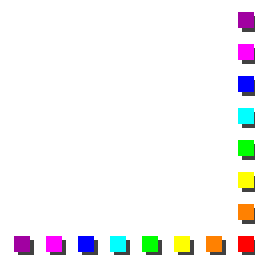


## *Mapeamento de Genes: Delineamentos com Famílias e Estudos Observacionais*

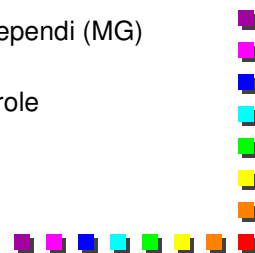
*Júlia Maria Pavan Soler*  
*pavan@ime.usp.br*

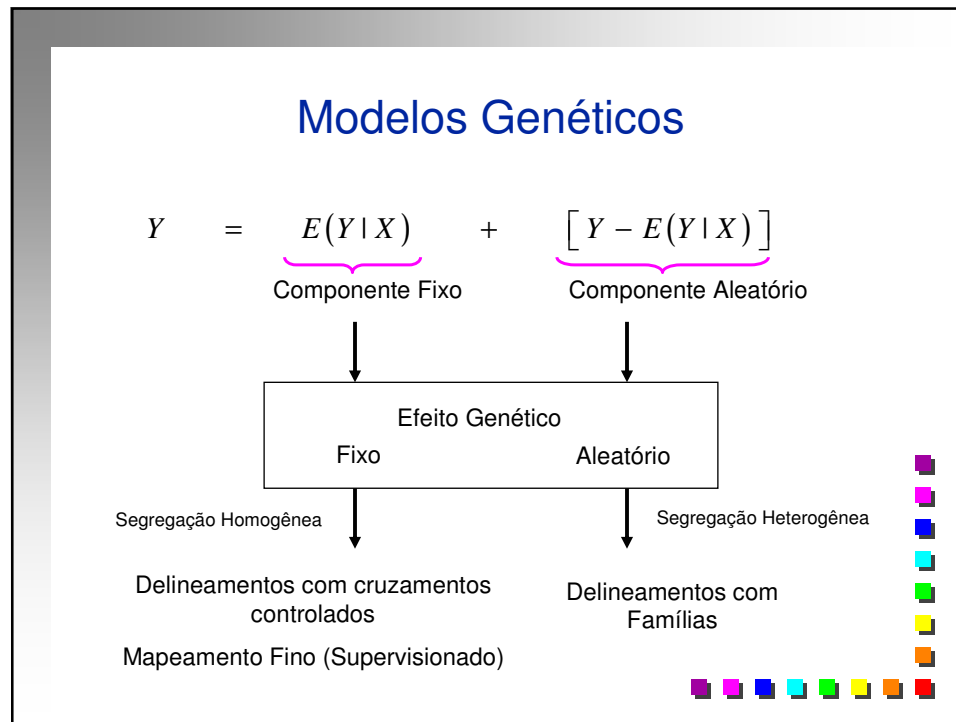
*DEST – U.F.P. – Curitiba*  
*Set / 2007*



## **Programa**

- ✓ Aula 1: Conceitos de Genética Quantitativa e Genômica  
Mapeamento de Genes  
Efeitos Fixos de Genes
- Aula 2: Delineamentos Familiares – Análise de QTL's
- Aula 3: Aplicação - Projeto Corações de Baependi (MG)
- Aula 4: Estudos Observacionais Caso-Controlle  
Perspectivas





### Modelo Genético Quantitativo

$y_{ij}$  : fenótipo do indivíduo i com genótipo j

- Efeito Fixo do Gene (Delin. Cruzamentos Controlados)
 

$$y_{ij} = \underbrace{\mu}_{\text{ef. fixo}} + \underbrace{g_j}_{\text{ef. aleatório}} + e_{ij}$$

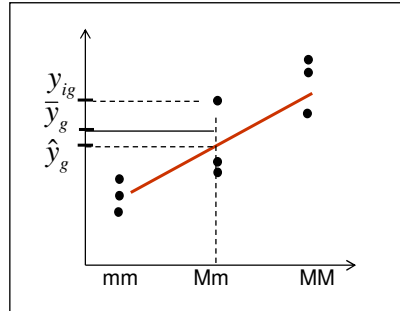
$\Rightarrow E(y_{ij}) = \mu + g_j \quad g_j = aX_{aij} + dX_{dij}$   
 $Var(y_{ij}) = Var(e_{ij}) = \sigma_e^2$   
 $\Rightarrow y_{ij} \sim N(\mu_j; \sigma_e^2)$
- Efeito Aleatório do Gene

$$y_{ij} = \underbrace{\mu}_{\text{ef. fixo}} + \underbrace{g_i}_{\text{ef. aleatório}} + e_{ij}$$

