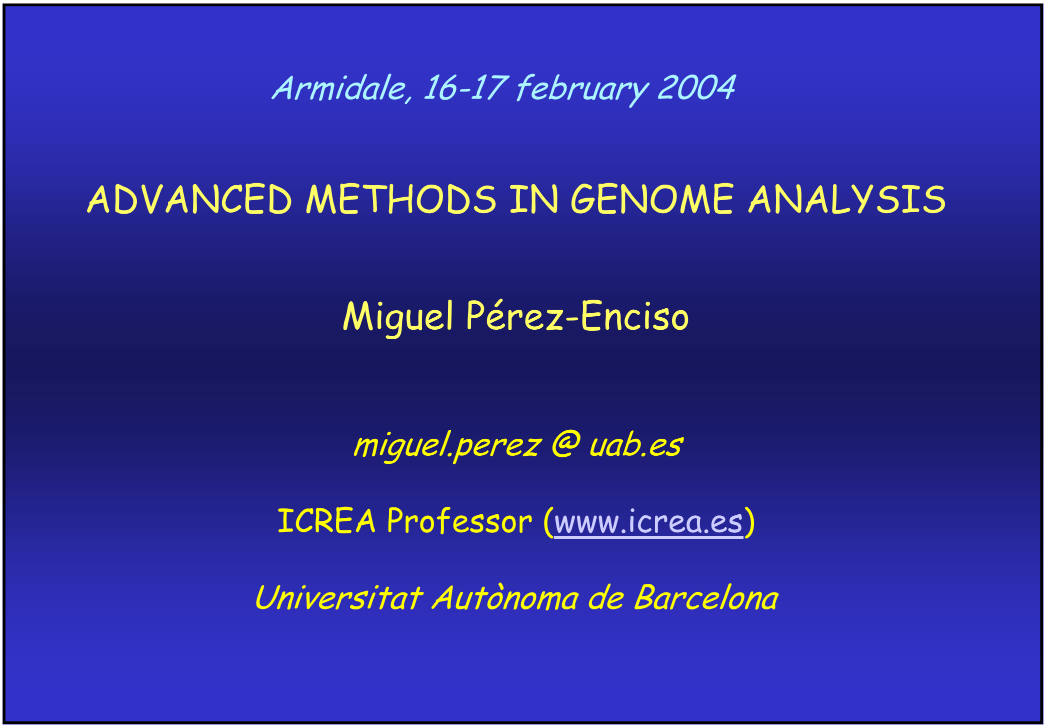


Armidale, 16-17 february 2004

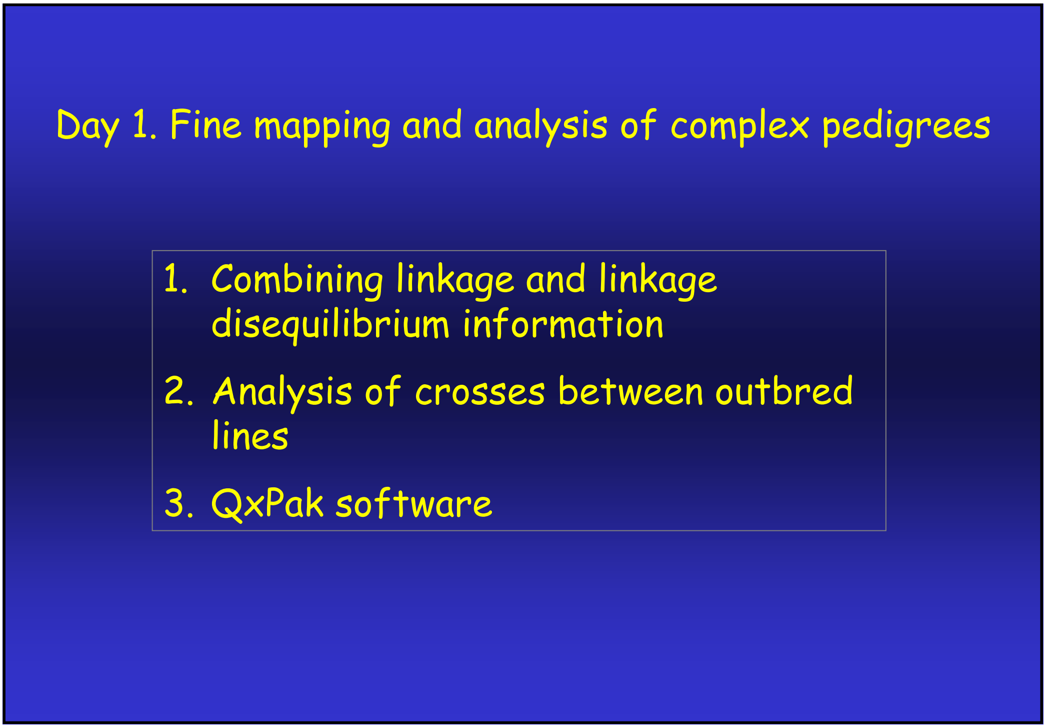
ADVANCED METHODS IN GENOME ANALYSIS

Miguel Pérez-Enciso

miguel.perez @ uab.es

ICREA Professor (www.icrea.es)

Universitat Autònoma de Barcelona



Day 1. Fine mapping and analysis of complex pedigrees

1. Combining linkage and linkage disequilibrium information
2. Analysis of crosses between outbred lines
3. QxPak software

Day 2. cDNA microarray analysis

1. Basic techniques

    Clustering

2. Prediction of phenotype given cDNA pattern

    Partial Least Squares

3. Genetical genomics

    Heat shock proteins (rats)

    Whole genome (yeast)

    Combining expression and markers for gene detection

Genetics has become a data rich science, where the limiting step already *NOW* is the data analysis, rather than in the obtention of the data themselves.

Recall the first QTL experiments, the rationale behind was to measure as many traits as possible because we wanted to maximize the output per marker typed.

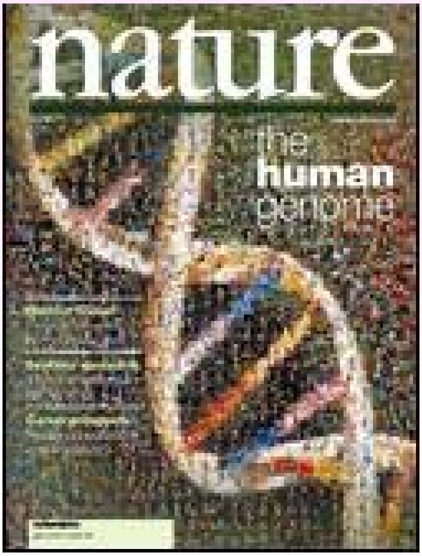
Three main streams of data:

DNA polymorphism (markers)

Expression data (functional genomics)

DNA Sequence

The culprit



What are quantitative traits?

Sensitive to the environment

Affected by several genes

↓

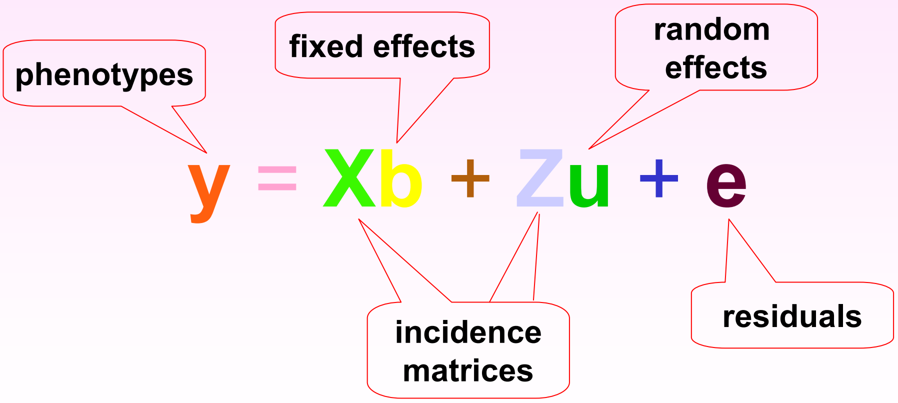
Traits showing a continuous distribution

The classical framework ...

is composed of two distinct parts:

1. The mixed model.
2. The infinitesimal genetic model (Fisher, 1918).

The mixed model



$$y = Xb + Zu + e$$

phenotypes

fixed effects

random effects

incidence matrices

residuals

From

the mixed model theory

+

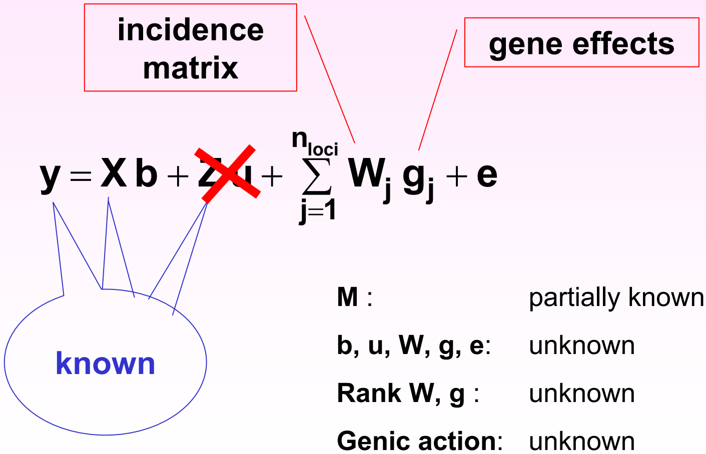
the infinitesimal genetic model

=

the mixed model equations (MME, Henderson, 1950)

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + \lambda G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

The problem:



incidence matrix

gene effects

$$y = X\,b + \cancel{Z\,u} + \sum_{j=1}^{n_{loci}} W_j\, g_j + e$$

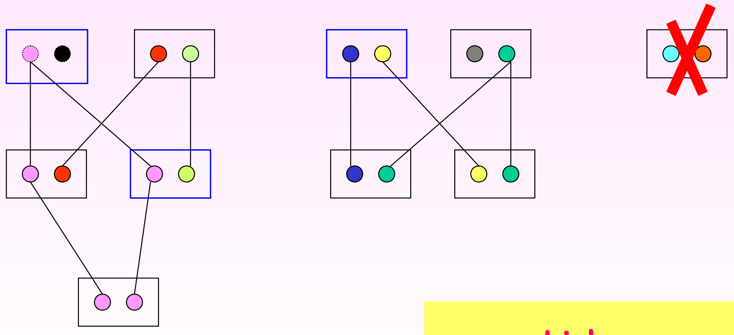
known

M :	partially known
b, u, W, g, e:	unknown
Rank W, g :	unknown
Genic action:	unknown

