

# BIOINFORMATYKA

---

1. Wykład wstępny
2. Struktury danych w badaniach bioinformatycznych
3. Bazy danych: projektowanie i struktura
4. Bazy danych: projektowanie i struktura
5. Zastosowanie przykładowych programów do analizy danych
6. Zastosowanie przykładowych programów do analizy danych
7. Powiązania pomiędzy genami: równ. Hardyego-Weinberga, wsp. rekombinacji
- 8. Analiza sprzężeń - teoria**
9. Analiza sprzężeń
10. Analiza asocjacyjna
11. Analiza asocjacyjna
12. Symulacje komputerowe, jackknife, bootstrap
13. Monte Carlo Markov Models
14. Metody klasyfikacyjne
15. Wykład podsumowujący

- 1. Genetyczna determinacja cech ciągłych**
- 2. Poszukiwanie genów przy pomocy analizy sprzężeń**
- 3. Estymacja prawdopodobieństwa genotypu genu**
- 4. Testowanie istotności efektów genów**

**GENETYCZNA  
DETERMINACJA CECH  
CIĄGŁYCH**

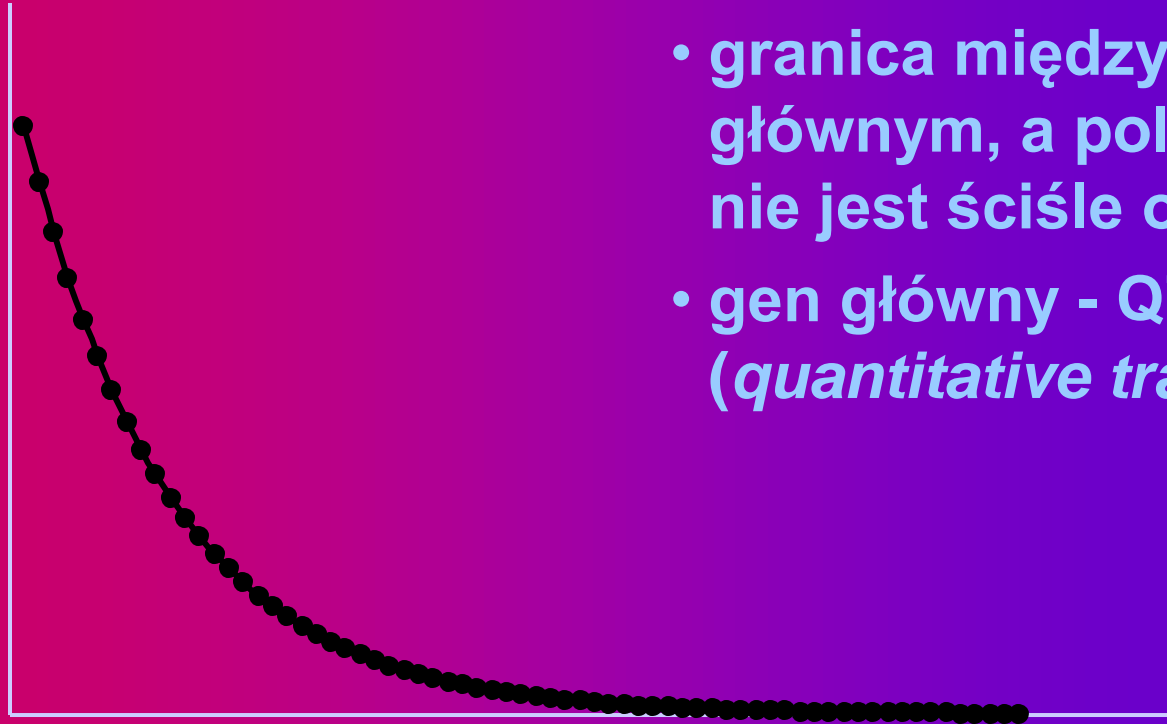
# GENETYCZNA DETERMINACJA CECH CIĄGŁYCH

$$y = \mu + s + g + i + e$$

GENY  
GŁÓWNE

POLIGENY

efekt pojedynczego genu

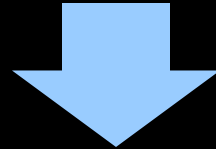


liczba genów

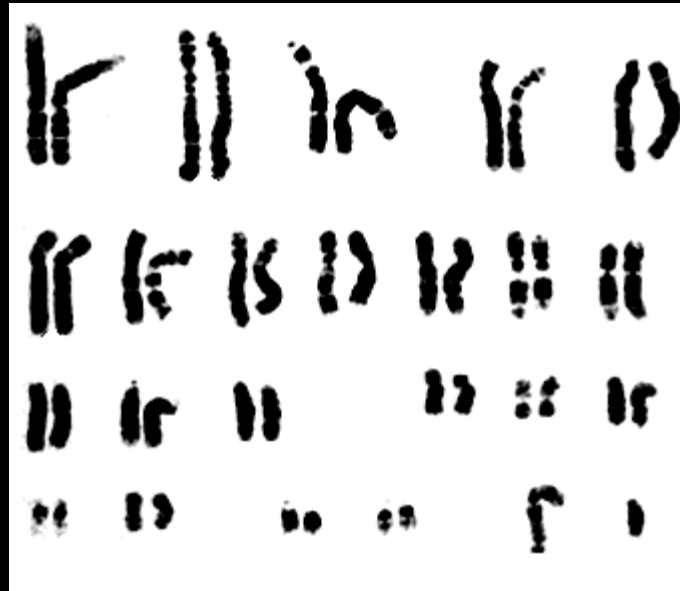
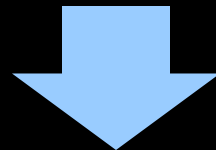
- granica między genem głównym, a poligenem nie jest ściśle określona
- gen główny - QTL (*quantitative trait locus*)