$\Rightarrow E(y_{ij}) = \mu$   
 $Var(y_{ij}) = Var(g_i + e_{ij}) = \sigma_g^2 + \sigma_e^2$

**Justificar!!**

## Componentes de Variação



Identidade útil

$$\begin{cases} y_{ig} - \bar{y} = (\bar{y}_g - \bar{y}) + (y_{jg} - \bar{y}_g) \\ = (\hat{y}_g - \bar{y}) + (\bar{y}_g - \hat{y}_g) + (y_{jg} - \bar{y}_g) \end{cases}$$

$$\sigma_y^2 = \underbrace{\sigma_a^2 + \sigma_d^2}_{\sigma_g^2} + \sigma_e^2$$

## Exemplos de Insucesso na Modelagem de Efeitos Genéticos

$y_{ij}$  : fenótipo do indivíduo  $i$  com genótipo  $j$

- Efeito Aleatório do Gene (indivíduos não relacionados)

$$y_{ij} = \mu_j + e_{ij} = \mu + g_j + e_{ij}$$

$$g_j \sim N(0; \sigma_g^2) \quad \mu_j = \mu + g_j \sim N(\mu; \sigma_g^2)$$

$$e_{ij} \sim N(0; \sigma_e^2) \quad g_j \perp e_{ij}$$

$$\Rightarrow y_{ij} \sim N(\mu; \sigma_g^2 + \sigma_e^2)$$

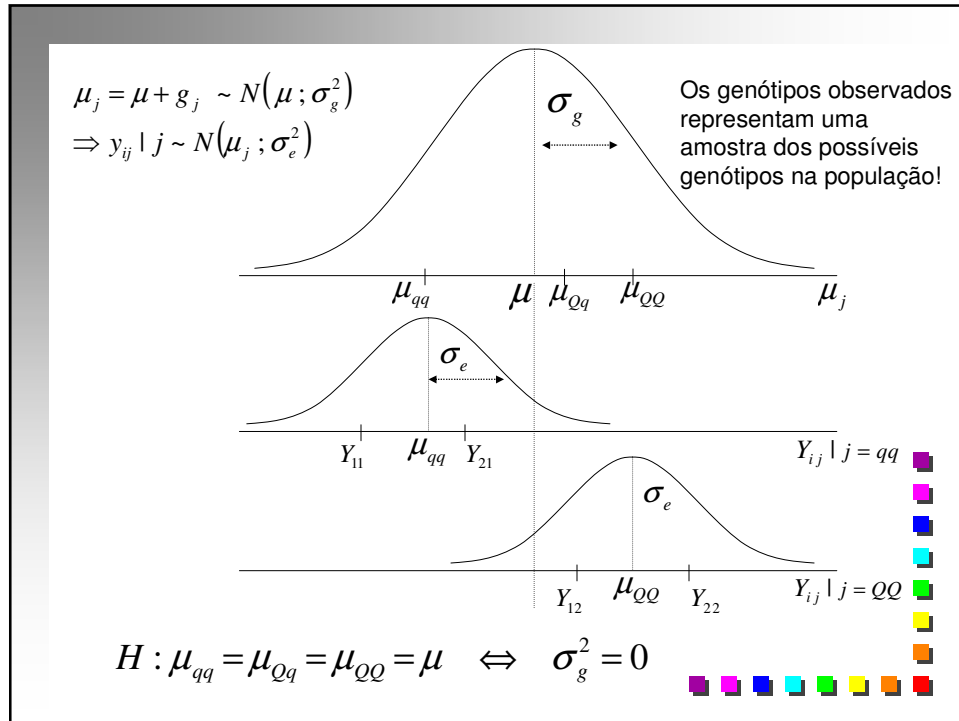
$$Cov(y_{ij}; y_{i'j'}) = \begin{cases} \sigma_g^2 & i \neq i' \quad j = j' \\ 0 & i \neq i' \quad j \neq j' \end{cases}$$

$$\Rightarrow Cov(\mathbf{Y}) = \mathbf{1}\mathbf{1}'\sigma_g^2 + \mathbf{I}\sigma_e^2$$

Genótipos idênticos por estado (ibs)

Estrutura de covariância uniforme

Não é apropriada!



