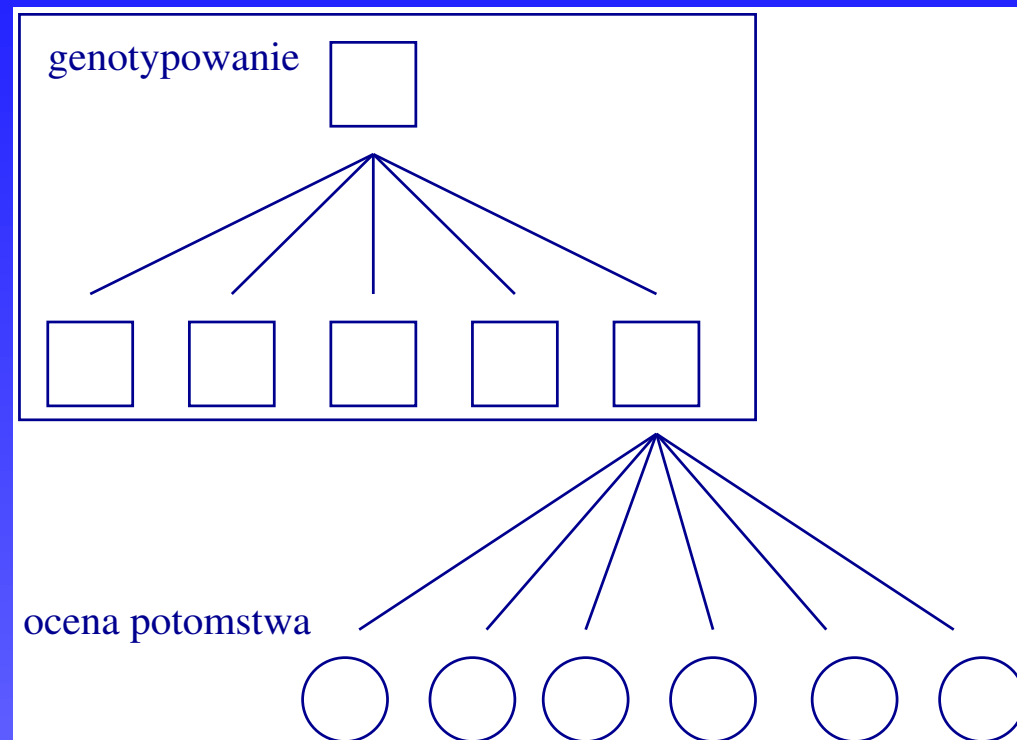
## Daughter design

Pozwala wykryć QTL w czystych rasach. Porównywane są córki krowy, które odziedziczyły po ojcu alterantywne allele markerowe. Genotypowane są buhaje, matki i córki.