# GENETYCZNA DETERMINACJA CECH CIĄGLYCH

## POWIĄZANIE WARTOŚCI FENOTYPOWEJ CECHY Z SEGREGACJĄ MARKERÓW



- wartość fenotypowa cechy
- genotypy markerów
- spokrewnienie osobników



# GENETYCZNA DETERMINACJA CECH CIĄGŁYCH

TTCTTTCTGTGGTGCTTTACAAAATCTTGAACCTTCTGGAATTGATCAAGCAGATAGACG  
AACGAAATACTGGAATAACAGTTAAAGATCGTGCTGCTTTTTAAAAAATTTTAGAAGCT  
ACCAAACAAAGCAAATTCAGTGTATTGCACCTAATTGCCAAAAACAAGTCTCTCCTTT  
ACAATATTCGAAAAATAATAACTTTATATATAATTCCGGTACTACAAAGGGTATAGTTT  
TGGATAACAGGCATGTGTTTAATATCTTACAAAATCTTCCACAAACGTTTAAATTATTG  
TTAACCCCTTCGAATGCTCATCAAATCGTATCTCCCGAAAATGTCTTTTATGCTAATAG  
TATCTTACTTCCACCACATAATCTACGAACTATCAATGTTTATGATGGTCAGGTTACGA  
GTTTGTTAACAAAGTGATTTGAATCTGATAATGCGAAGAGTTGCTAATAATGAGACAAAT  
GCAAAAATACAAAAATCTTGGATTCTATCGATAACAGCCGAGGTGCCAATCCATATGC  
TACAAATAAAAAGCTTACTTTGGATACTTTGACAGGTGGACACTCAAAGAATCTTATT  
TGCGAAGTTATATTAATGGCAAACGTATTCCTGAGACTGCCAGAGCTGTAATCGAACCC  
TCTATGAATAAAACTGGCTTTATTGAAGTACCATCTTACATTTTAAACAAGTTAAGAGA  
TGTTGTCTTTTATAATCACGTTACGAAAGATAACATACTCAAAGTCTTCAAACGAAC  
AAGCTTTTCTAACATATATCAAAGTGATCATAATTCTGAAAATCCTTATATGGTTTAT  
GATTTAGCACAGAAGAATGGATATTTAACCTTGGCTCCTAATTTCCGGTGATATTTTCGA  
AAAAGGAAAGAGGAAGGTGGTTTTGTAACATTTGCAGACATCCATCTATCTGGTTAA  
CTAATATCCAATCTGGTATAATAAAAAGATCAGAAGGGTTTACTATTAACATCCCAACC  
ACAATTTGCACATCTTTTAAATGCTGATTTTGATGGAGATGAGATGACAATATATTCTTT  
CAAATCCCCATGTGCCAATCTCGAACAAGCTTTGATTATGAACCTCACGAAATCTCTTCA  
AAAATCTATAACAAGCAATCCAATGTTCCGGCTTGGTCCAAGATCAAATACCAGCCTTG  
AATAAGTTATATAGACGACAAAATTATACATATAACGATGCGTTGGTGATTTTAGGACA  
ATTCGGATTTCTGTTAACACCTGGAAAAGATAATTATACCGGAAAAGATATACTTTCTT  
GTGTATTCCCAAACATTATACACTCAAAGGAATTGTTGAAAATGGCGAACTTATTTTG  
GAGAATTTTACAAATAAACTCGTTTCCGCAAATTCCTCAAAGTCCATCTTTGGGCATCT  
TGTTTTATTTTATGGACAAGAGTATGGTTTGACTATATTGGATACAATGCGAGATATTG  
TTCAAATTTTATTACACATTTTGGTTTTCAGTGTA AAAATCCGAGATATGATCCCAAGC  
CCAAAATTTTGGATATTCTAGAAAAGATCGTAGACCAAGAAGTGGATAAAAATTGATAA  
ACAAACAAAACCTTCTATATGACGATATCGAACAAGGTAAGGTTATAATCAACTCTTATG  
ATGATATTTCTGAGTTCAGATTA AAAAATGTGGCTATTATGAAAAAGAACTAGAAAGC  
AACTTTTGGAACTTTTGGATGAATATTATGATGAAGACAATAATTTCTAGAGATGTA  
TAGAACGGGATATAAGGTCAACATTAACGAACTTCTCTCTATTATGTGTTTCTCGGGTT  
TTAAAAATTATGGAAATATCGAAATGATTACACCGGGTCTTAATGGTAAACATCTTTG  
TTTAGCTTACCAGATTCTATAAACTTACAAGATTATGGGTTCATCAAAGCTCTATTGC  
CAAAGGGTTAACGTTTGAAGAATATGCTACAATCGTAAAACAAGAAGCTTTTCCACAAA  
TTGTTAATGTTACAACCTGGTACTTCACAAACAGGATTTTTGGGGAAAAAATGGTTAAA  
ATGGCTTCTGAATTC

