

# **SYSTEMY INFORMATYCZNE WSPOMAGAJĄCE HODOWLĘ**

---

**Detekcja genów głównych w populacjach zwierząt  
hodowlanych – wprowadzenie**

- 1. Modelowanie zmienności genetycznej**
- 2. Modele statystyczne**
- 3. Analiza sprzężeń**

# MODELOWANIE ZMIENNOŚCI GENETYCZNEJ



# MODELOWANIE ZMIENNOŚCI GENETYCZNEJ

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

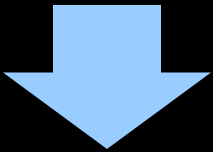
**GENY  
GŁÓWNE**

**POLIGENY**

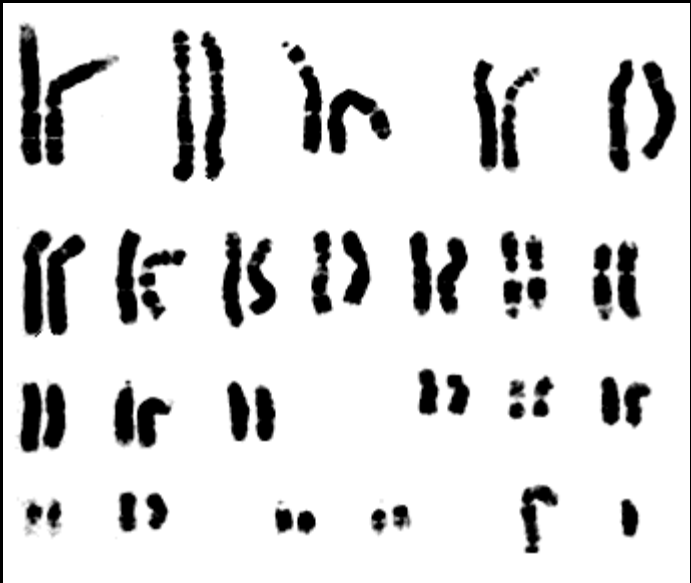
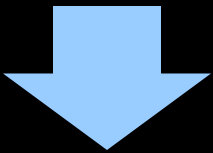


# MODELOWANIE ZMIENNOŚCI GENETYCZNEJ

## POWIĄZANIE WARTOŚCI FENOTYPOWEJ CECHY Z SEGREGACJĄ MARKERÓW



- wartość fenotypowa cechy
- genotypy markerów
- spokrewnienie osobników

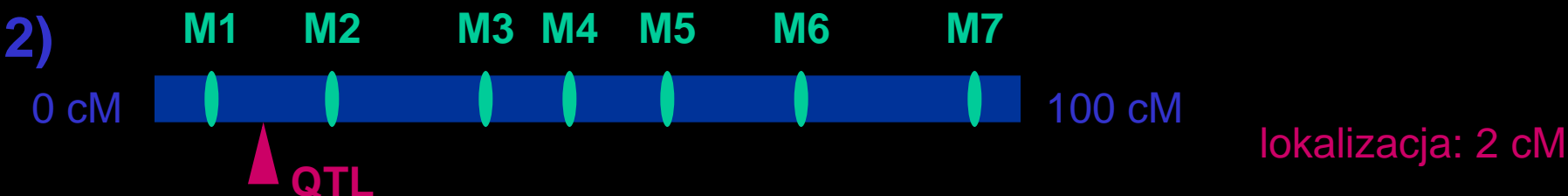
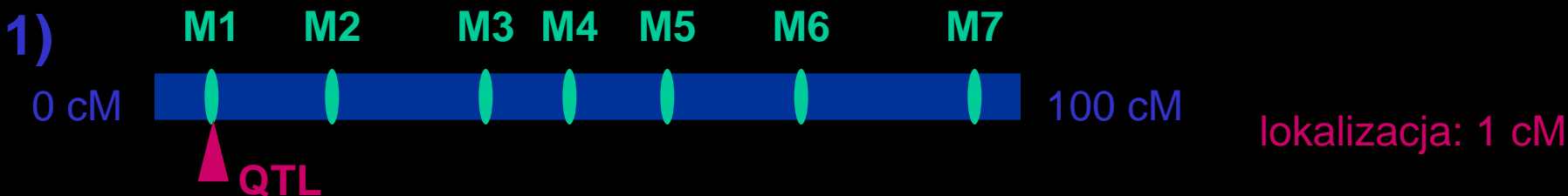