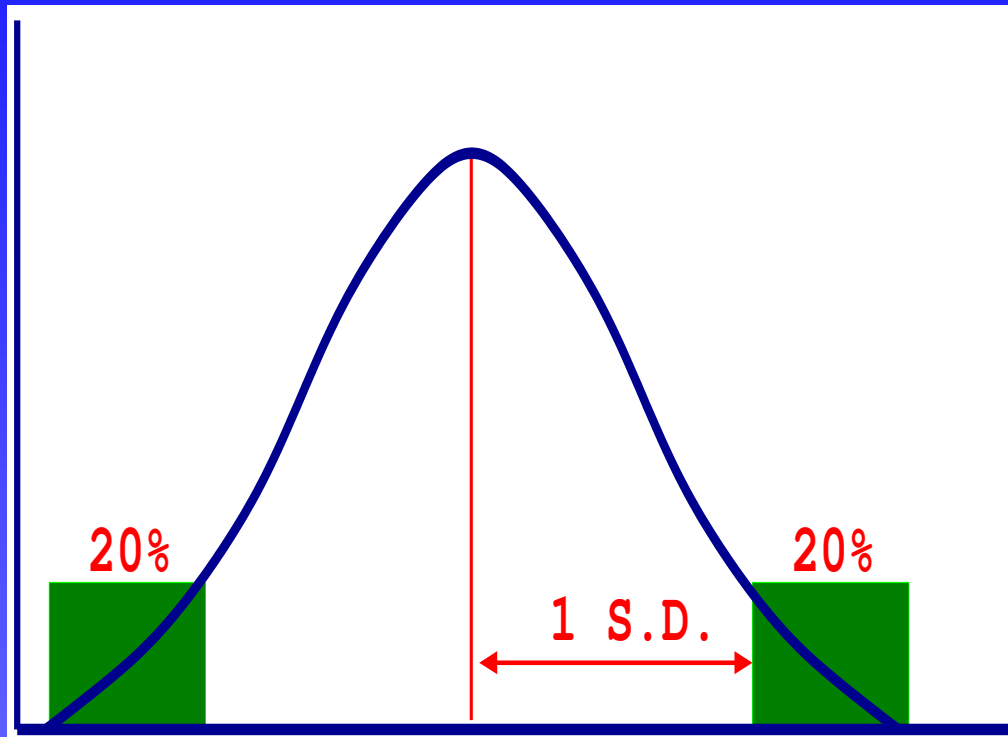


## Granddaughter design

Genotypowane są tylko buhaje (duża oszczędność). Zamiast fenotypów bada się wartość hodowlaną buhajów, ocenioną na podstawie dużej liczby córek. “ $h^2$ ” wartości hodowlanej aż 80%.



## Genotypowanie wybiórcze



Genotypowanie osobników odbiegających 1 S.D. od średniej pozwala zredukować liczbę osobników o 60% bez straty mocy doświadczenia.

## Dokładność analizy sprzężeniowej

- Wymagana - 1cM (klonowanie pozycyjne genów kandydatów)
- Ograniczona
  - ★ dostępnością loci markerowych
  - ★ ilością i rozkładem crossing-over w badanej rodzinie
- Osiągana
  - ★ cechy proste - kilka cM (400 mejoz daje 90% szans na lokalizację genu z dokładnością 1cM)
  - ★ cechy złożone - kilkadziesiąt cM

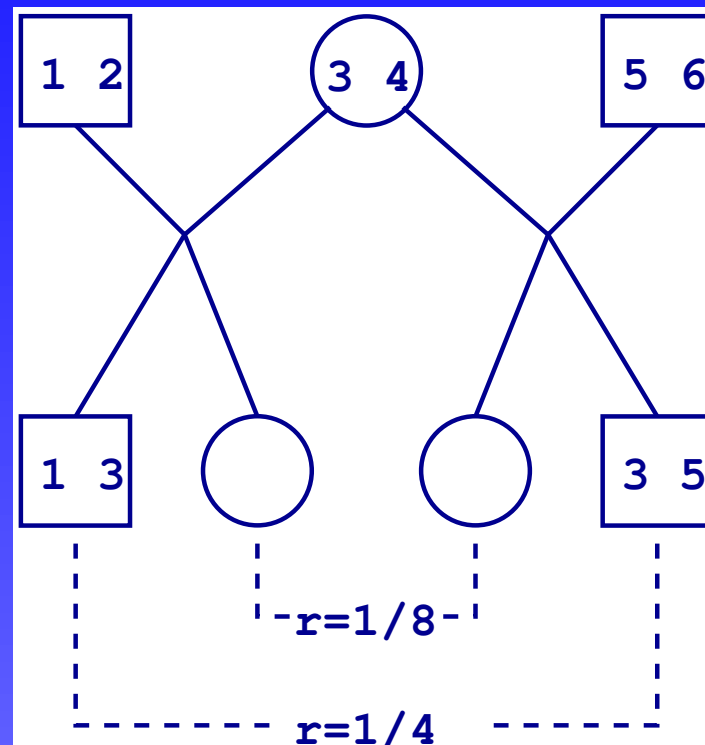
## Mapowanie porównawcze

- Pozwala zawęzić grupę genów kandydatów we wskazanym regionie (50cM to setki genów).
- Wykorzystuje zjawisko konserwatyzmu genetycznego - tu podobieństwa ułożenia genów u różnych gatunków.
- Najważniejsze źródła informacji to mapy człowieka i myszy.