Day 1. Fine mapping and analysis of complex pedigrees

1. Combining linkage and linkage disequilibrium information
2. Analysis of crosses between outbred lines
3. QxPak software

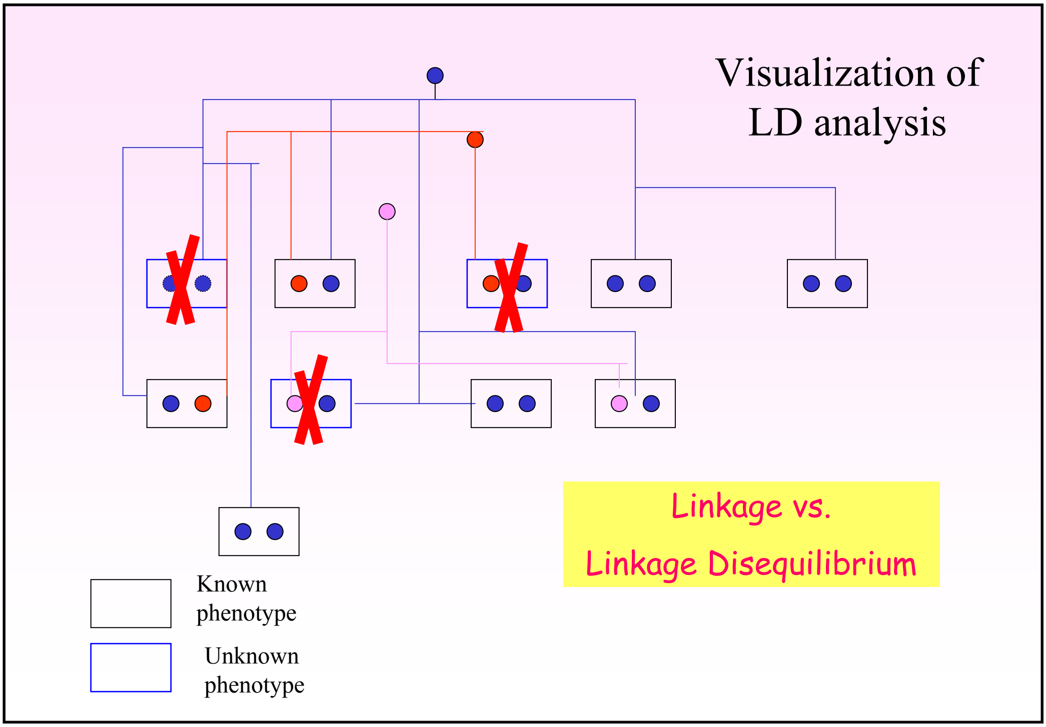
Visualization of linkage analysis



Linkage vs Linkage Disequilibrium

□ Known phenotype

□ Unknown phenotype



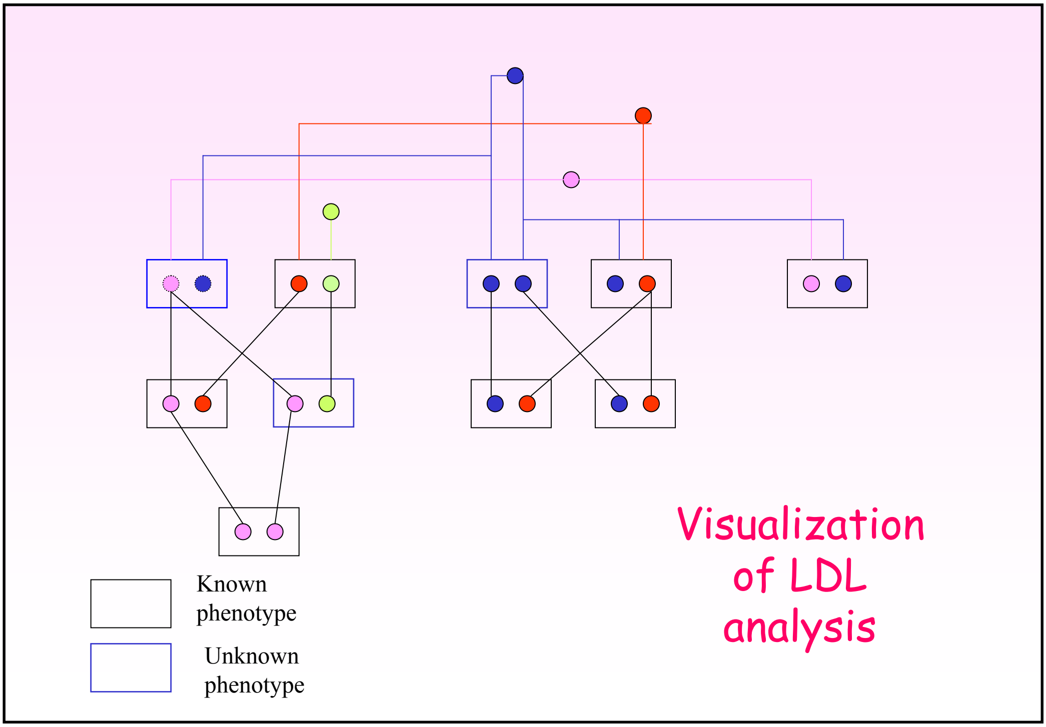
Visualization of LD analysis

Linkage vs.

Linkage Disequilibrium

□ Known phenotype

□ Unknown phenotype



Visualization of LDL analysis

□ Known phenotype

□ Unknown phenotype

Combining linkage and LD

Linkage analysis:

- Robust, but not very accurate.
- Assumes all alleles from base population are different.

Linkage disequilibrium:

- It can be very accurate but very sensitive to departures from model assumptions.
- Risk of false positives.
- The region of maximum LD may not coincide with the QTL position.
- A pure LD analysis disregards pedigree structure.

Potential advantages of LDL

Homozygous genotypes contribute information.

Non related individuals with phenotype also.

Offspring phenotypes contribute to assess the likely genotype of their parents, thus making it more robust than simply LD.

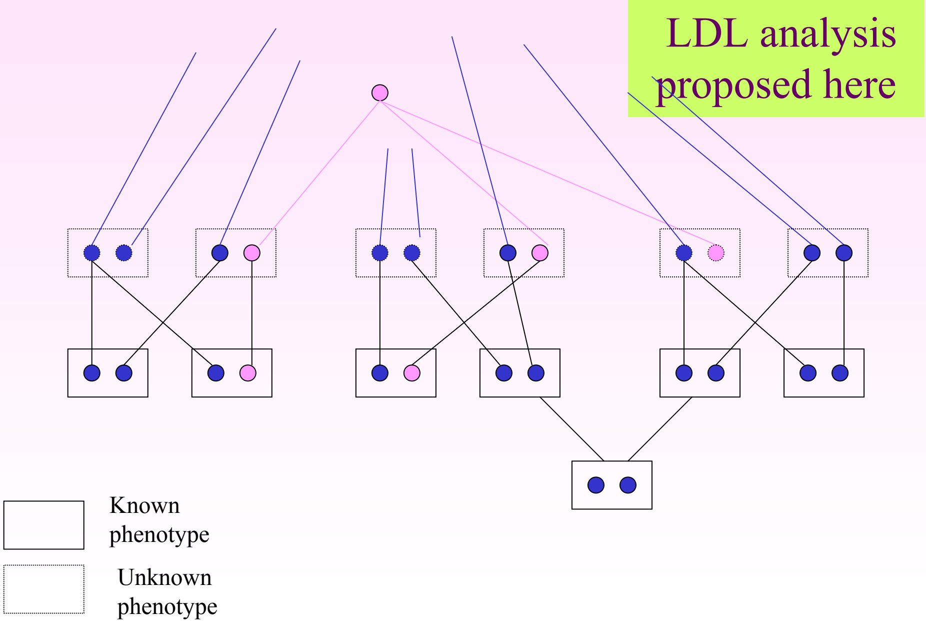
By comparing L, LD, and LDL estimates we can verify the assumptions in modelling LD decay.

LDL mapping

(Pérez-Enciso, Genetics, 2003)

Assumptions:

- QTL identified within a predetermined region.
- A biallelic QTL with a mutant allele appearing t generations ago on a single haplotype.
- A star shape (exponential growth) genealogy.

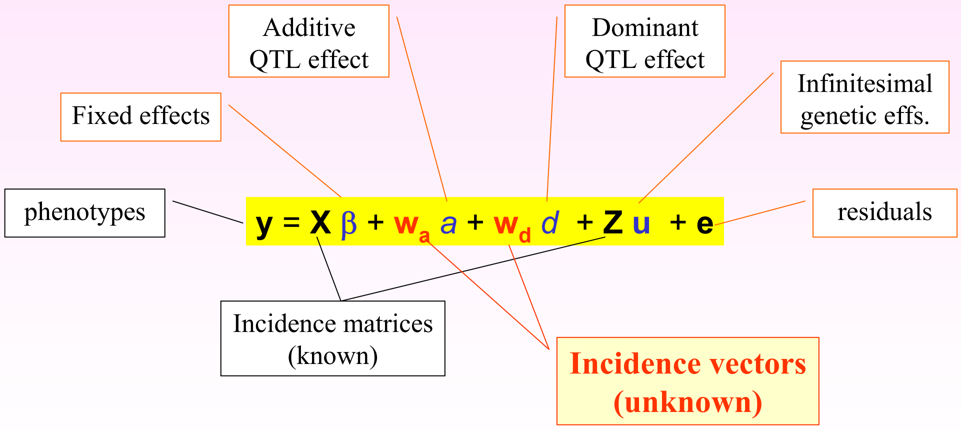


LDL analysis proposed here

Known phenotype

Unknown phenotype

The model



$$\mathbf{y} = \mathbf{X}\,\beta + \mathbf{w_a}\,a + \mathbf{w_d}\,d + \mathbf{Z}\,\mathbf{u} + \mathbf{e}$$

Additive QTL effect

Dominant QTL effect

Fixed effects

Infinitesimal genetic effs.