# MODELOWANIE ZMIENNOŚCI GENETYCZNEJ

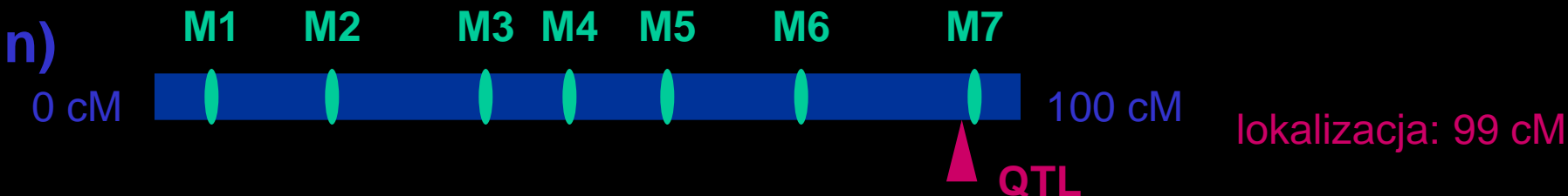
TTCTTTCTGTGGTGCTTTACAAAATCTTGAACCTTCTGGAATTGATCAAGCAGATAGACG  
AACGAAATACTGGAATAACAGTTAAAGATCGTGCTGCTTTTAAAAAAATTTTAGAAGCT  
ACCAAACAAAGCAAATTCAGTGTATTGCACCTAATTGCCAAAAACAAGTCTCTCCTTT  
ACAATATTCGAAAAATAATAACTTTATATATAAATTCGGTACTACAAAGGGTATAGTTT  
TGGATAACAGGCATGTGTTTAATATCTTACAAAATCTTCCACAAACGTTTAAATTATG  
TTAACCCCTTCGAATGCTCATCAAATCGTATCTCCCGAAAATGTCTTTTATGCTAATAG  
TATCTTACTTCCACCACATAATCTACGAACTATCAATGTTTATGATGGTCAGGTTACGA  
GTTTGTTAACAAGTGATTTGAATCTGATAATGCGAAGAGTTGCTAATAATGAGACAAAT  
GCAAAAATACAAAAATCTTGGATTCTATCGATAACAGCCGAGGTGCCAATCCATATGC  
TACAAATAAAAAGCTTACTTTGGATACTTTGACAGGTGGACACTCAAAGAATCTTATT  
TGCGAAGTTATATTAATGGCAAACGTATTCCTGAGACTGCCAGAGCTGTAATCGAACCC  
TCTATGAATAAACTGGCTTTATTGAAGTACCATCTTACATTTTAAACAAGTTAAGAGA  
TGTTGTCTTTTATAATCACGTTACGAAAGATAACATACTCAAAGTCTTCAAACGAAC  
AAGCTTTTCTAACATATATCAAAGTGATCATAATTCTGAAAATCCTTATATGGTTTAT  
GATTTAGCACAGAAGAATGGATATTTAACCTTGGCTCCTAATTTCCGGTGATATTTTCGA  
AAAAGGAAAGAGGAAGGTGGTTTTGTAACATTTGCAGACATCCATCTATCTGGTTAA  
CTAATATCCAATCTGGTATAATAAAAAGATCAGAAGGGTTTACTATTAACATCCCAACC  
ACAATTTGCACATCTTTTAAATGCTGATTTTGATGGAGATGAGATGACAATATATCTTT  
CAAATCCCATGTGCCAATCTCGAACAAAGCTTTGATTATGAACTCACGAAATCTCTTCA  
AAAATCTATAACAAGCAATCCAATGTTCCGGCTTGGTCCAAGATCAAATACCAGCCTTG  
AATAAGTTATATAGACGACAAAATTATACATATAACGATGCGTTGGTGATTTTAGGACA  
ATTCGGATTTCTGTTAACACCTGGAAAAGATAATTATACCGGAAAAGATATACTTTCTT  
GTGTATTCCAAACATTATACACTCAAAGGAATTGTTGAAAATGGCGAACTTATTTTG  
GAGAATTTTACAAATAAATCGTTTTCCGCAAATTCCTCAAAGTCCATCTTTGGGCATCT  
TGTTTTATTTTATGGACAAGAGTATGGTTTGACTATATTGGATACAATGCGAGATATG  
TTCAAATTTTATTACACATTTTGGTTTTAGTGTAAAAATCCGAGATATGATCCCAAGC  
CCAAAATTTTGGATATTCTAGAAAAGATCGTAGACCAAGAAGTGGATAAAATTGATAA  
ACAAACAAAACCTTCTATATGACGATATCGAACAAAGGTAAGGTTATAATCAACTCTTATG  
ATGATATTTCTGAGTTCAGATTAATAAATGTGGCTATTATGAAAAAGAACTAGAAAGC  
AACTTTTGGAACTTTTGGATGAATATTATGATGAAGACAATAATTTCTAGAGATGTA  
TAGAACGGGATATAAGGTCAACATTAACGAACTTCTCTCTATTATGTGTTTCTCGGGTT  
TTAAAAATTATGGAAATATCGAAATGATTACACCGGGTCTTAATGGTAAACATCTTTG  
TTTAGCTTACCAGATTCTATAAATTAACAAGATTATGGGTTCATCAAAGCTCTATTGC  
CAAAGGGTTAACGTTTGAAGAATATGCTACAATCGTAAAACAAGAAGCTTTTCCACAAA  
TTGTTAATGTTACAACCTGGTACTTCACAAACAGGATTTTTGGGGAAAAAAATGGTTAAA  
ATGGCTTCTGAATTC