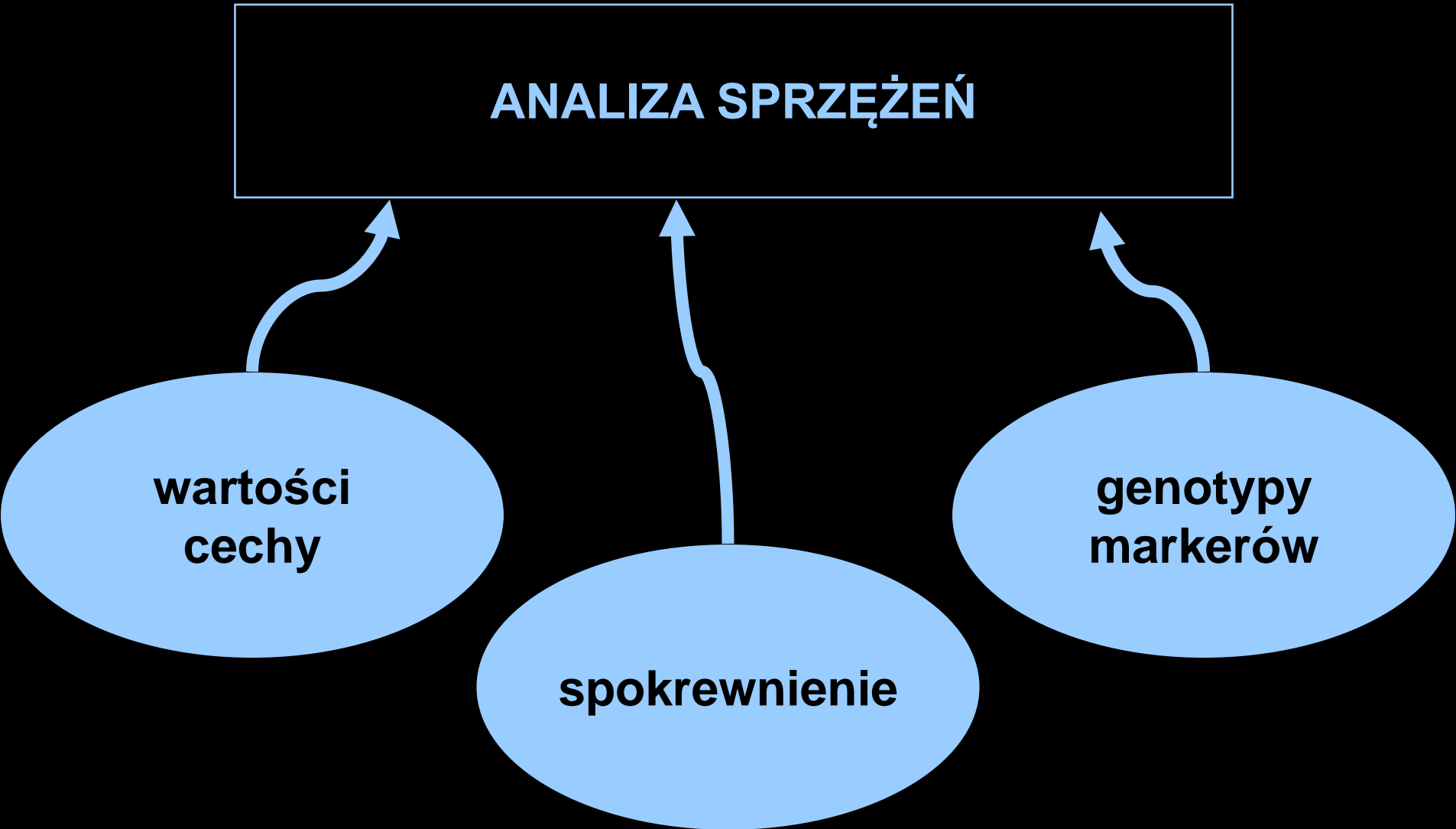
# POSZUKIWANIE GENÓW - ANALIZA SPRZĘŻEŃ

## ANALIZA SPRZĘŻEŃ

1. Statystyczna procedura poszukiwania genów
2. Poszukiwanie najbardziej prawdopodobnej lokalizacji QTL na chromosomie
3. Wykorzystująca prawdopodobieństwo odziedziczenia **(niewidzialnego)** genu głównego
4. Obliczone na podstawie częstości rekombinacji pomiędzy **(widzialnymi)** markerami