phenotypes

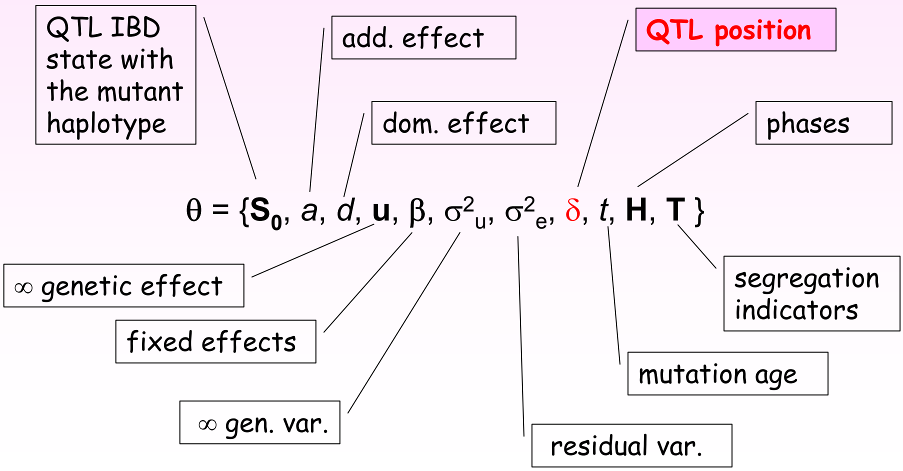
residuals

Incidence matrices (known)

Incidence vectors (unknown)

Parameters to be estimated

$$\mathbf{y} = \mathbf{X}\,\beta + \mathbf{w_a}\,a + \mathbf{w_d}\,d + \mathbf{Z}\,\mathbf{u} + \mathbf{e}$$



$$\theta = \{\mathbf{S_0}, a, d, \mathbf{u}, \beta, \sigma^2_u, \sigma^2_e, \delta, t, \mathbf{H}, \mathbf{T}\}$$

QTL IBD state with the mutant haplotype

add. effect

dom. effect

QTL position

phases

∞ genetic effect

fixed effects

∞ gen. var.

residual var.

mutation age

segregation indicators

Bayesian inference

$$\mathbf{y} = \mathbf{X}\,\beta + \mathbf{w_a}\,a + \mathbf{w_d}\,d + \mathbf{Z}\,\mathbf{u} + \mathbf{e}$$

Parameters to be estimated

$$\theta = \{\mathbf{S_0}, a, d, \mathbf{u}, \beta, \sigma^2_u, \sigma^2_e, \delta, t, \mathbf{H}, \mathbf{T}\}$$

Bayes posterior

$$p(\theta \mid \mathbf{y}, \mathbf{M}) \propto p(\mathbf{y}, \mathbf{M} \mid \theta)\, p(\theta) = p(\mathbf{y} \mid \theta)\, p(\mathbf{M} \mid \theta)\, p(\theta)$$

Marginal Bayes posterior

$$p(\theta_l \mid \mathbf{y}, \mathbf{M}) = \int_{\theta_{-l}} p(\theta_l, \boldsymbol{\theta}_{-l} \mid \mathbf{y}, \mathbf{M})\ \partial \boldsymbol{\theta}_{-l}$$

Bayesian inference principles

(joint) posterior

likelihood

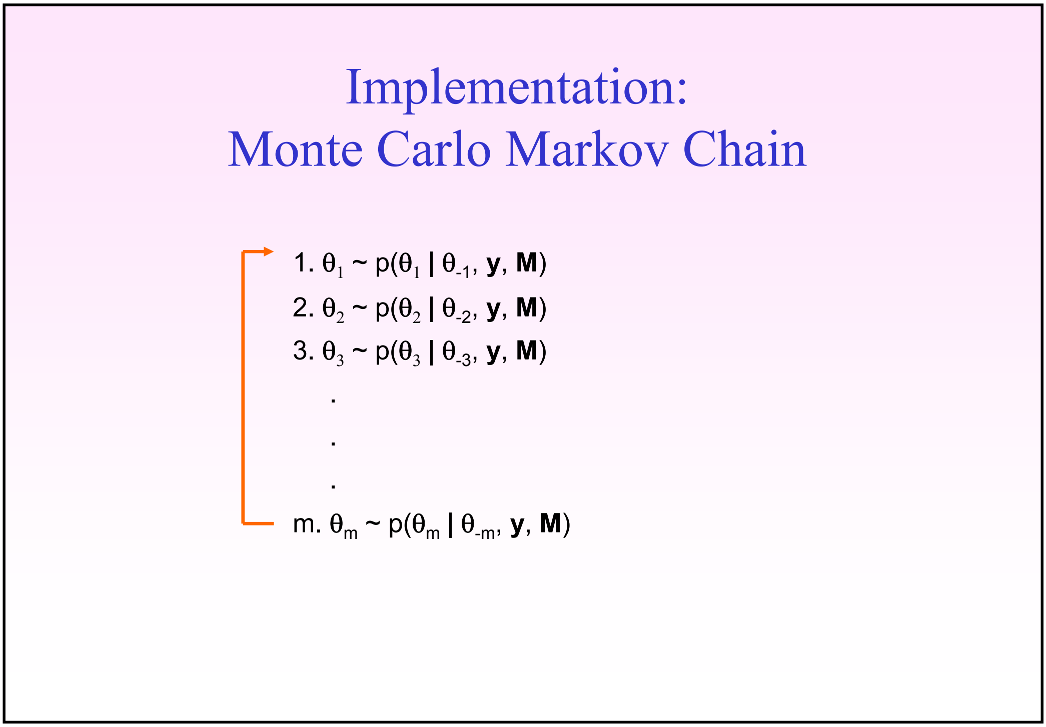
prior

$$p(\theta \mid y) = \frac{p(y \mid \theta)\, p(\theta)}{p(y)}$$

data density

marginal posterior

$$p(\theta_l \mid \mathbf{y}, \mathbf{M}) = \int_{\theta_{-l}} p(\theta_l, \boldsymbol{\theta}_{-l} \mid \mathbf{y}, \mathbf{M})\ \partial \boldsymbol{\theta}_{-l}$$



Implementation:
Monte Carlo Markov Chain

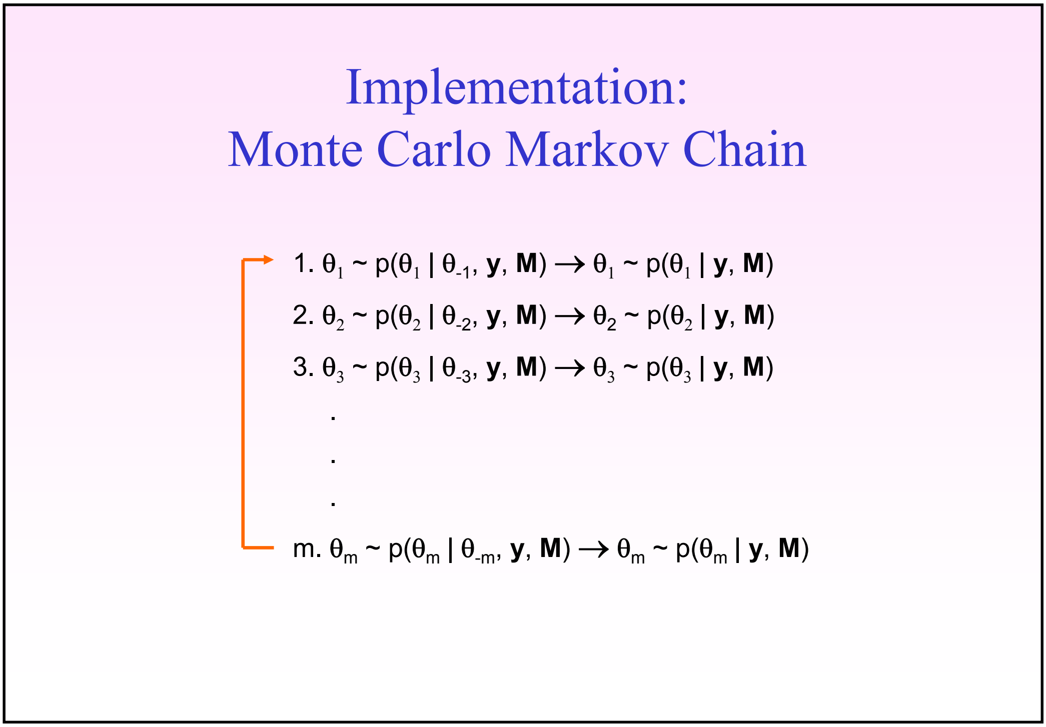
1. $\theta_1 \sim p(\theta_1 \mid \theta_{-1}, \mathbf{y}, \mathbf{M})$
2. $\theta_2 \sim p(\theta_2 \mid \theta_{-2}, \mathbf{y}, \mathbf{M})$
3. $\theta_3 \sim p(\theta_3 \mid \theta_{-3}, \mathbf{y}, \mathbf{M})$

.

.

.

m. $\theta_m \sim p(\theta_m \mid \theta_{-m}, \mathbf{y}, \mathbf{M})$



Implementation:
Monte Carlo Markov Chain

1. $\theta_1 \sim p(\theta_1 \mid \theta_{-1}, \mathbf{y}, \mathbf{M}) \rightarrow \theta_1 \sim p(\theta_1 \mid \mathbf{y}, \mathbf{M})$
2. $\theta_2 \sim p(\theta_2 \mid \theta_{-2}, \mathbf{y}, \mathbf{M}) \rightarrow \theta_2 \sim p(\theta_2 \mid \mathbf{y}, \mathbf{M})$
3. $\theta_3 \sim p(\theta_3 \mid \theta_{-3}, \mathbf{y}, \mathbf{M}) \rightarrow \theta_3 \sim p(\theta_3 \mid \mathbf{y}, \mathbf{M})$

.

.