# MODELOWANIE ZMIENNOŚCI GENETYCZNEJ

## POSZUKIWANIE NAJBARDZIEJ PRAWDOPODOBNEJ LOKALIZACJI GENU GŁÓWNEGO NA CHROMOSOMIE



...



# MODELE STATYSTYCZNE

---

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

$$y = \mu + s + q + p + i + e$$

## LICZBA GENÓW GŁÓWNYCH

- pojedynczy gen główny
- kilka sprzężonych genów głównych
- epistaza pomiędzy genami

## WPŁYW GENU GŁÓWNEGO

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

## OKREŚLENIE POŁOŻENIA GENU GŁ.

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

# MODELE STATYSTYCZNE

## POSZUKIWANIE NAJBARDZIEJ PRAWDOPODOBNEJ LOKALIZACJI GENU GŁÓWNEGO NA CHROMOSOMIE

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

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

...

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

---

## ANALIZA SPRZĘŻEŃ

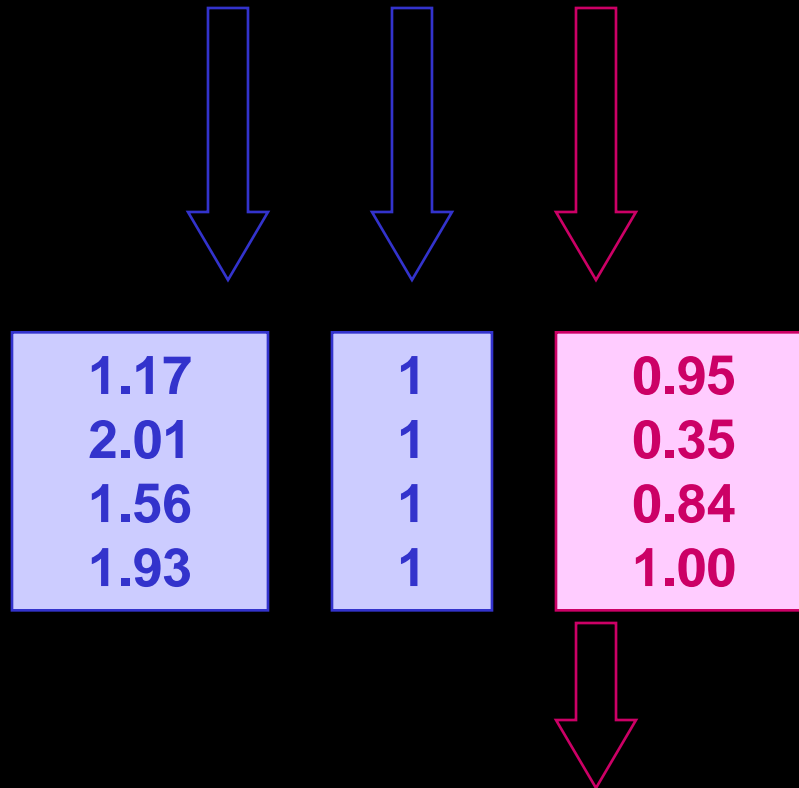
1. Statystyczna procedura poszukiwania genów
2. Wykorzystująca prawdopodobieństwo odziedziczenia **(niewidzialnego)** genu głównego
3. Obliczone na podstawie częstości rekombinacji pomiędzy **(widzialnymi)** markerami

## ANALIZA ASOCJACYJNA

1. Statystyczna procedura poszukiwania genów
2. Wykorzystująca korelacje pomiędzy wartościami cechy, a genotypami markerów
3. **UWAGA:** korelacje nie muszą wynikać z bliskości markera i genu na chromosomie

# ANALIZA SPRZĘŻEŃ

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



- prawdopodobieństwo, że dany osobnik odziedziczył określony genotyp genu głównego (QQ, Qq, qq)
- prawdopodobieństwo jest obliczone na podstawie genotypu markerów