# ANALIZA SPRZĘŻEŃ - źródła informacji

---



# ANALIZA SPRZĘŻEŃ - modele statystyczne

---

$$y = \mu + s + g + i + e$$

## LICZBA QTL

- pojedynczy
- kilka sprzężonych QTL
- epistaza pomiędzy genami

## WPŁYW QTL

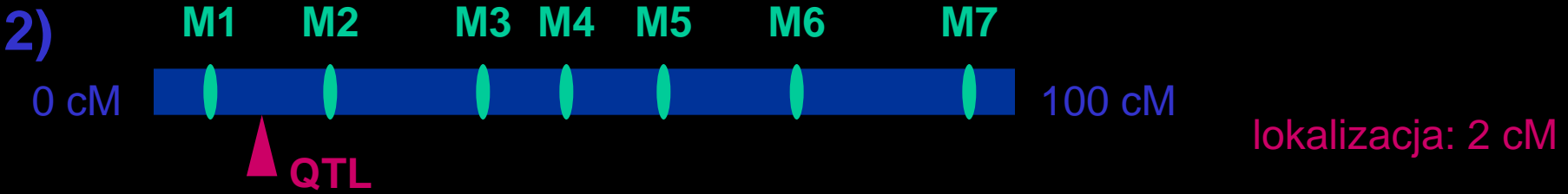
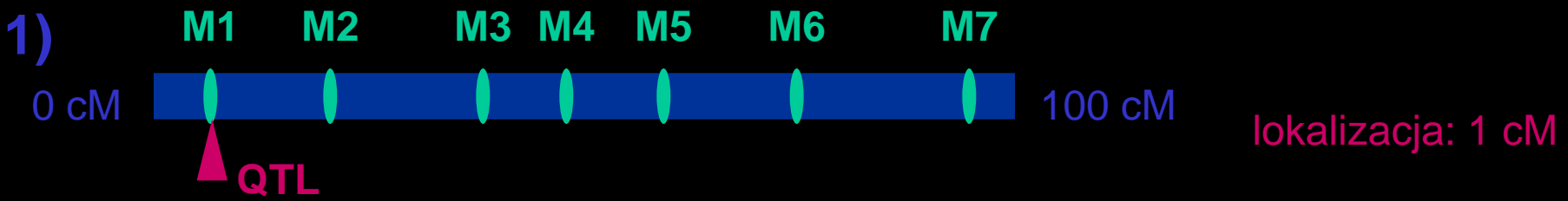
- wpływa na jedną cechę
- wpływa na kilka cech
- efekty epistatyczne

## OKREŚLENIE POŁOŻENIA QTL

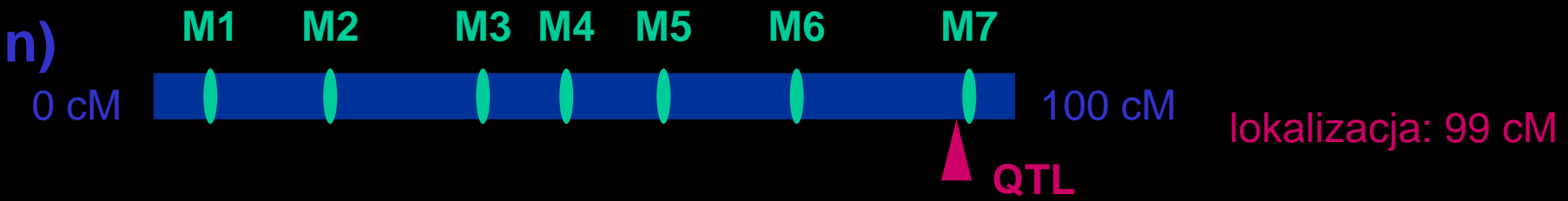
- na podstawie pojedynczego markera
- na podstawie dwu sąsiadujących markerów (przedział markerowy)
- na podstawie kilku markerów

# ANALIZA SPRZĘŻEŃ - poszukiwanie genów

## POSZUKIWANIE NAJBARDZIEJ PRAWDOPODOBNEJ LOKALIZACJI QTL NA CHROMOSOMIE



...



# ANALIZA SPRZĘŻEŃ - poszukiwanie genów

## POSZUKIWANIE NAJBARDZIEJ PRAWDOPODOBNEJ LOKALIZACJI QTL NA CHROMOSOMIE

1)  $y = \mu + q_1 + e$  → prawdopodobieństwo

2)  $y = \mu + q_2 + e$  → prawdopodobieństwo

...