### Exemplos de Insucesso na Modelagem de Efeitos Genéticos

$y_{ij}$  : fenótipo do indivíduo i com genótipo j

#### ▪ Efeito Aleatório do Gene (dados de famílias)

$$y_{ij} = \mu_j + e_{ij} = \mu + g_j + e_{ij}$$

$$g_j \sim N(0; \sigma_g^2) \quad e_{ij} \sim N(0; \sigma_e^2)$$

$$Cov(y_{ij}; y_{i'j'}) = \begin{cases} \left(\frac{1}{2}\right)^r \sigma_g^2 & i \neq i' \quad j = j' \\ 0 & i \neq i' \quad j \neq j' \end{cases}$$

$$\Rightarrow Cov(\mathbf{Y}) = 2\Phi \sigma_g^2 + I\sigma_e^2$$

$\downarrow$   
 $\left(\frac{1}{2}\right)^r$

Genótipos idênticos por descendência (ibd)

Estrutura de covariância depende do grau de parentesco (r).

Ainda assim o modelo Não é apropriado!

Controlar o efeito de outros Locus

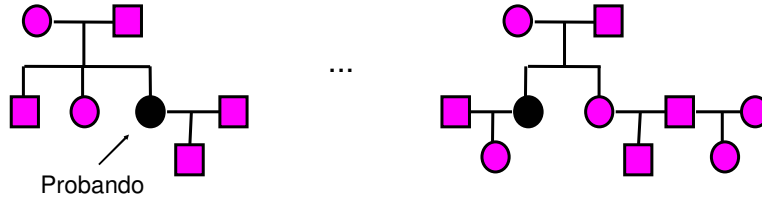
Permitir que a "semelhança" entre os indivíduos dependa do número de alelos ibd que eles compartilham



## Delineamentos com Famílias

Dados de Pedigrees, Heredogramas

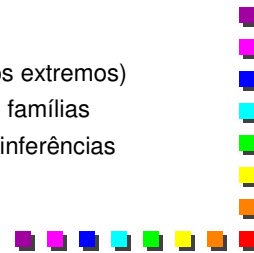
Fundadores



- Segregação Heterogênea dos genes reguladores de fenótipos quantitativos
- Amostragem Aleatória das famílias
- Amostragem por Probandos (Seleção de indivíduos extremos)

⇒ Não Aleatório ⇒ diminui correlação dentro das famílias

⇒ Correções nas inferências



## Modelos Genéticos - Dados Familiares

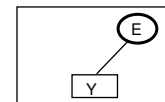
### Modelo Esporádico

$$Y_f = \mu + X_f \beta + e_f; \quad e_f \sim N(0; \Omega_f)$$

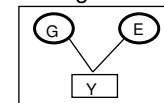
Valor esperado de Y para a família f (covariáveis)

$$\Rightarrow \Omega_f = I_f \sigma_e^2$$

Componente de variância residual



C. V. residual e Poligênico



- Modelo Poligênico: muitos genes de pequeno efeito dispersos pelo genoma

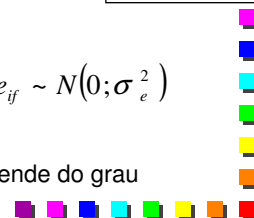
$$Y_f = \mu + X_f \beta + g_f + e_f \quad g_f = (g_{if})_{n_f \times 1}$$

Poligene: pool de genes

$$g_{if} = \sum_{k=1}^m Q_{ifk} \sim N\left(0; \sigma_g^2 = \sum_{k=1}^m \sigma_{Q_k}^2\right) \quad e_{if} \sim N(0; \sigma_e^2)$$

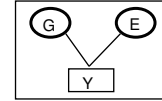
$$\Rightarrow \Omega_f = 2\Phi_f \sigma_g^2 + I_f \sigma_e^2$$

Covariância: depende do grau de parentesco



- Note que no Modelo Poligênico temos:

$$Var(Y_{if}) = \sigma_y^2 = \sigma_g^2 + \sigma_e^2 = \sum_{k=1}^m \sigma_{Q_k}^2 + \sigma_e^2$$



$$Cov(Y_{if}; Y_{i'f}) = Cov(\mu + x_i^t \beta + \sum_{k=1}^m Q_{ikf} + e_{if}; \mu + x_{i'}^t \beta + \sum_{k=1}^m Q_{i'kf} + e_{i'f})$$

$$= Cov(\sum_{k=1}^m Q_{ikf}; \sum_{k=1}^m Q_{i'kf}) = \sum_{k=1}^m Cov(Q_{ikf}; Q_{i'kf})$$