.

m. $\theta_m \sim p(\theta_m \mid \theta_{-m}, \mathbf{y}, \mathbf{M}) \rightarrow \theta_m \sim p(\theta_m \mid \mathbf{y}, \mathbf{M})$

Priors assumed

Flat unbounded priors

a, d, β

Flat bounded priors

δ, t

Naive ignorance priors (hyperparameter $\nu = 0$)

σ^2_u, σ^2_e

Binomial priors

S_0 (prior QTL frequency), **H, T**

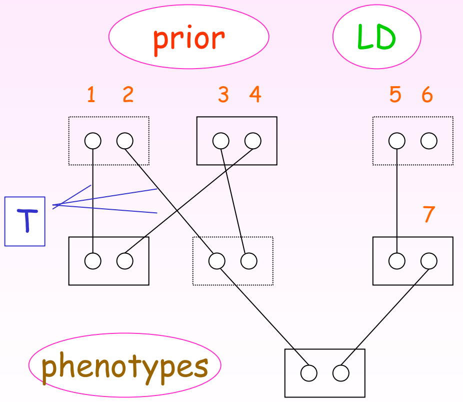
Transmission indicators

Phases

Multivariate normal

$p(\mathbf{u}) = N(\mathbf{0}, \mathbf{A}\, \sigma^2_u)$

Sampling QTL alleles
$(S_0 \mid a, d, \beta, \mathbf{u}, \mathbf{T}, \mathbf{H}, \mathbf{M}, \mathbf{y}, t, \sigma^2_e)$



$p(S_{0i} \mid \theta_-) \propto$

$p(S_0) \times$

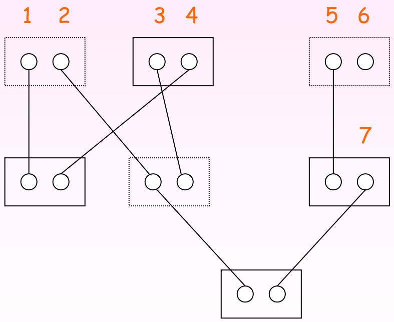
$p(\mathbf{M}_i \mid \theta_-) \times$

$\Pi_{j\in Ri}\, p(y_j \mid \theta_-)$

prior allele frequency

individuals having received origin i

Contribution from phenotypes
$(\mathbf{y} \mid S_0, a, d, \beta, \mathbf{u}, \sigma^2_e)$

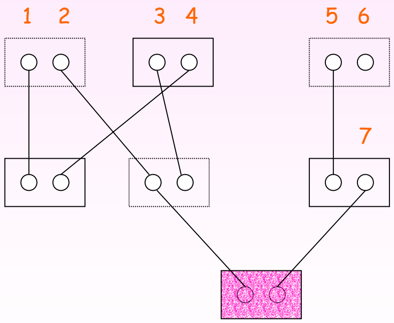


$\Pi_{j\in Ri}\, p(y_j \mid \theta_-)$

$p(y_i \mid S_{0j}, a, d, u_i, \beta, \sigma^2_e) =$

$N(y_j - \mathbf{x}_i' \beta - u_j - w_{aj}\, a - w_{dj}\, d\,,\, \sigma^2_e)$

Contribution from phenotypes
$(\mathbf{y} \mid S_0, a, d, \beta, \mathbf{u}, \sigma^2_e)$



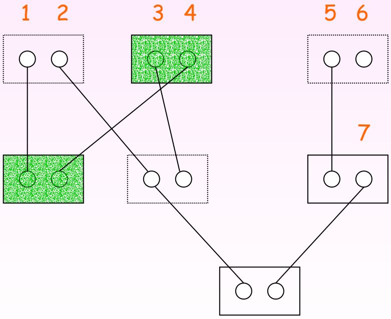
$\Pi_{j\in Ri}\, p(y_j \mid \theta_-)$

$p(y_i \mid S_{0j}, a, d, u_i, \beta, \sigma^2_e) =$

$N(y_j - \mathbf{x}_i' \beta - u_j - w_{aj}\, a - w_{dj}\, d\,,\, \sigma^2_e)$

Example: origin 2

Contribution from phenotypes
$(\mathbf{y} \mid S_0, a, d, \beta, \mathbf{u}, \sigma^2_e)$



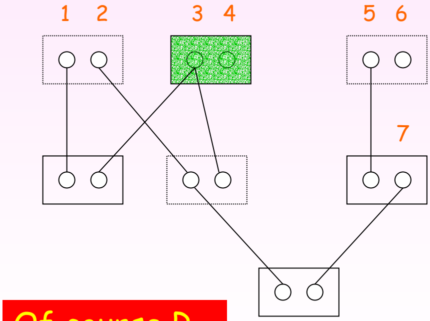
$\Pi_{j\in Ri}\; p(y_j \mid \theta_{-})$

$p(y_i \mid S_{0j}, a, d, u_i, \beta, \sigma^2_e) =$

$N(y_j - \mathbf{x}_i'\, \beta - u_j - w_{aj}\, a - w_{dj}\, d\, , \sigma^2_e)$

Example: origin 4

Contribution from phenotypes
$(\mathbf{y} \mid S_0, a, d, \beta, \mathbf{u}, \sigma^2_e)$



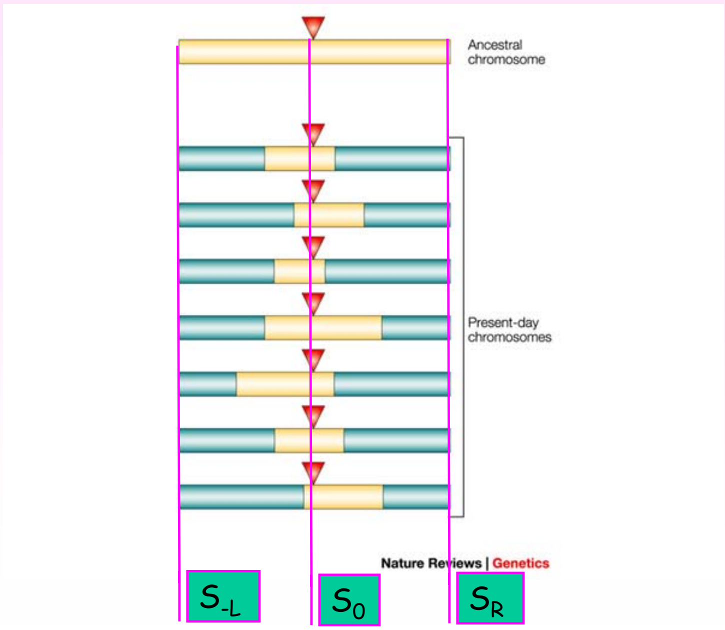
$\Pi_{j\in Ri}\; p(y_j \mid \theta_{-})$

$p(y_i \mid S_{0j}, a, d, u_i, \beta, \sigma^2_e) =$

$N(y_j - \mathbf{x}_i'\, \beta - u_j - w_{aj}\, a - w_{dj}\, d\, , \sigma^2_e)$

Of course R_i depends on T!

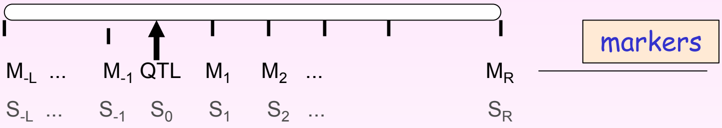
Example: origin 4



Nature Reviews | Genetics

Contribution from disequilibrium: $p(M \mid \theta)$

Haplotype i



S_k is IBD indicator for position k
S_k =1 ≡ kth marker allele is IBD with initial mutant haplotype
S_k =0 ≡ kth marker allele not IBD with initial mutant haplotype
k = 0 ≡ QTL position

based in Morris et al. (2000)

Contribution from disequilibrium: $p(M \mid \theta)$

$p(\mathbf{M}_i \mid S_{0i}, t, \mathbf{H}, \delta) =$

$= p(M_{i-L}, \ldots, M_{i-1}, M_{i1}, M_{i2}, \ldots M_{iR} \mid \theta) =$

$= p(M_{i-L}, \ldots, M_{i-1} \mid \theta)\ p(M_{i1}, M_{i2}, \ldots M_{iR} \mid \theta) = Q_{iL}\ Q_{iR}$

'left' part

'right' part

$Q_{iR} = p(M_1, M_2, \ldots M_R \mid \theta) =$

$= \Sigma_{S1=0,1}\ p(M_1, M_2, \ldots M_R \mid S_1, S_0, \theta\_) =$

$= \Sigma_{S1=0,1}\ p(M_2, \ldots M_R \mid S_1, \theta\_)\ p(M_1 \mid S_1, \theta\_)\ p(S_1 \mid S_0, \theta\_)$

remaining haplotype

prob. allele | IBD state

transition probabilities

Contribution from disequilibrium: $p(M \mid \theta)$

$p(\mathbf{M}_i \mid S_{0i}, t, \mathbf{H}, \delta) =$

$= p(M_{i-L}, \ldots, M_{i-1}, M_{i1}, M_{i2}, \ldots M_{iR} \mid \theta) =$

$= p(M_{i-L}, \ldots, M_{i-1} \mid \theta)\ p(M_{i1}, M_{i2}, \ldots M_{iR} \mid \theta) = Q_{iL}\ Q_{iR}$

'left' part

'right' part

$Q_R = p(M_1, M_2, \ldots, M_R \mid S_0, \ldots) =$

$= \sum_{S_1} p(M_2, \ldots, M_R \mid S_1)\ p(M_1 \mid S_1)\ p(S_1 \mid S_0) =$

$= \sum_{S_1} \sum_{S_2}\ p(M_3, \ldots, M_R \mid S_2)\ \underline{p(M_2 \mid S_2)\ p(S_2 \mid S_1)}\ \underline{p(M_1 \mid S_1)\ p(S_1 \mid S_0)} =$

$= \sum_{S_1} \sum_{S_2} \cdots \sum_{S_R} \prod_{k=1}^{R} p(M_k \mid S_k)\ p(S_k \mid S_{k-1})$

Sk are integrated out

Contribution from disequilibrium: p(M | θ)

Marker allele probs | IBD state

$$Q_R = p(M_1, M_2, \ldots M_R \,|\, \boldsymbol{\theta}) = \sum_{S_1} \sum_{S_2} \ldots \sum_{S_R} \prod_{k=1}^{R} p(M_k \,|\, S_k)\, p(S_k \,|\, S_{k-1})$$

transition IBD prob.