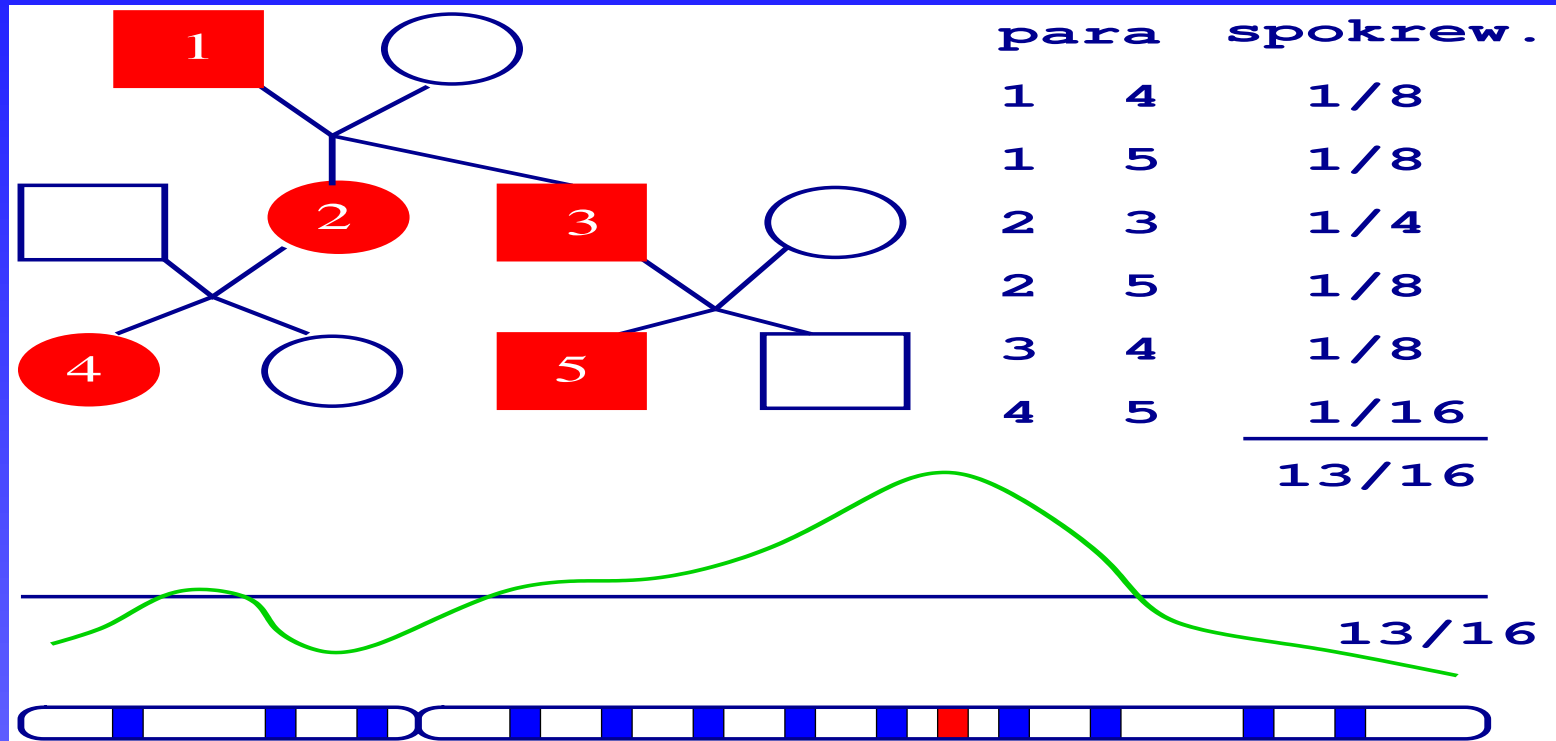
## Mapowanie genów i.p.p.

- Przy b. bliskim sąsiedztwie markera i genu crossing-over prawie nie zachodzi, co daje dużą nierównowagę gametyczną.
- Jeżeli przyczyną choroby jest 1 zmutowany gen, chorzy będą dzielili sąsiadujące markery i.p.p.