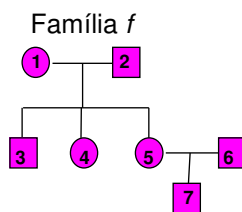
$$= \sum_{k=1}^m \underbrace{(1/2 k_{1k} + k_{2k})}_{\text{coeficiente de Cotterman (1940)}} \sigma_{Q_k}^2 \approx E\left(\frac{1}{2} k_{1i} + k_{2i}\right) \sum_{i=1}^n \sigma_{Q_i}^2 = 2\phi_{jj'} \sigma_g^2$$

Efeito aleatório linear do loco

$k_{1i}$  é a probabilidade de compartilhar 1 alelo IBD no  $i$ -ésimo QTL  
 $k_{2i}$  é a probabilidade de compartilhar 2 alelos IBD no  $i$ -ésimo QTL

A aproximação é feita sobre todo o genoma  
 Suposição: Efeito de dominância é desprezível

### Estruturas de Covariância



**Matriz de Relacionamento  
(Grau de Parentesco)**

$$2\phi_{ij} = \left(\frac{1}{2}\right)^r \quad r: \text{ grau de parentesco}$$

⇒ Para o cálculo é necessário conhecer a estrutura familiar

$$2\Phi = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} & \begin{bmatrix} 1 & 0 & 1/2 & 1/2 & 1/2 & 0 & 1/4 \\ 0 & 1 & 1/2 & 1/2 & 1/2 & 0 & 1/4 \\ 1/2 & 1/2 & 1 & 1/2 & 1/2 & 0 & 1/4 \\ 1/2 & 1/2 & 1/2 & 1 & 1/2 & 0 & 1/4 \\ 1/2 & 1/2 & 1/2 & 1/2 & 1 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/2 \\ 1/4 & 1/4 & 1/4 & 1/4 & 1/2 & 1/2 & 1 \end{bmatrix} \end{matrix}$$

$$Cov(y_{ij}; y_{i'j'})$$

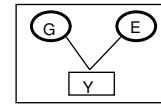
$$= \begin{cases} \left(\frac{1}{2}\right)^r \sigma_g^2 & i \neq i' \quad j = j' \\ 0 & i \neq i' \quad j \neq j' \end{cases}$$

## Modelo de Componentes de Variância Genético

### Modelo Poligênico:

$$Y_f = X_f \beta + g_f + e_f$$

$$Y_f = X_f \beta + \varepsilon_f$$



Dois componentes de var.: poligene e resíduo

⇒ Resíduo Condicional:

$$e = Y - E(Y | g) = Y - (X\beta + g) \quad \text{Resíduo confundido}$$

⇒ Resíduo Marginal:

$$\varepsilon = Y - E(Y) = Y - X\beta = g + \varepsilon \quad \text{Resíduo puro}$$

⇒ Efeitos Aleatórios:

$$g_{jf} = E(Y_{jf} | g_{jf}) - E(Y_{jf})$$

## Modelo de Componentes de Variância

### Estimação de Parâmetros

$$Y_f \sim N(X \beta; \Omega_f) \Rightarrow \Omega_f = 2\Phi_f \sigma_g^2 + I_f \sigma_e^2$$

*N famílias independentes ⇒ Produto de Normais Multivariadas*

$$\ln L(\beta, \sigma_g^2, \sigma_e^2 | Y) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln |\Omega| - \frac{1}{2} \Delta' \Omega^{-1} \Delta$$



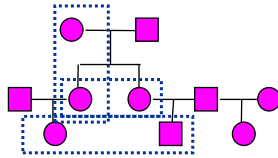
↑  
Matriz bloco diagonal

Estimação simultânea dos efeitos fixos e aleatórios

Algoritmo de Newton-Raphson

⇒ Aplicativo SOLAR (for Linux: [www.sfbr.org](http://www.sfbr.org))

## Dados Familiares - Herdabilidades



### Modelo Poligênico

$$\mathbf{Y}_f = \mathbf{X}_f \boldsymbol{\beta} + \mathbf{g}_f + \mathbf{e}_f$$

$$\Rightarrow \boldsymbol{\Omega}_f = 2\boldsymbol{\Phi}_f \sigma_g^2 + \mathbf{I}_f \sigma_e^2$$

$$\text{Cov}(y_{ij}; y_{i'j'}) = \begin{cases} \left(\frac{1}{2}\right)^r \sigma_g^2 & i \neq i' \quad j = j' \\ 0 & i \neq i' \quad j \neq j' \end{cases}$$

$$\rho(Y_{fi}; Y_{fi'}) = \frac{\text{Cov}(Y_{fi}; Y_{fi'})}{\sigma_Y^2} \cong 2\phi_{ii'} \frac{\sigma_g^2}{\sigma_Y^2} = 2\phi_{ii'} h_g^2$$

$$h_g^2 = \frac{\sigma_g^2}{\sigma_Y^2}$$

Herdabilidade do fenótipo devido ao efeito do poligene

Há interesse em estimar herdabilidades



## Modelo de Componentes de Variância

(Almasy and Blangero, 1998; de Andrade et al., 1999)