n)  $y = \mu + q_{99} + e$  → prawdopodobieństwo

### ELEMENTY MODELU:

$y$       wartość cechy

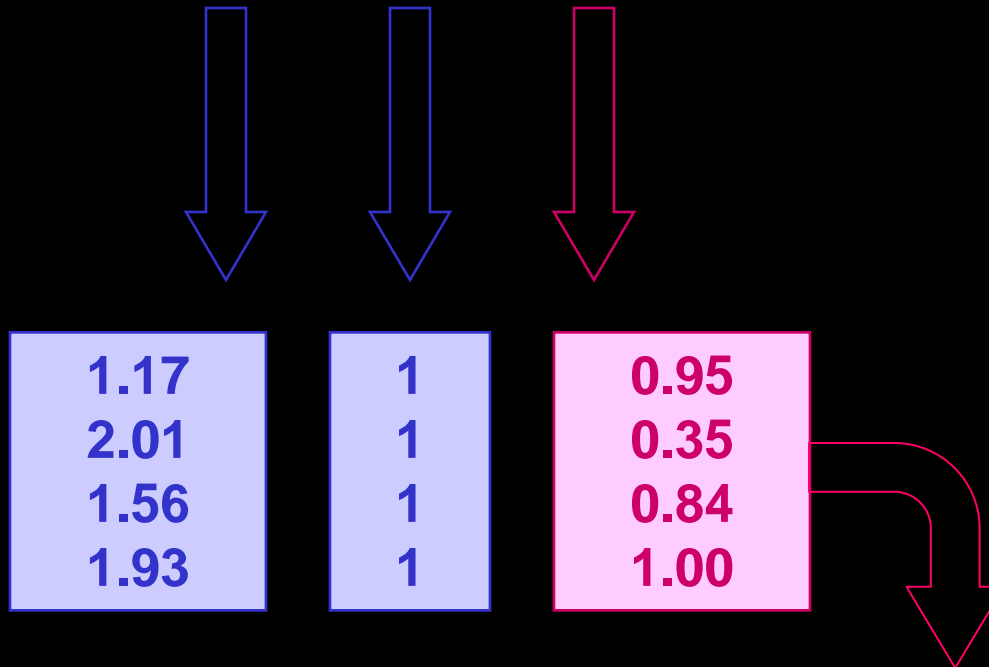
$\mu$       efekty wspólne dla wszystkich osobników

$e$       efekty niemierzalne (błąd)

$q_i$       efekt genu głównego zlokalizowanego w pozycji "i"

# ANALIZA SPRZĘŻEŃ - poszukiwanie genów

$$y = \mu + q_i + e$$



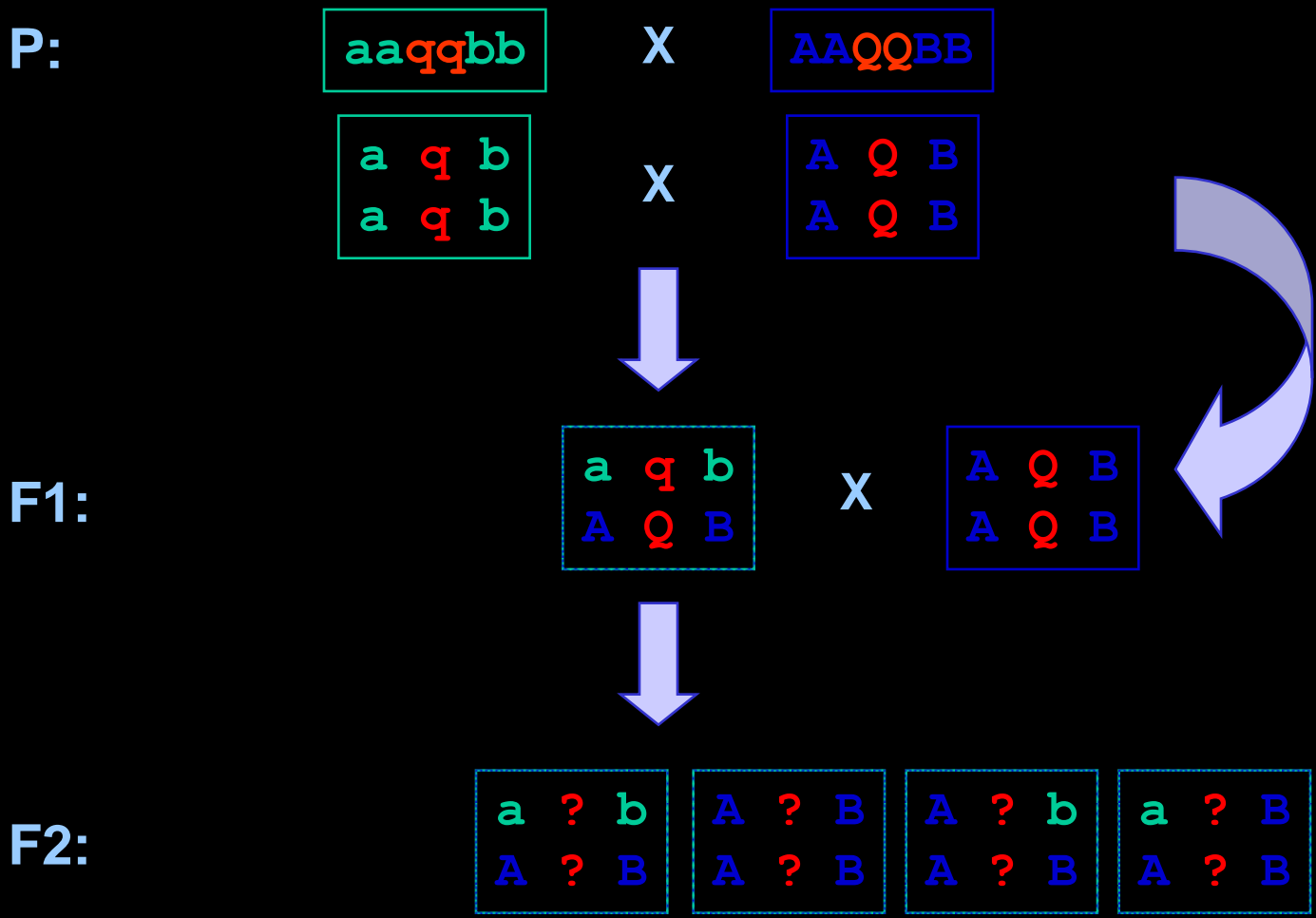
- prawdopodobieństwo, że dany osobnik odziedziczył określony genotyp QTL (QQ, Qq, qq)
- prawdopodobieństwo jest obliczone na podstawie genotypu markerów i informacji o spokrewnieniu
- prawdopodobieństwo jest funkcją częstości rekombinacji pomiędzy markerami