## Mapowanie genów i.p.p.



## Mapowanie genów i.p.p.





## Asocjacja allel-cecha

- test  $\chi^2$  dla danego allelu

	testowany allel	
	obecny	brak
gr. dośw.	$n_1$	$n_3$
gr. kontrolna	$n_2$	$n_4$

- siła asocjacji - ryzyko relatywne

$$R = \frac{n_1 \times (n_2 + n_4)}{n_2 \times (n_1 + n_3)}$$

## Przypadek cukrzycy

Badano plemiona Pima i Tohono O'odham (pd. Arizona) w których często występuje cukrzyca typu 2. Wykonano proste porównanie udziału cukrzyków w dwóch grupach haplotypowych. Badanie wykazało, że pewien allel  $Gm^+$  jest negatywnie skorelowany z cukrzycą.

$Gm^+$	ogółem	% cukrzyków
obecny	293	8%
brak	4627	29%

## Populacje łączone

Populacja A - 100 osobników

allel	N	chorych	% chorych
obecny	80	16	20%
brak	20	4	20%

Populacja B - 100 osobników

allel	N	chorych	% chorych
obecny	30	15	50%
brak	70	35	50%

A+B - 200 osobników

allel	N	chorych	% chorych
obecny	110	16+15=31	28%
brak	90	4+35=39	43%

## Przypadek cukrzycy - c.d.

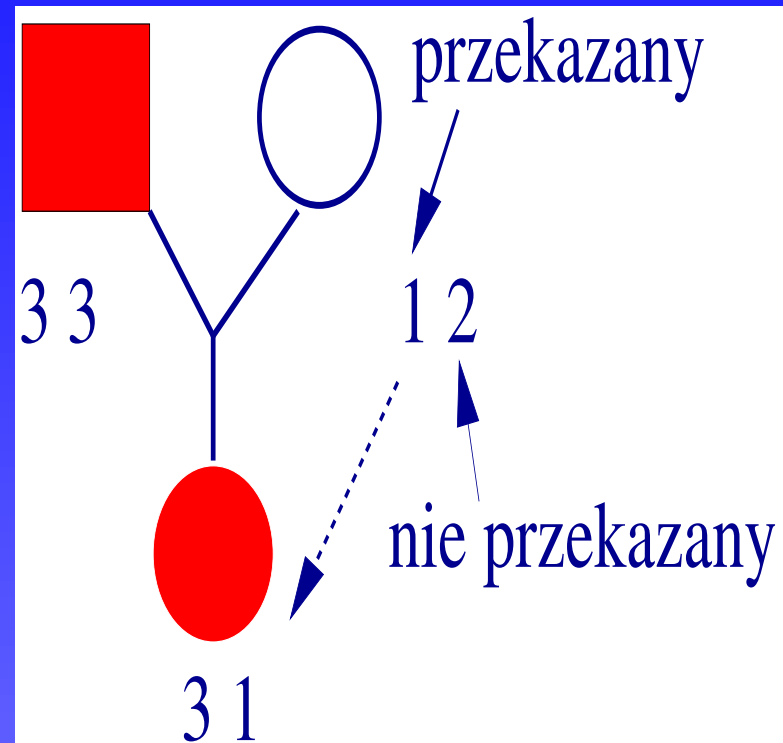
Dalsze badanie wykazało, że asocjacja jest spowodowana przemieszaniem plemion z Kaukazami. U Kaukazów allel  $Gm^+$  występuje z częstotliwością 67%, a u rodowitych Pima i Tohono <1%. Badania ograniczone do rodowitych Pima i Tohono nie wykazały asocjacji.