- Testes de Hipóteses sob o modelo Poligênico:

$$\boldsymbol{\Omega} = 2\boldsymbol{\Phi} \sigma_g^2 + \mathbf{I}_f \sigma_e^2 \quad H_0: \sigma_g^2 = 0$$

→ Estatística Razão de Verossimilhanças

$$\Lambda = 2 \left[ \ln L(\hat{\theta}) - \ln L(\theta_0) \right] \sim \text{Mistura } \frac{1}{2} \chi_0^2 + \frac{1}{2} \chi_1^2$$

Estatística Lod Score

$$\text{LOD} = \log_{10} L(\hat{\theta}) - \log_{10} L(\theta_0) = \frac{\Lambda}{2 \ln 10}$$

↑
↑

Modelo Poligênico      Modelo Esporádico

Rejeição de  $H_0 \Rightarrow$   
herdabilidade poligênica  
significativa  $\Rightarrow$  Identificar  
"genes" candidatos





# Mapeamento de Genes

## Delineamentos com Famílias e Estudos Observacionais

### Projeto Corações de Baependi, MG

Ajuste de Modelos de Componentes de Variância Poligênico (SOLAR)

Estimativas de Herdabilidades:  $\sigma_g^2 / (\sigma_e^2 + \sigma_g^2)$

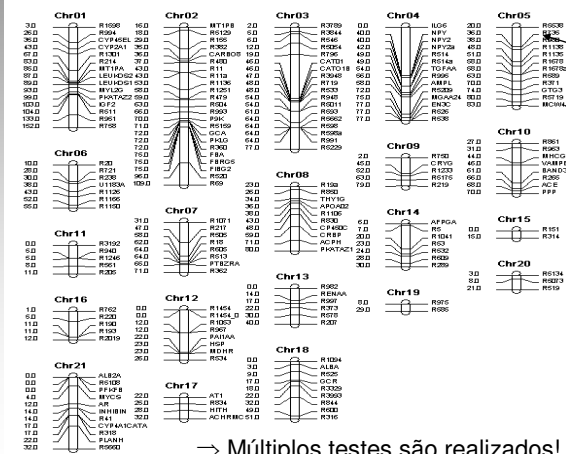
Fenótipo	Herdabilidade <sup>a</sup>	Prop. Cov	Covariáveis						n
			Sexo	Idade	Idade <sup>2</sup>	Sex*Idad	Ant-Hipert	Hipoglic	
<b>Var. Quantitativas</b>									
CIRCABD*	0.3924022	0.2128177		x	x	x	x	x	1666
CHDL*	0.3502449	0.0220878	x				x		1661
GLICOSE	0.2895079	0.1412034		x	x	x	x	x	1622
TRIG*	0.2798660	0.1157518		x	x		x		1655
mediaPAS*	0.2202316	0.3416865	x	x	x	x	x		1680
mediaPAD*	0.2099344	0.1938822	x	x	x	x	x		1678
<b>Fatores Principais</b>									
F1	0.2436401	0.2718471	x	x	x		x	x	1586
F2	0.3172038	0.1596869		x	x	x		x	1586
F3	0.3502158	0.0177955	x				x		1587
<b>Var. de Controle</b>									
Idade	0.0618386b	0.2777962	x				x	x	1685
Altura	0.8248354	0.5837355	x	x					1669

\*: transformação logarítmica (base 10) Significância: <sup>a</sup> (p<0.000) <sup>b</sup> (p = 0.010)

⇒ Próximo Passo: Coleta de Dados Genotípicos (Mapa de Marcadores) para encontrar regiões cromossômicas candidatas

### Mapeamento Genético Clássico

Mapa de Marcadores Moleculares ⇒ posição e genótipos conhecidos



Testar os “efeitos (aleatórios) genéticos” de todos os locos em um mapa de marcadores moleculares

Usar a informação do mapa para testar posições intermediárias

Identificar regiões cromossômicas candidatas!

⇒ Múltiplos testes são realizados!