**ESTYMACJA  
PRAWDOPODOBIENSTWA  
GENOTYPU QTL**

# OBLICZANIE PRAWDOPODOBIENSTWA GENOTYPU QTL

## PRAWDOPODOBIENSTWO ODZIEDZICZONEGO GENOTYPU QTL

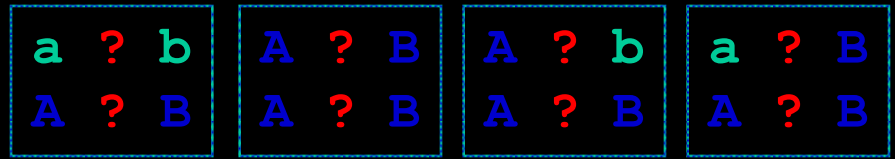
P: linie zinbredowane → homozygotyczne pod względem wszystkich loci (markerkiery i geny główne)



# OBLICZANIE PRAWDOPODOBIEŃSTWA GENOTYPU QTL

## PRAWDOPODOBIEŃSTWO ODZIEDZICZONEGO GENOTYPU QTL

F2:

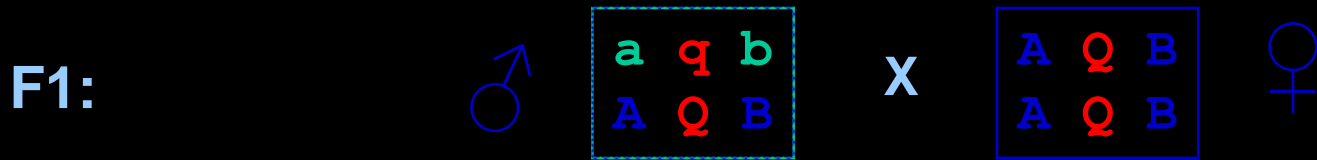


$P ( ?? \rightarrow QQ ) =$   
 $P ( ?? \rightarrow Qq ) =$   
 $P ( ?? \rightarrow qq ) =$

} obliczone  
na podstawie  
markerów

# OBLICZANIE PRAWDOPODOBIEŃSTWA GENOTYPU QTL

## PRZYKŁAD 1:

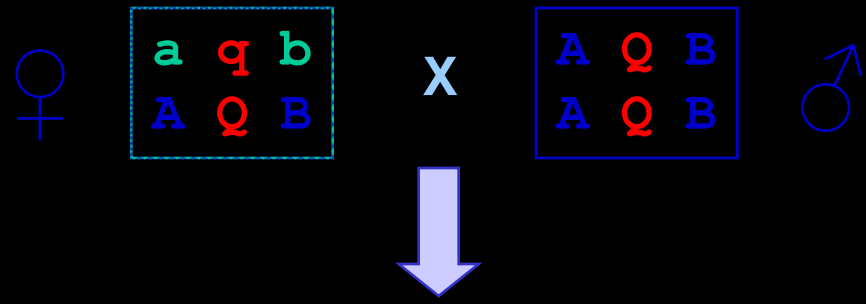


- od matki ♀ zawsze dziedziczony jest genotyp **A–Q–B**
- prawdopodobieństwo genotypu **??** zależy od ojca ♂:
  - $P ( ?? = QQ ) \rightarrow$  rekombinacja pomiędzy **a–?**  
rekombinacja pomiędzy **?–b**  
 $P ( QQ ) = r_{a?} * r_{?b} / ( 1-r_{ab} )$
  - $P ( ?? = Qq ) \rightarrow$  brak rekombinacji pomiędzy **a–?**  
brak rekombinacji pomiędzy **?–b**  
 $P ( Qq ) = ( 1 - r_{a?} ) * ( 1 - r_{?b} ) / ( 1-r_{ab} )$
  - $P ( ?? = qq ) \rightarrow$  genotyp niemożliwy  $P ( qq ) = 0$