$p(M_k \mid S_k = +)$ is simply the population allele frequencies
$p(M_k \mid S_k = -) = 1$ if that allele was carried by the mutant haplotype
$p(M_k \mid S_k = -) = 0$ for the remaining alleles

t, time since mutation

r, recombination rate btw markers

$p(S_k = - \mid S_{k-1} = -) = \exp(- \, t\, r_{k,k+1}) + [1 - \exp(- \, t\, r_{k,k+1})]\, \alpha$

$p(S_k = - \mid S_{k-1} = +) = [1 - \exp(- \, t\, r_{k,k+1})]\, \alpha$

$p(S_k = + \mid S_{k-1} = -) = [1 - \exp(- \, t\, r_{k,k+1})]\, (1-\alpha)$

$p(S_k = + \mid S_{k-1} = +) = \exp(- \, t\, r_{k,k+1}) + [1 - \exp(- \, t\, r_{k,k+1})]\, (1 - \alpha)$

α, prob of recombining with a haplotype carrying the mutation

Morris et al. (2000)

Rearranging...

$$Q_R = p(M_1, M_2, \ldots M_R \,|\, \boldsymbol{\theta}) = \sum_{S_1} \sum_{S_2} \ldots \sum_{S_R} \prod_{k=1}^{R} p(M_k \,|\, S_k)\, p(S_k \,|\, S_{k-1})$$

$$= \sum_{S_1} p(M_1 | S_1)\, p(S_1 | S_0) \ldots \sum_{S_{R-1}} p(M_{R-1} | S_{R-1})\, p(S_{R-1} | S_{R-2}) \sum_{S_R} p(M_R | S_R)\, p(S_R | S_{R-1})$$

Then:

$$Q_R = \sum_{k=R}^{1} q_k$$

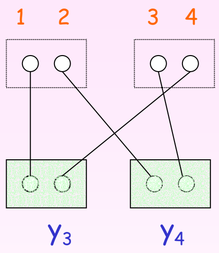
$$Q_L = \sum_{k=-L}^{1} q_k$$

Where

$q_k = \Sigma_{Sk}\, p(M_k \mid S_k)\, p(S_k \mid S_{k-1})\, q_{k-1}$

$q_R = q_{-L} = 1$

Block sampling QTL alleles



$$p(S_{01,}\, S_{02} \mid \theta_{-}) \propto$$

$$p(S_0) \text{ x}$$

$$p(\mathbf{M}_i \mid \theta_{-}) \text{ x}$$

$$\Pi_{j\in Ri}\, p(y_j \mid \theta_{-})$$

S_{01}	S_{02}	$p(S_0)$	$p(M\mid\theta)$	$p(y\mid\theta)$
+	+	$\propto p(S_0{=}+)^2$	$p(M_1\mid S_0{=}+)p(M_2\mid S_0{=}+)$	$p(y_3\mid S_0{=}+)p(y_4\mid S_0{=}+)$
+	-	$\propto p(S_0{=}+)p(S_0{=}-)$	$p(M_1\mid S_0{=}+)p(M_2\mid S_0{=}-)$	$p(y_3\mid S_0{=}+)p(y_4\mid S_0{=}-)$
-	+	$\propto p(S_0{=}+)p(S_0{=}-)$	$p(M_1\mid S_0{=}-)p(M_2\mid S_0{=}+)$	$p(y_3\mid S_0{=}-)p(y_4\mid S_0{=}+)$
-	-	$\propto p(S_0{=}-)^2$	$p(M_1\mid S_0{=}-)p(M_2\mid S_0{=}-)$	$p(y_3\mid S_0{=}-)p(y_4\mid S_0{=}-)$

Sampling mixed model effects (a, d, β, **u**)

Conditional on $\mathbf{w}_a$, $\mathbf{w}_d$, and variances

$\boldsymbol{\beta}^* = (\boldsymbol{\beta}', a, d)'$

$\mathbf{X}^* = (\mathbf{X}, \mathbf{w}_a, \mathbf{w}_d)$

$$\begin{bmatrix} \mathbf{X^{*\prime} X^*} & \mathbf{X^{*\prime} Z} \\ \mathbf{Z' X^*} & \mathbf{Z' Z} + \mathbf{A}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^* \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X^{*\prime} y} \\ \mathbf{Z y} \end{bmatrix} \equiv \mathbf{C\, b} = \mathbf{r}$$

$\lambda = \sigma_e^? / \sigma_u^?$

$$(b_i \mid \boldsymbol{\theta}_{-}, \mathbf{y}, \mathbf{M}) \sim N(r_i - \sum_{j=1, j\neq i}^{\text{rank(C)}} c_{ij}\, r_j\,,\, \sigma_e^2/c_{ii})\,, \quad \forall i$$

← Normal f(x)

any of β, **u**, a, or d parameters

Sampling variances

$$p(\sigma^2_u \mid S_0, a, d, \mathbf{u}, \beta, \sigma^2_e, \mathbf{y}) = (\mathbf{u}' \mathbf{A}^{-1} \mathbf{u})\ \chi_m^{-2}$$

Chi squared

$$p(\sigma^2_e \mid S_0, a, d, \mathbf{u}, \beta, \sigma^2_u, \mathbf{y}) =$$

$$= (\mathbf{y} - \mathbf{X}^* \beta^* - \mathbf{Z}\,\mathbf{u})'\ (\mathbf{y} - \mathbf{X}^* \beta^* - \mathbf{Z}\,\mathbf{u})\chi_n^{-2}$$

LD parameters: t

Metropolis Hastings

The new proposed t_{new} is accepted with probability:

$$\min\ \{1, \frac{p(\mathbf{M} \mid \mathbf{S_0}, t^{new}, \mathbf{H},\ \delta)}{p(\mathbf{M} \mid \mathbf{S_0}, t, \mathbf{H},\ \delta)}\}$$

LD parameters: p(M|S)

Marker non IBD with founder mutant haplotype

Marker IBD status

= 1 if i indiv, h hap,has k allele at j marker ; 0 otherwise

$$p(M_{kj}\,|\,S_k{=}+,\boldsymbol{\theta}_{\text{-}}) = \sum_{i=1}^{N_b}\sum_{h=1}^{2} p(S_{kih} = +\,|\,S_{0ih}\,)\,\eta_{ihjk}\ /\ (2\,N_b)$$

Marker k, allele j

ith indiv among base popn.

h-th haplotype

# indivs. in base popn.

Marker IBD with founder mutant haplotype

$$p(M_{kj}\,|\,S_k{=}-,\boldsymbol{\theta}_{\text{-}}) = \sum_{i=1}^{N_b}\sum_{h=1}^{2} p(S_{kih} = -\,|\,S_{0ih}\,)\,\eta_{ihjk}\ /\ (2\,N_b)$$

Phases, H

are sampled in blocks of n_h phases of the same individual jointly via Gibbs sampling (see above this chapter).

Segregation indicators, T

are sampled jointly with QTL position following Mendelian rules and using available marker information.

QTL position { δ }

Sampling δ is, together with updating QTL alleles, the most critical aspect of QTL Bayesian implementation. This occurs because $\mathbf{S_0}$, **T**, **H** and δ are highly interdependent and it is difficult t update them all simultaneously.

Conditional on $\mathbf{S}_0$ and **T**, updating δ is a straighforward, and it is like a standard linkage analysis.

But this simple approach is very prone to get δ stuck within a marker bracket because, conditional on a given set of crossovers, it is very unlikely to 'jump' to the next marker bracket.

A mixed approach was followed here (Uimari and Sillanpaa, 2001)

A new δ is accepted with prob.

$$\text{Min}\,\{1, \frac{p(\mathbf{T}\,|\,\delta^{new},\ \mathbf{H})}{p(\mathbf{T}\,|\,\delta,\ \mathbf{H})}\}$$

This ratio depends on the xovers that have occurred when the QTL is assumed to be in position δ or in δ^{new}

or, every ñ iterations, with prob.

$$\text{Min}\,\{1, \frac{p(\mathbf{y}\,|\ \mathbf{T}^{new}, \mathbf{S}_0, a, d, \mathbf{u}, \beta,\ \sigma_e^2)}{p(\mathbf{y}\,|\,\mathbf{T}, \mathbf{S}_0, a, d, \mathbf{u}, \beta,\ \sigma_e^2)}\}$$

This ratio is computed after sampling a new **T** set conditional on δ^{new}, the ratio depends on how fit new genotypes to observed phenotypes

Example: Simulation study

'Simple' pedigree (n = 480):

40 fullsib families; 10 offspring / family

'Complex' pedigree (n = 480)

4 generation pedigree; 80 parents; 5 fullsibs / family

Region explored = 25 cM

6 microsatellites and 11 SNPs

Additive effect = 1

Dominant eff = 0

Residual var = 1

Complete association

All haps with SNP_{18} = 2 had mutant QTL allele

Incomplete association

Initially, 42% of had SNP_{18} = 1 had mutant QTL allele

In all cases, star shape genealogy

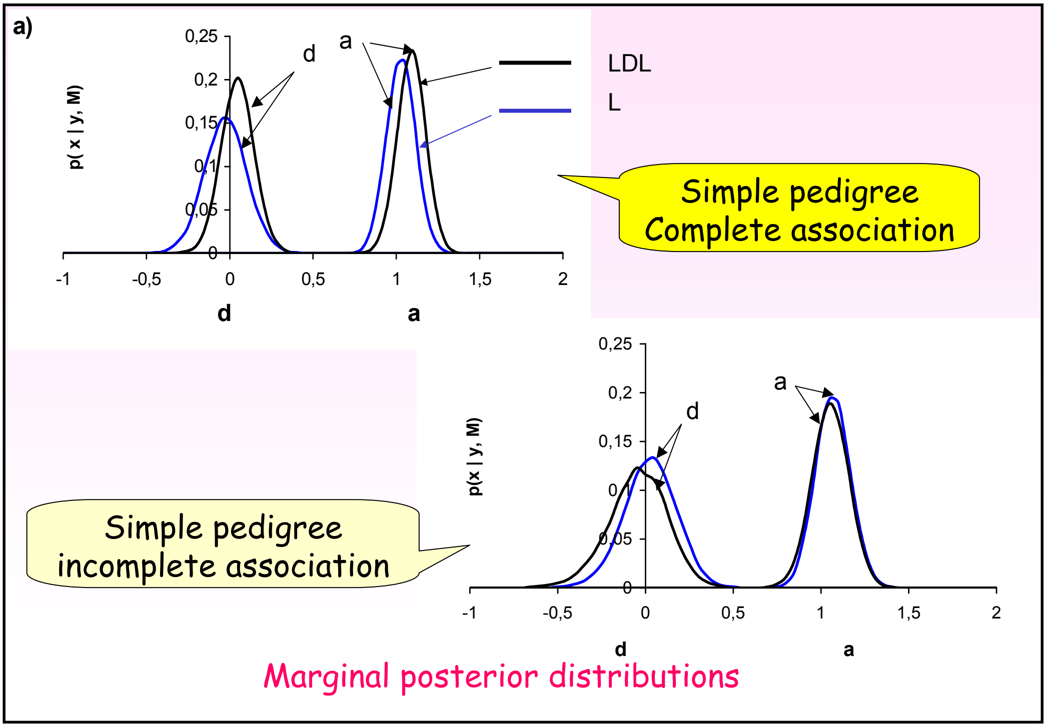
Main results

Complete association

Popn.	Method	$E(a\|y)$	$E(d\|y)$	$E(\delta\|y)$	$Var(\delta\|y)^{0.5}$
Simple	LDL	1.07	0.06	0.177	0.024
	L	1.04	0.03	0.161	0.045
Complex	LDL	0.96	0.00	0.179	0.020
	L	0.91	0.02	0.175	0.035
Simple	LDL	1.03	-0.08	0.170	0.041
	L	1.04	-0.01	0.144	0.060
Complex	LDL	0.88	0.06	0.185	0.026
	L	0.89	0.10	0.180	0.032
True		**1.00**	**0.00**	**0.18**	

Incomplete assoc.