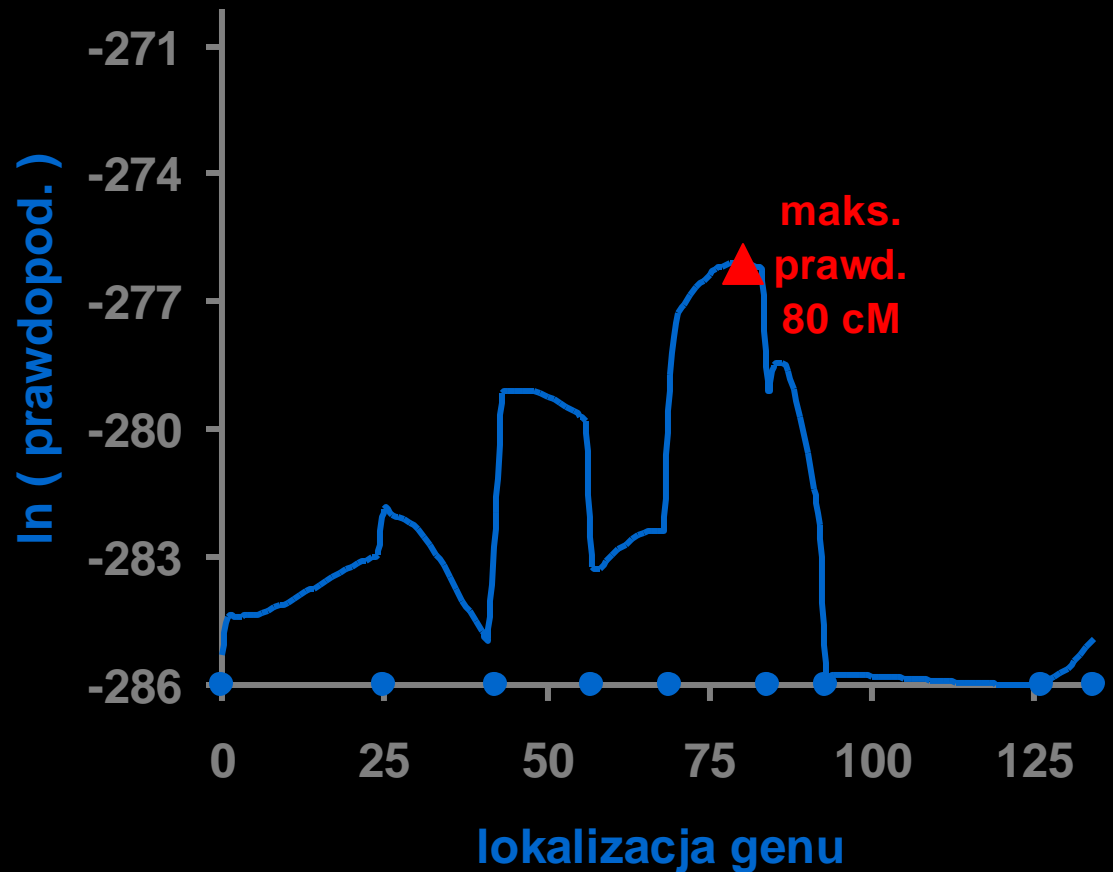
# ANALIZA SPRZĘŻEŃ

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*)

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

# ANALIZA SPRZĘŻEŃ

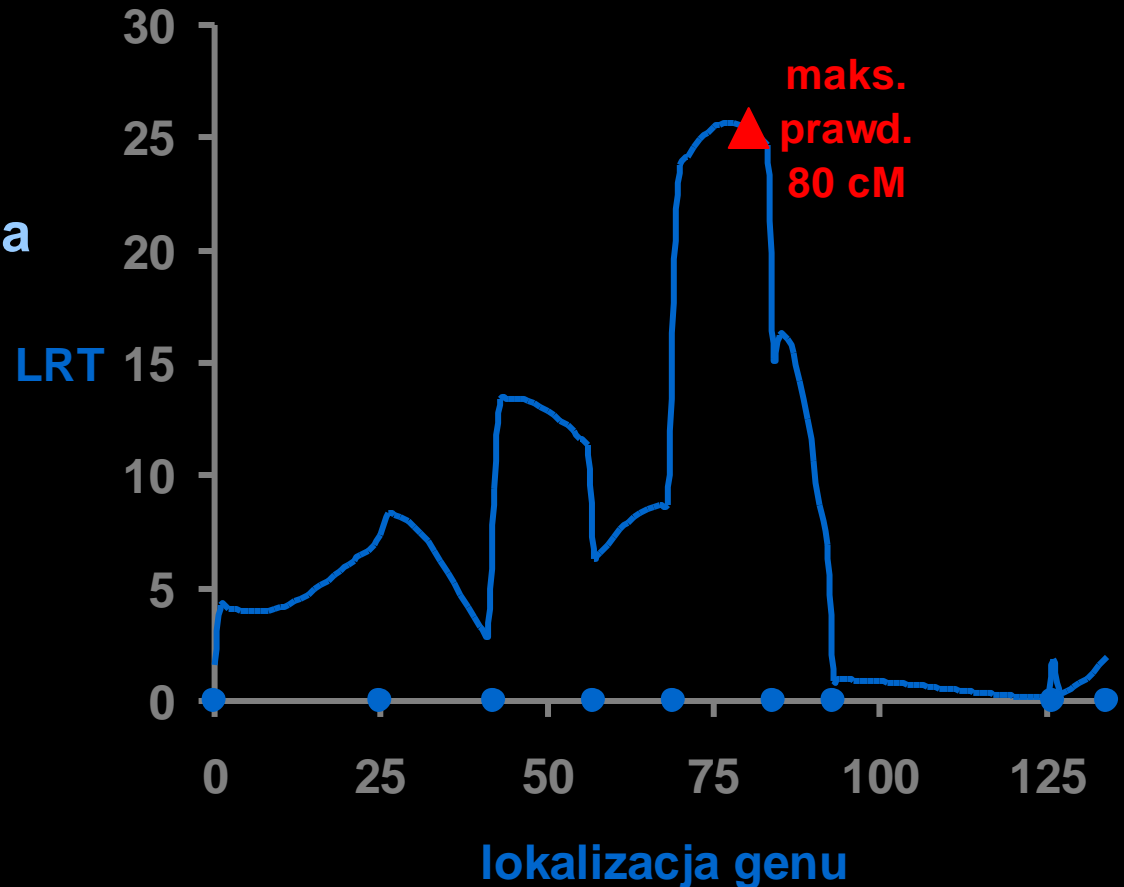
- najbardziej prawdopodobna lokalizacja genu: 80 cM
- dla pozycji 80 cM TLR = 25.46
- założone maksymalne prawdopodobieństwo błędu  $\alpha_{MAX} = 0.01$
- prawdopodobieństwo błędu dla LRT=25.46 wynosi  $\alpha_T = 0.0000045$