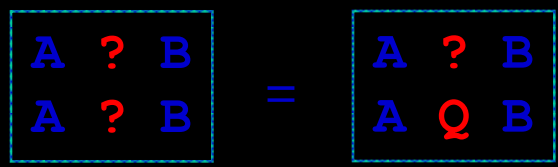
# OBLICZANIE PRAWDOPODOBIENSTWA GENOTYPU QTL

## PRZYKŁAD 2:

F1:



F2: np. osobnik o genotypie



$P(?? = QQ) \rightarrow$

brak rekombinacji pomiędzy A-?  
 brak rekombinacji pomiędzy ?-B  
 $P(QQ) = (1 - r_{a?}) * (1 - r_{?b}) / (1 - r_{ab})$

$P(?? = Qq) \rightarrow$

rekombinacja pomiędzy A-?  
 rekombinacja pomiędzy ?-B  
 $P(Qq) = r_{a?} * r_{?b} / (1 - r_{ab})$

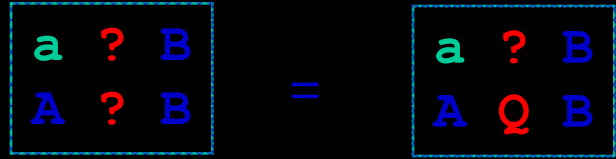
# OBLICZANIE PRAWDOPODOBIENSTWA GENOTYPU QTL

## PRZYKŁAD 3:

F1:



F2: np. osobnik o genotypie



$P(?? = QQ) \rightarrow$

$P(QQ) = r_{a?} * (1 - r_{?b}) / r_{ab}$

$P(?? = Qq) \rightarrow$

$P(Qq) = (1 - r_{a?}) * r_{?b} / r_{ab}$

**TESTOWANIE**

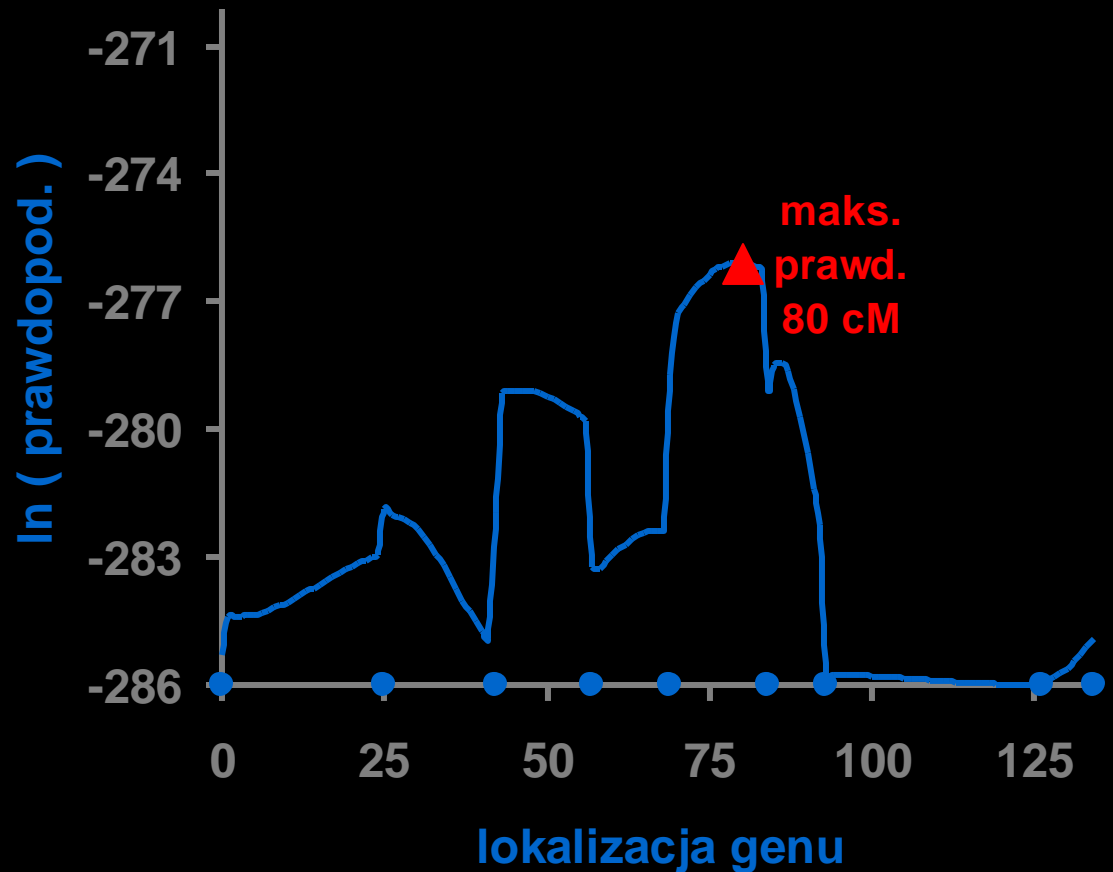
# TESTOWANIE - profil prawdopodobieństwa