## Modelo de Componentes de Variância Poligênico $\Rightarrow$ Oligogênico

$$\mathbf{Y}_f = \mathbf{X}_f \boldsymbol{\beta} + \sum_{k=1}^m \mathbf{Q}_{fk} + \mathbf{e}_f$$

$\mathbf{g}_f$  : poligene

$$\sigma_Y^2 = \sigma_g^2 + \sigma_e^2 = \sum_{k=1}^m \sigma_{Q_k}^2 + \sigma_e^2$$

poligene e resíduo



Isolar no modelo o efeito aleatório de um Loco (de “grande” efeito)

$$\mathbf{Y}_f = \mathbf{X}_f \boldsymbol{\beta} + \mathbf{Q}_{f1} + \sum_{k=2}^m \mathbf{Q}_{fk} + \mathbf{e}_f$$

$$\sigma_Y^2 = \sigma_{q1}^2 + \sum_{k=2}^m \sigma_{Q_{ki}}^2 + \sigma_e^2$$

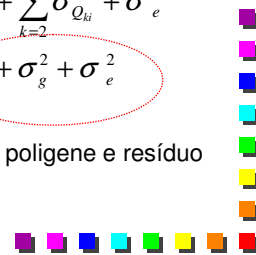
$$= \sigma_{q1}^2 + \sigma_g^2 + \sigma_e^2$$

QTL, poligene e resíduo

$$\text{Cov}(Y_{if}; Y_{i'f}) \approx \pi_{ii'} \sigma_{q1}^2 + 2\phi_{ii'} \sigma_g^2$$

↑  
Estimada como a proporção de alelos ibd no Loco 1

$\Rightarrow$  Dados de genótipo

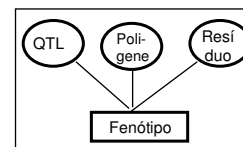


## Modelo de Componentes de Variância Oligogênico

$$\mathbf{Y}_f = \boldsymbol{\mu} + \mathbf{X}_f \boldsymbol{\beta} + \mathbf{q}_{1f} + \mathbf{g}_f + \mathbf{e}_f$$

efeitos aleatórios: QTL poligene ambiente

C.V.: QTL, Poligene e Resíduo



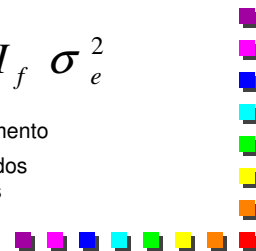
$$\text{Cov}(Y_{fi}; Y_{fi'}) = \begin{cases} \sigma_q^2 + \sigma_g^2 + \sigma_e^2 & i = i' \\ \pi_{ii'} \sigma_q^2 + 2\phi_{ii'} \sigma_g^2 & i \neq i' \end{cases}$$