•  $\alpha_{MAX} > \alpha_T$

→  $H_1$

→ na dziedziczenie cechy ma wpływ gen główny

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



## AnimalQTLdb

<http://www.animalgenome.org/cgi-bin/QTLdb/index>

This Animal Quantitative Trait Locus (QTL) database (**AnimalQTLdb**) is designed to house all publicly available QTL data on livestock animal species for easily locating and making comparisons within and between species. The database tools are also added to link the QTL data to other types of genome information, such as RH maps, physical maps, and human genome maps. Besides the QTL data from species listed below, we also plan to apply the tool to other animal species where feasible. **JAS** among other journals require that new QTL data be entered into this database before they may be published.



### Pig QTL

There are **4,928** QTLs from **205** publications curated into the database. Those QTLs represent **499** different traits (see **data summary** for details).



### Cattle QTL

There are **2,344** QTLs from **114** publications curated into the database. Those QTLs represent **185** different traits (see **data summary** for details).



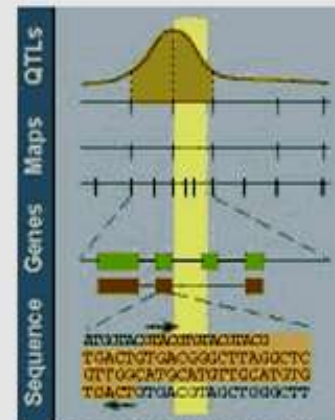
### Chicken QTL

There are **1,290** QTLs from **70** publications curated into the database. Those QTLs represent **164** different traits (see **data summary** for details).



### Sheep QTL

There are **84** QTLs from **15** publications curated into the database. Those QTLs represent **30** different traits. Currently it is under active development at the Faculty of Veterinary Science, The University of Melbourne, Australia (see **data summary** for most recent updates).



This graph is adopted from the Rat GDB with kind permission



**Frequently asked questions:** It is designed to serve as a User's Guide to QTLdb, addressing the use, application and related issues.



**New Data:** If you have new QTL data that has not been included in the current release, please either (1) register as a QTLdb curator to input it directly by yourself (see "Database Curators / Editors" below), or (2) inform our Bioinformatics Team so that we can include it for curation before the next release.

- 1. Modelowanie zmienności genetycznej**
- 2. Modele statystyczne**
- 3. Analiza sprzężeń**