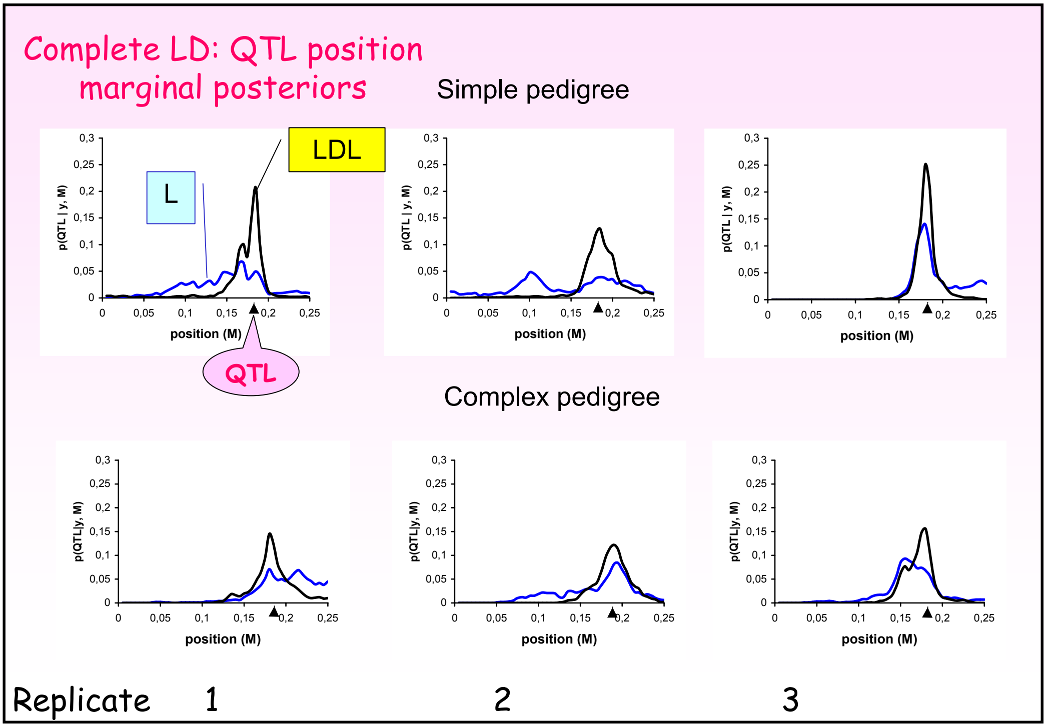
Average of 3 replicates



Simple pedigree
Complete association

Simple pedigree
incomplete association

Marginal posterior distributions



Complete LD: QTL position
marginal posteriors

Simple pedigree

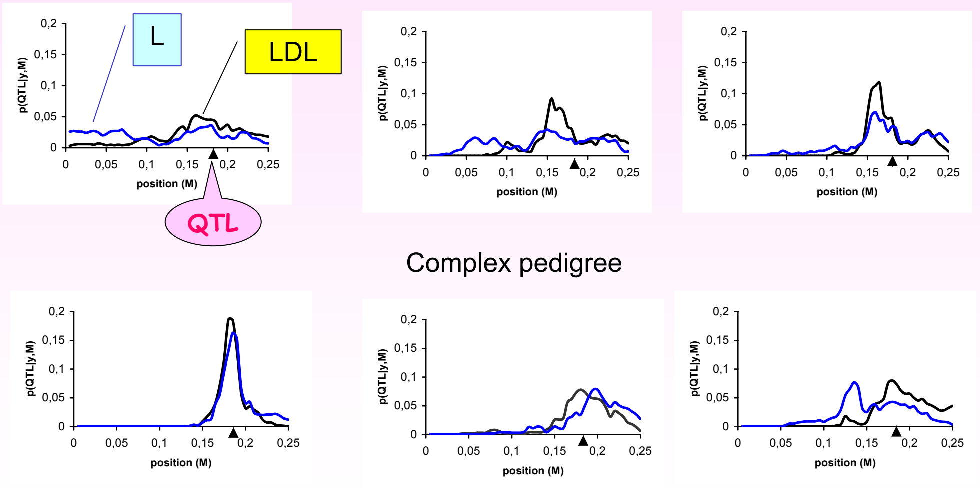
Complex pedigree

Replicate 1 2 3

Incomplete LD: QTL position marginal posteriors

Simple pedigree

Complex pedigree

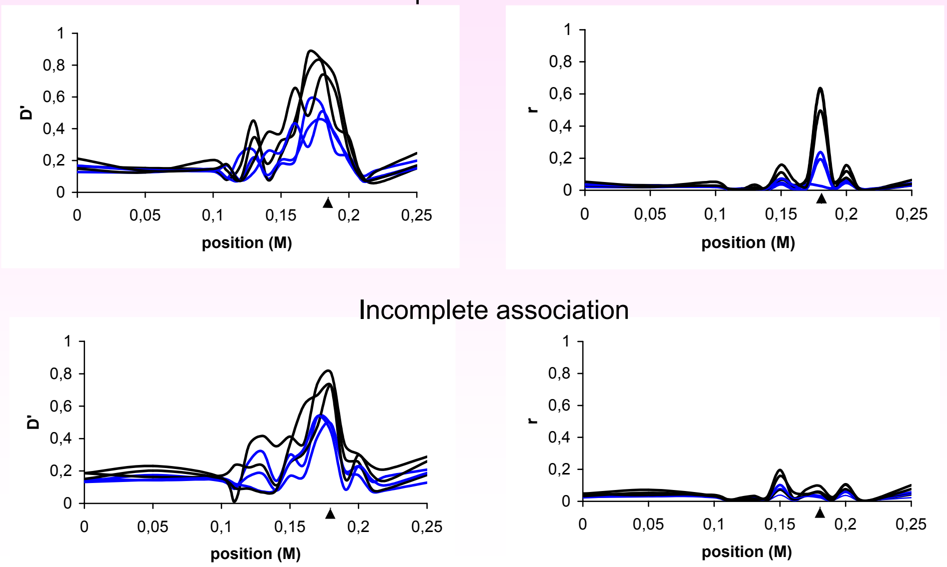


Replicate 1 2 3

Disequilibrium measures: simple pedigree

Complete association

Incomplete association



Conclusions

The advantage of LDL over linkage only will depend on the structure of the population as well as on the validity of the LD model.

Uncertainty on phases and on QTL alleles makes it LDL to perform much poorer than expected.

It seems that LDL increase in accuracy should not be overestimated.

A very exciting and timely area of research, many open fronts and approaches.

Other approaches

Allison, D. B., Heo, M., Kaplan, N., & Martin, E. R. (1999). Sibling-based tests of linkage and association for quantitative traits. *Am J Hum Genet* **64,** 1754-1763.

Farnir, F., Grisart, B., Coppieters, W., Riquet, J., Berzi, P., Cambisano, N., Karim, L., Mni, M., Moisio, S., Simon, P., Wagenaar, D., Vilkki, J., & Georges, M. (2002). Simultaneous Mining of Linkage and Linkage Disequilibrium to Fine Map Quantitative Trait Loci in Outbred Half-Sib Pedigrees. *Genetics* **161,** 275-287.

➡ Meuwissen, T. H., Karlsen, A., Lien, S., Olsaker, I., & Goddard, M. E. (2002). Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* **161,** 373-379.

Meuwissen & Goddard’s approach

The goal is to compute the probabilities that two haplotypes are identical by descent (IBD) at a given position or segment

For any given position $G=\{g_{ij}\}$ contains these P_{IBD}

G is later used in a maximum likelihood approach

Meuwissen & Goddard’s approach

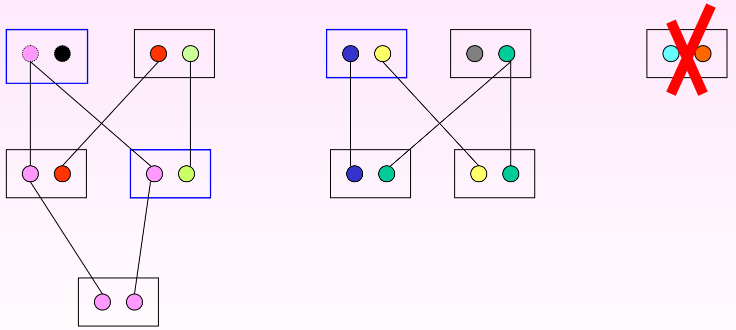
In a usual analysis, base population individuals are assumed to be unrelated.

But if we use LD, this is no longer true.

M&G present a method to compute the relationship (IBD probs) between base population individuals.

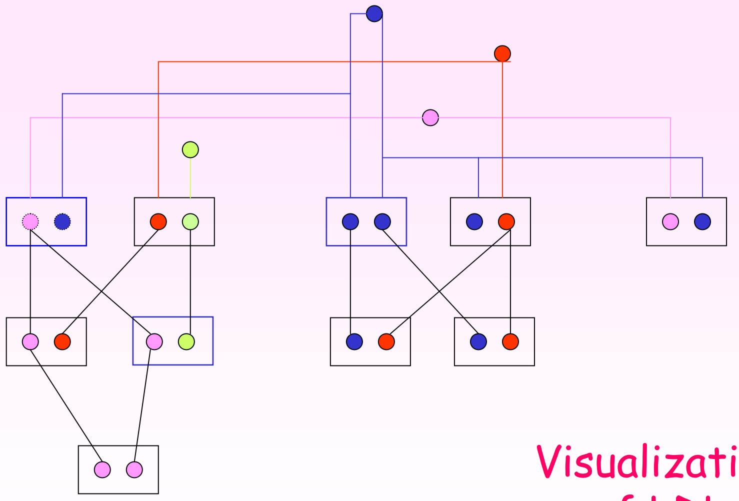
The usual relationship between descendants is increased according to this base IBD probs.