$Gm^+$	ogółem	% cukrzyków
obecny	17	59%
brak	1764	60%

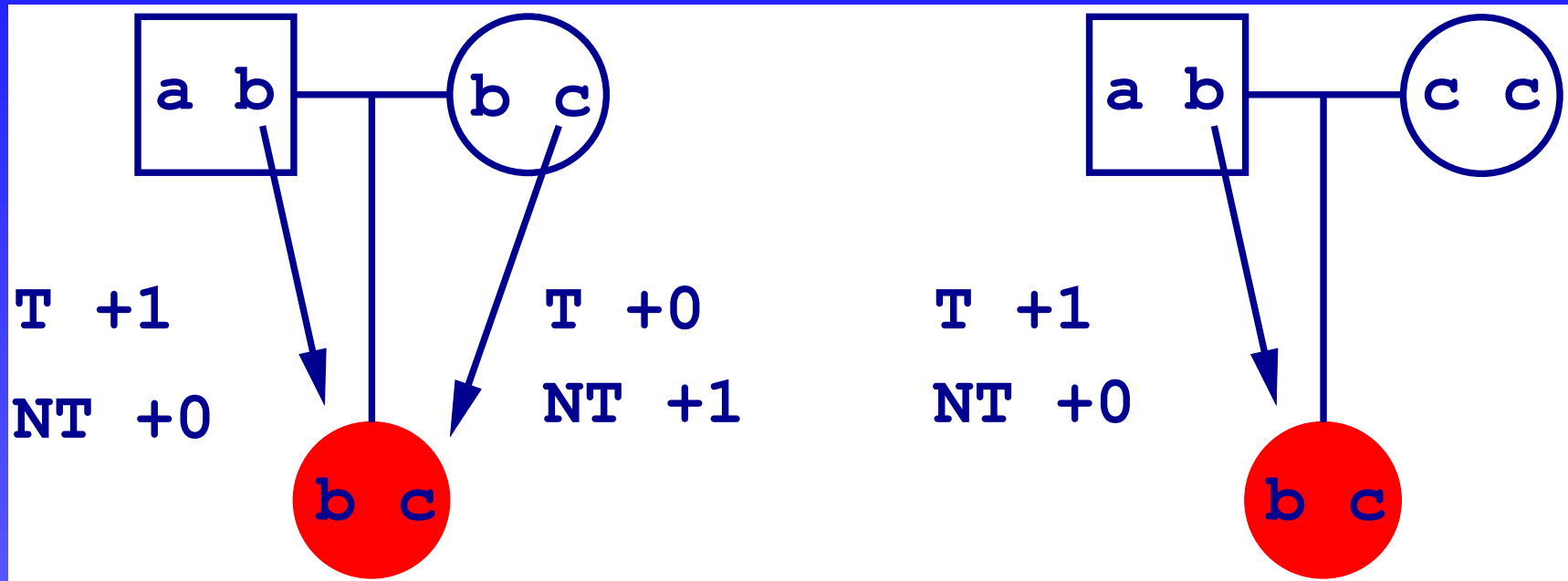
## Test TDT

### Transmission disequilibrium test

- porównuje ile razy dany allel został przekazany (T) chorym a ile nie (NT)
- przy braku sprzężenia  $T \simeq NT$
- $\chi_1^2 = (T - NT)^2 / (T + NT)$



## Test TDT



## Test TDT

Przetestowano 21 markerów u 455 rodzin cukrzyków typu 1. Jeden z trzech alleli markera *D2S152* jest mocno powiązany z cukrzycą. Obecność tego allelu może wskazywać na ryzyko zachorowania.

allel	T	NT	$\chi^2$	<i>p</i>
228	81	45	10.29	0.001
230	59	73	1.48	0.223
240	36	24	2.40	0.121

## Podsumowanie

- Określenie pozycji genu jest pierwszym krokiem do jego identyfikacji.
- Najłatwiej określić pozycję genów względem genów markerowych.
- Dla wielu gatunków powstały (markerowe) mapy sprzężeniowe stanowiące macierz do lokalizacji ważnych genów.
- Najważniejsze źródła markerów genetycznych to RFLP, mikro- i minisatelity oraz SNP.
- Analiza sprzężeniowa jest najważniejszym narzędziem mapowania. Bazuje na obserwacji rekombinacji genów.



- Geny sprzężone mają tempo rekombinacji  $< 0.5$ .
- Funkcja mapowa pozwala wyliczyć odległość genetyczną (cM) z oszacowanego tempa rekombinacji.
- Test lod score testuje czy dwa geny są sprzężone.
- QTL można wykryć analizując mieszańce mocno zróżnicowanych ras lub badając poszczególne rodziny.
- Do zmapowania QTL potrzeba setek osobników, dla cech prostej tylko kilkanaście.
- Wstępna lokalizacja QTL daje bardzo niedokładny wynik. Dokładna lokalizacja QTL stanowi poważne wyzwanie.

- Informacja genomowa może być przenoszona między gatunkami przez mapowanie porównawcze.
- Bliskie sprzężenie utrzymuje asocjacje allel (markerowy) - cecha przez wiele pokoleń.