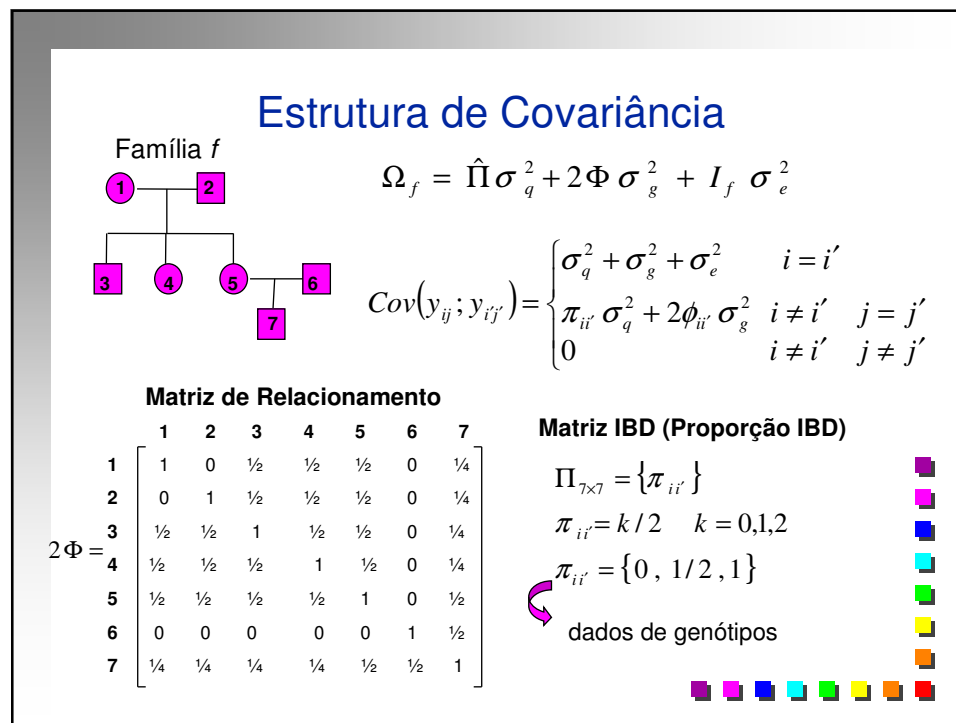
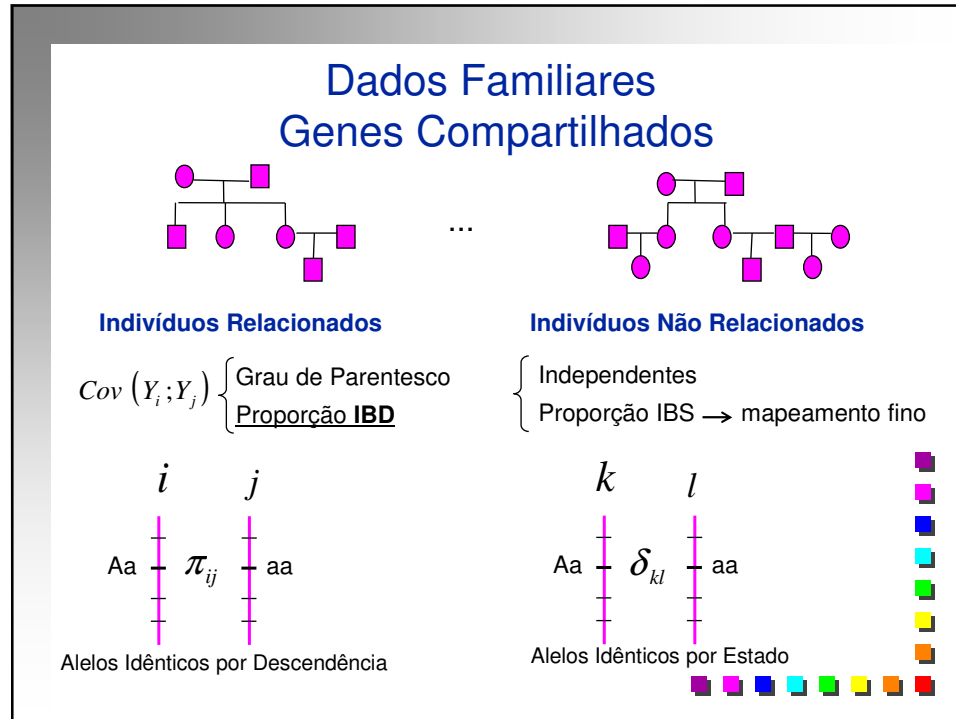


$$\boldsymbol{\Omega}_f = \hat{\Pi} \sigma_q^2 + 2\Phi \sigma_g^2 + \mathbf{I}_f \sigma_e^2$$

Matriz IBD  
 $\rightarrow \Pi$ : estimada dos dados do Mapa de Marcadores  
 $\rightarrow$  Para locos observados e não observados

Matriz de Relacionamento  
 $\rightarrow 2\Phi$ : estimada dos dados familiares





## Estratégia de Análise

Modelo	$\sigma_e^2$	$\sigma_g^2$	$\sigma_q^2$
Esporádico	$\sigma_e^2$	0	0
Poligênico	$\sigma_e^2$	$\sigma_g^2$	0
1 QTL	$\sigma_e^2$	$\sigma_g^2$	$\sigma_q^2$

### Teste do Modelo Poligênico

$$LOD_{Polig} = \log_{10} L(\hat{\sigma}_e^2, \hat{\sigma}_g^2 | Y) - \log_{10} L(\hat{\sigma}_e^2, \hat{\sigma}_g^2 = 0 | Y)$$

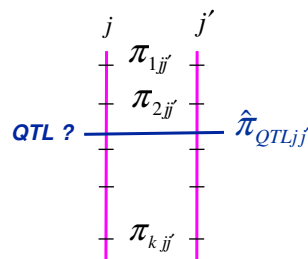
### Teste do Modelo Oligogênico (1 QTL de maior efeito)

$$LOD_{QTL} = \log_{10} L(\hat{\sigma}_e^2, \hat{\sigma}_g^2, \hat{\sigma}_q^2 | Y) - \log_{10} L(\hat{\sigma}_e^2, \hat{\sigma}_g^2, \hat{\sigma}_q^2 = 0 | Y)$$

$$\Lambda = 2 \left[ \ln L(\hat{\theta}) - \ln L(\theta_0) \right] \sim \text{Mistura } \frac{1}{2} \chi_0^2 + \frac{1}{2} \chi_1^2$$