Visualization of linkage analysis



□ Known phenotype

□ Unknown phenotype

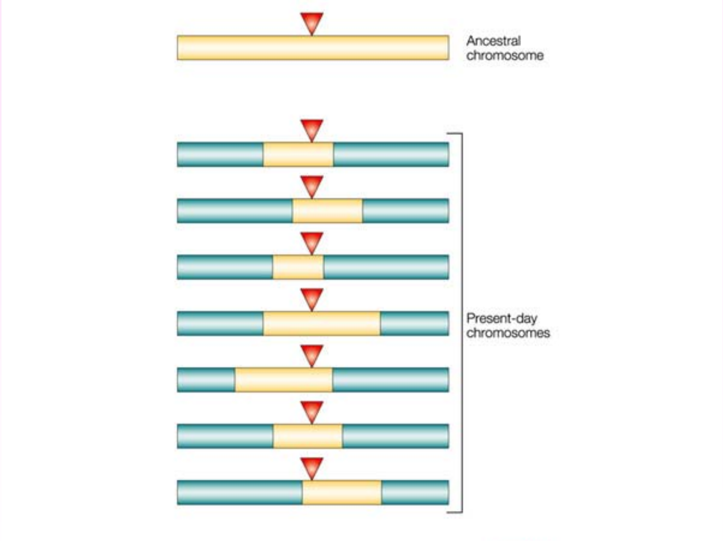


□ Known phenotype

□ Unknown phenotype

Visualization of LDL analysis

Haplotype erosion



Nature Reviews | Genetics

Ardlie et al. 2002

Prob. of two individuals sharing an intact chromosome segment

Depends on:

c ≡ chr. length (↓)

t ≡ time in generations since most recent common ancestor (MRCA) (↓)

Ne ≡ effective size (↓)

Prob. of two individuals sharing an intact chromosome segment

P MRCA in gen. t

P no common ancestor gens t-1

P no recombination in 2t gens.

$$\frac{1}{2Ne}\left(1-\frac{1}{2\mathrm{Ne}}\right)^{t-1}\left[\exp(\text{-c})\right]^{2t}$$

$$\approx \frac{1}{2Ne}\exp\left[-\frac{\mathrm{t}-1}{2\mathrm{Ne}}-2\mathrm{ct}\right]$$

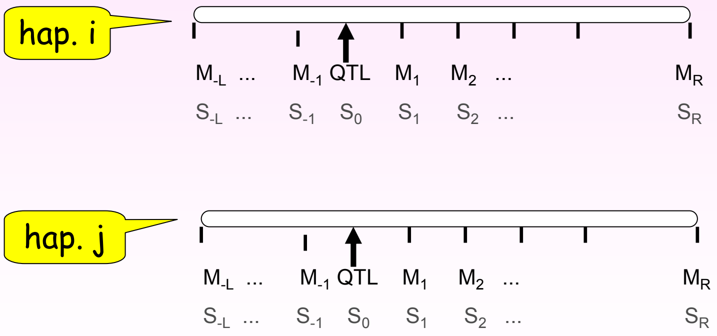
Prob. of two individuals sharing an intact chromosome segment of length c

reference time

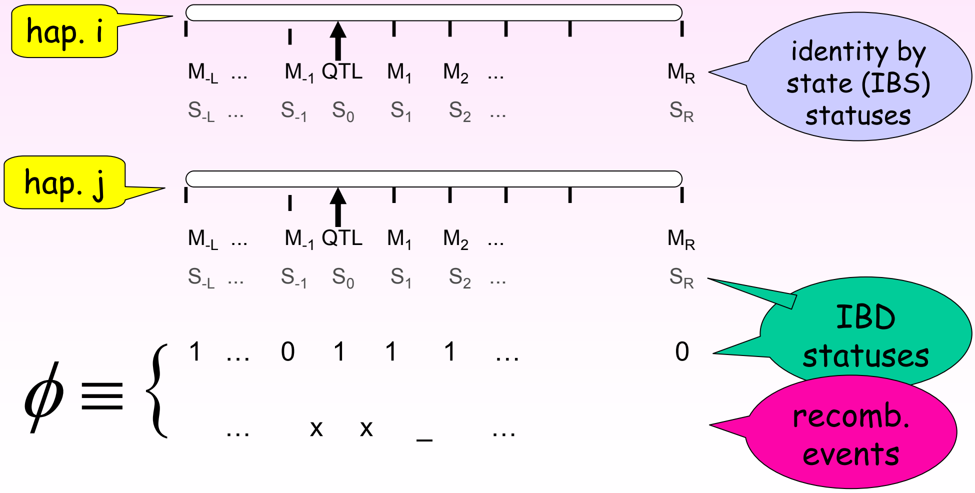
$$Pc=\frac{\exp(-2c)}{2Ne}\sum_{\mathrm{t}=1}^{\mathrm{T}}\exp\left(-(\mathrm{t}-1)\left(\frac{1}{2\mathrm{Ne}}+2c\right)\right)=$$

$$\frac{\exp(-2c)}{2Ne}\;\frac{1-\exp\left[-\mathrm{T}(2\mathrm{c}+\frac{1}{2\mathrm{Ne}})\right]}{1-\exp\left[-(2\mathrm{c}+\frac{1}{2\mathrm{Ne}})\right]}$$

Prob. of two individuals being IBD


hap. i
M-L ... M-1 QTL M1 M2 ... MR
S-L ... S-1 S0 S1 S2 ... SR
hap. j
M-L ... M-1 QTL M1 M2 ... MR
S-L ... S-1 S0 S1 S2 ... SR

Prob. of two individuals being IBD


hap. i
M-L ... M-1 QTL M1 M2 ... MR
S-L ... S-1 S0 S1 S2 ... SR
identity by state (IBS) statuses
hap. j
M-L ... M-1 QTL M1 M2 ... MR
S-L ... S-1 S0 S1 S2 ... SR
IBD statuses
φ ≡ { 1 ... 0 1 1 1 ... 0
... x x _ ...
recomb. events

M&G’s original idea

S_0 S_1	M_1	p(M|S)	P(S)
0x0	0	$1-p^2$	
1x0	0	$1-p^2$	
0x1	1	1	
0x0	1	p^2	
1_1	1	1	P_c
1x1	1	1	
1x0	1	p^2	

see eqn. before

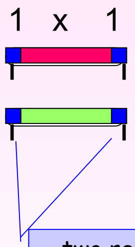
Goal:

$$P(S_0\,|\,M_1) = \frac{P(S_0, M_1)}{P(M_1)} = \frac{\sum_{S_1} P(M_1\,|\,S_{0,1})\,P(S_{0,1})}{P(M_1)}$$

freq. 2 allele marker 1

M&G’s original idea

S_0 S_1	M_1	p(M|S)	P(S)
0x0	0	$1-p^2$	
1x0	0	$1-p^2$	
0x1	1	1	
0x0	1	p^2	
1_1	1	1	P_c
1x1	**1**	**1**	**?**
1x0	1	p^2	

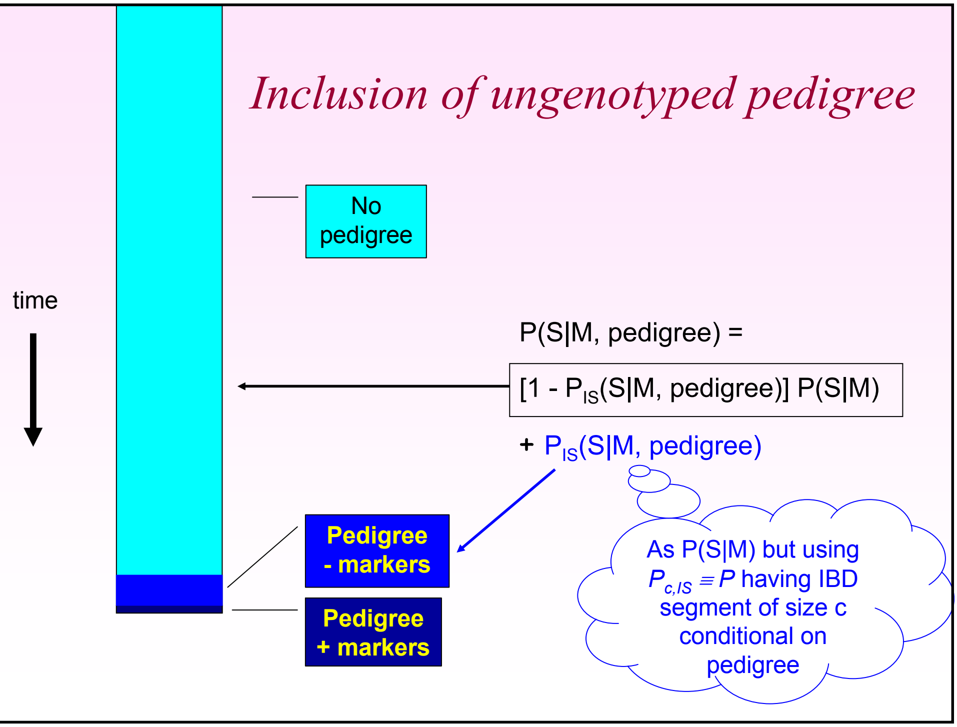


two regions of size 0 bracketing a region of size c

$$P\,(\text{‘}1x1\text{’}) = P_0\,(1-P_c)\,P_0 = E(F)^2\,(1-P_c)$$

1-exp(-T/2Ne) = E(F)

Inclusion of ungenotyped pedigree



P(S|M, pedigree) =

$[1 - P_{IS}(S|M, \text{pedigree})]\ P(S|M)$

$+\ P_{IS}(S|M, \text{pedigree})$

As P(S|M) but using $P_{c,IS} \equiv P$ having IBD segment of size c conditional on pedigree

What do we do next?