1)  $y = \mu + q_1 + e$  → prawdopodobieństwo

2)  $y = \mu + q_2 + e$  → prawdopodobieństwo

...

n)  $y = \mu + q_{99} + e$  → prawdopodobieństwo



## HIPOTEZY

- $H_0$ : w pozycji "i" brak genu głównego  $\rightarrow q_i = 0$
- $H_1$ : w pozycji "i" występuje gen główny  $\rightarrow q_i \neq 0$

## MODELE STATYSTYCZNE

- MODEL0:  $y = \mu + e$
- MODEL1:  $y = \mu + q_i + e$

## TEST

- LRT (*likelihood ratio test*)

$$\text{LRT} = -2 [ \ln \text{Pr} ( \text{MODEL 0} ) - \ln \text{Pr} ( \text{MODEL 1} ) ] \sim \chi^2_{M1-M0}$$

# TESTOWANIE - wnioskowanie

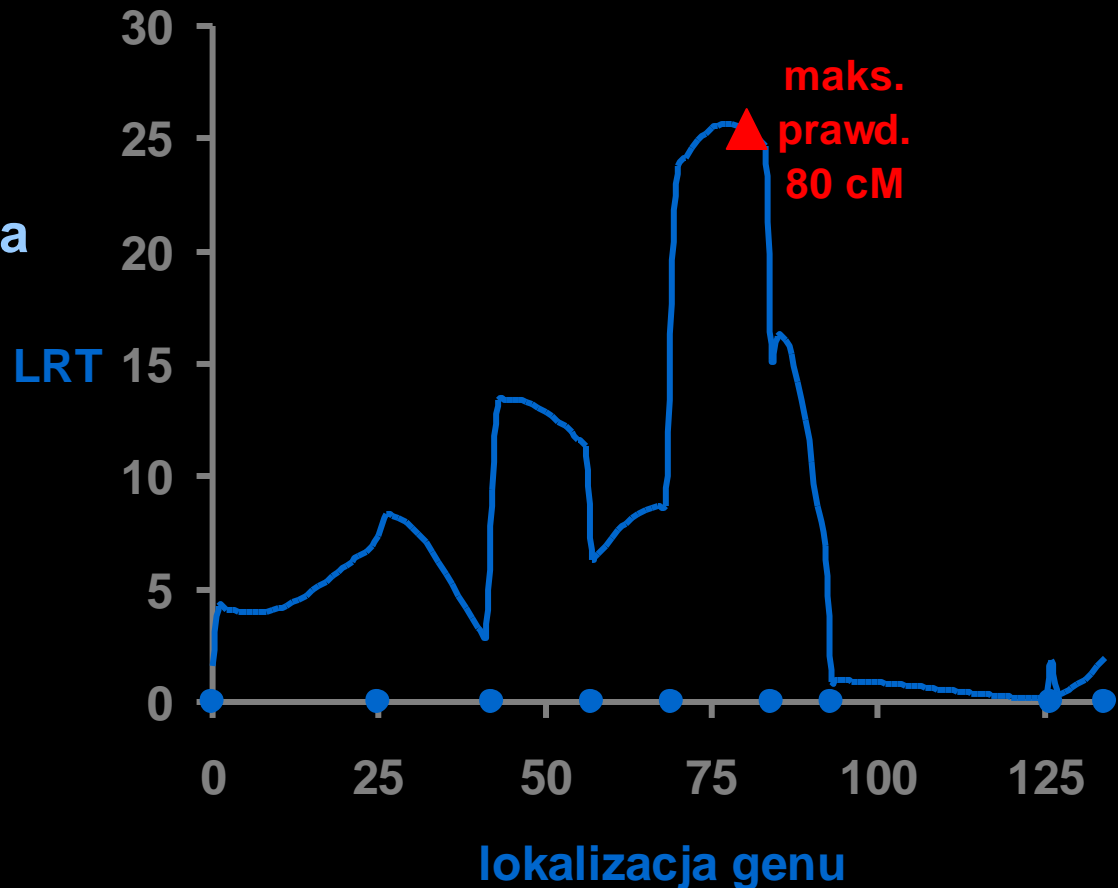
- najbardziej prawdopodobna lokalizacja genu: 80 cM
- dla pozycji 80 cM TLR = 25.46
- założone maksymalne prawdopodobieństwo błędu  $\alpha_{MAX} = 0.01$
- prawdopodobieństwo błędu dla LRT=25.46 wynosi  $\alpha_T = 0.0000045$

•  $\alpha_{MAX} > \alpha_T$

→  $H_1$

→ na dziedziczenie cechy ma wpływ gen główny

→ gen jest zlokalizowany na chrom. 6 ok. 80 cM od pierwszego markera





- 1. Genetyczna determinacja cech ciągłych**
- 2. Poszukiwanie genów przy pomocy analizy sprzężeń**
- 3. Estymacja prawdopodobieństwa genotypu genu**
- 4. Testowanie istotności efektów genów**