## Predição da Proporção IBD

- Método via Regressão (Fulker et al., 1995; Almasy and Blangero, 1998)



$\pi$  - Autocorrelação : sinal de ligação M-QTL

$\pi$  - Variabilidade: herdabilidade do traço

Mapa de referência (Dados de Marcadores)



Fixar uma posição Candidata (QTL)

$$\hat{\pi}_{qij} = 2\phi_{ij} + \hat{\beta}_1 \pi_{M1} + \dots + \hat{\beta}_k \pi_{Mk}$$

$$\hat{\beta} = V(\pi_M)^{-1} \text{Cov}(\pi_q; \pi_M)$$

$$\text{Cov}(\pi_m, \pi_q | r_{jj'}; \theta_{mq}) \approx p^r(\theta)$$

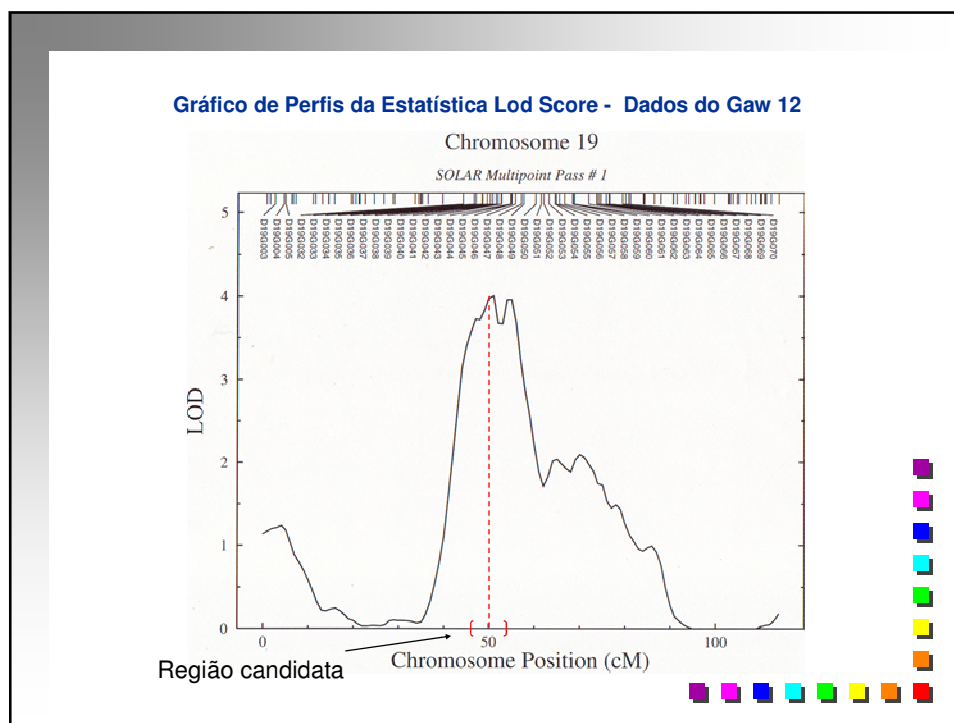


Algoritmos baseados em Simulações (MCMC)

→ dados missing de marcadores

# Mapeamento de Genes

## Delineamentos com Famílias e Estudos Observacionais



## Bibliografia

Almasy, L.; Blangero, J. (1998). Multipoint Quantitative-Trait Linkage Analysis in General Pedigrees. *Am. J. Hum. Genet.* **62**: 1198-1211.

Balding, D.J.; Bishop, M. and Cannings, C. (2003). **Handbook of Statistical Genetics**. 2th Ed. Wiley.

Blangero, J.; Williams, J.; Almasy, L. (2000). Quantitative Trait Locus Mapping Using Human Pedigrees. *Human Biology* **72**(1): 35-62

De Andrade, M; Amos, C. and Thiel, T. (1999). Methods to Estimate Genetic Components of variance for Quantitative traits in Family Studies. *Genetic Epidemiology* **17**:64-76.

Fulker, D.W.; Cherny, S.S.; Sham, P.C.; Hewitt, J.K. (1999). Combined Linkage and Association Sib-Pair Analysis for Quantitative Traits. *Am. J. Hum. Genet.* **64**: 259-267.

Leon, J.M. (2002). **Mapeamento Genético em Populações Humanas via Modelos de Componentes de Variância**. Dissertação de Mestrado. São Paulo: Universidade de São Paulo, Departamento de Estatística.

Lynch, M. and Walsh, B. (1998). **Genetics and Analysis of Quantitative Traits**. Sinauer Associates.