1. In the end, we obtain P(IBD|M) at any desired genome positions
2. ML estimates can be obtained maximizing

$$\ln L = -1/2\ [\text{Constant} + \log|\mathbf{V}| + (\mathbf{y}\text{-}\mathbf{X}\ \mathbf{b})'\ \mathbf{V}^{-1}\ (\mathbf{y} - \mathbf{X}\ \mathbf{b})],$$

with

$$\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$$

where

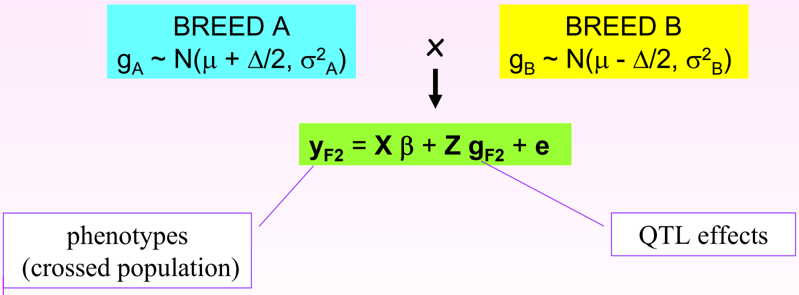
$$\mathbf{G} = \{\ P_{ij}(\text{IBD}|\text{M})\ \}$$

Day 1. Fine mapping and analysis of complex pedigrees

1. Combining linkage and linkage disequilibrium information
2. Analysis of crosses between outbred lines
3. QxPak software

Analysis of crosses between outbred lines

Pérez-Enciso & Varona (2000)



BREED A
$g_A \sim N(\mu + \Delta/2, \sigma^2_A)$

×

BREED B
$g_B \sim N(\mu - \Delta/2, \sigma^2_B)$

$$\mathbf{y}_{F2} = \mathbf{X}\,\beta + \mathbf{Z}\,\mathbf{g}_{F2} + \mathbf{e}$$

phenotypes
(crossed population)

QTL effects

Additive genic action is assumed between and within breeds.

Linkage equilibrium supposed within purebred individuals

y may contain also purebred records or any combination F2, BC,F3

$$\mathbf{y} = \mathbf{X}\,\beta + \mathbf{Z}\,\mathbf{g} + \mathbf{e}$$

$$\begin{pmatrix}\mathbf{y}\\ \mathbf{g}\\ \mathbf{e}\end{pmatrix} \sim N\left[\begin{pmatrix}\mathbf{Xb}+\mathbf{P}\,\Delta\\ \mathbf{P}\Delta\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{V} & \mathbf{GZ'} & \mathbf{R}\\ \mathbf{ZG} & \mathbf{G} & \mathbf{0}\\ \mathbf{R} & \mathbf{0} & \mathbf{R}\end{pmatrix}\right]$$

prob. of indiv i having received and allele of breed A origin at QTL j

$\mathbf{P} = \{ (p_{ij} - 1/2) \}$

j-th allellic effect

$\Delta = \{ \Delta_j \}$

$\mathbf{V} = \mathbf{ZGZ'} + \mathbf{R}$

$\mathbf{G} = \Sigma\, \mathbf{G}_j \,;\ \mathbf{G}_j = \{\mathrm{Cov}(g_{ij}, g_{i'j})\}$

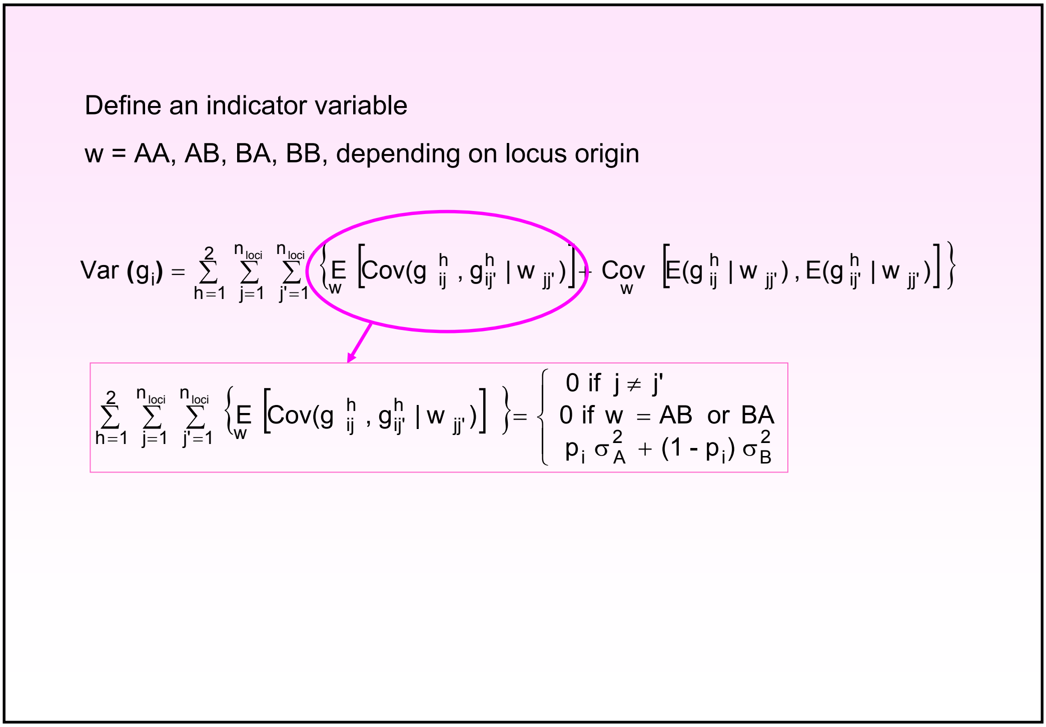
$\mathbf{R} = \mathbf{I}\, \sigma^2_e$

j-th QTL covariance (IBD) matrix

Define an indicator variable

w = AA, AB, BA, BB, depending on locus origin

$$\mathrm{Var}\ (\mathbf{g}_i) = \sum_{h=1}^{2} \sum_{j=1}^{n_{loci}} \sum_{j'=1}^{n_{loci}} \left\{ \underset{w}{E} \left[\mathrm{Cov}(g^{h}_{ij}, g^{h}_{ij'} \mid w_{jj'}) \right] + \underset{w}{\mathrm{Cov}} \left[E(g^{h}_{ij} \mid w_{jj'}), E(g^{h}_{ij'} \mid w_{jj'}) \right] \right\}$$

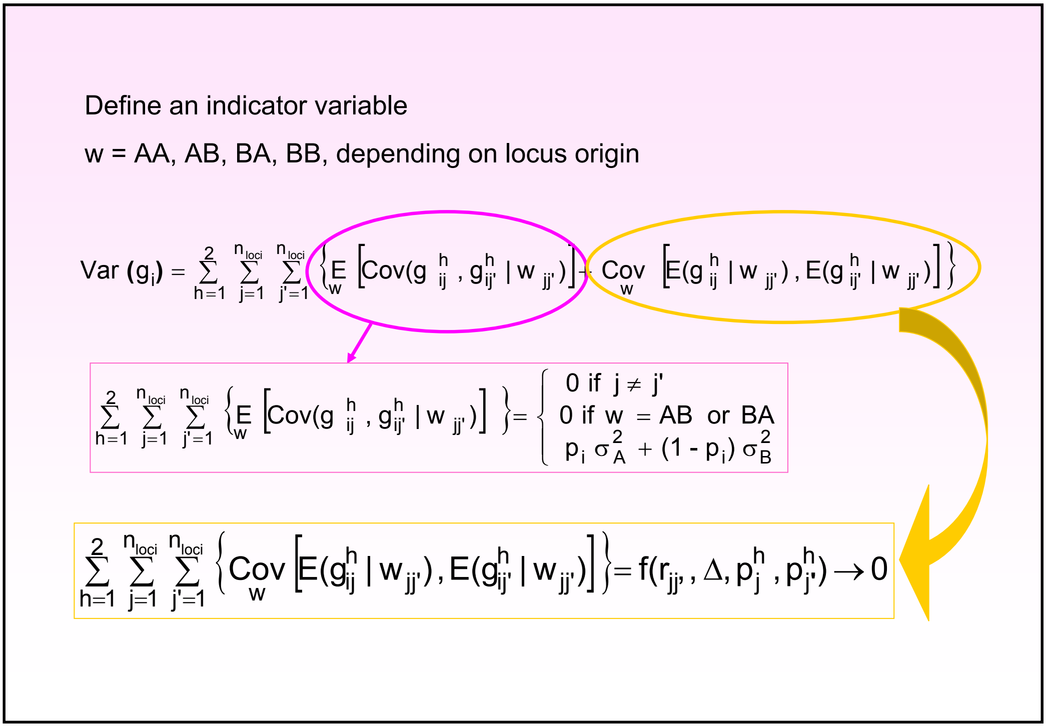


Define an indicator variable

w = AA, AB, BA, BB, depending on locus origin

$$\text{Var}\ (g_i) = \sum_{h=1}^{2}\sum_{j=1}^{n_{loci}}\sum_{j'=1}^{n_{loci}} \left\{ \underset{w}{E}\left[\text{Cov}(g^{h}_{ij}, g^{h}_{ij'} \mid w_{jj'})\right] + \underset{w}{\text{Cov}}\left[E(g^{h}_{ij} \mid w_{jj'}), E(g^{h}_{ij'} \mid w_{jj'})\right] \right\}$$

$$\sum_{h=1}^{2}\sum_{j=1}^{n_{loci}}\sum_{j'=1}^{n_{loci}} \left\{ \underset{w}{E}\left[\text{Cov}(g^{h}_{ij}, g^{h}_{ij'} \mid w_{jj'})\right] \right\} = \begin{cases} 0 \text{ if } j \neq j' \\ 0 \text{ if } w = AB \text{ or } BA \\ p_i\, \sigma_A^2 + (1 - p_i)\, \sigma_B^2 \end{cases}$$



Define an indicator variable

w = AA, AB, BA, BB, depending on locus origin

$$\text{Var}\ (g_i) = \sum_{h=1}^{2}\sum_{j=1}^{n_{loci}}\sum_{j'=1}^{n_{loci}} \left\{ \underset{w}{E}\left[\text{Cov}(g^{h}_{ij}, g^{h}_{ij'} \mid w_{jj'})\right] + \underset{w}{\text{Cov}}\left[E(g^{h}_{ij} \mid w_{jj'}), E(g^{h}_{ij'} \mid w_{jj'})\right] \right\}$$

$$\sum_{h=1}^{2}\sum_{j=1}^{n_{loci}}\sum_{j'=1}^{n_{loci}} \left\{ \underset{w}{E}\left[\text{Cov}(g^{h}_{ij}, g^{h}_{ij'} \mid w_{jj'})\right] \right\} = \begin{cases} 0 \text{ if } j \neq j' \\ 0 \text{ if } w = AB \text{ or } BA \\ p_i\, \sigma_A^2 + (1 - p_i)\, \sigma_B^2 \end{cases}$$

$$\sum_{h=1}^{2}\sum_{j=1}^{n_{loci}}\sum_{j'=1}^{n_{loci}} \left\{ \underset{w}{\text{Cov}}\left[E(g^{h}_{ij} \mid w_{jj'}), E(g^{h}_{ij'} \mid w_{jj'})\right] \right\} = f(r_{jj'}, \Delta, p^{h}_{j}, p^{h}_{j'}) \rightarrow 0$$

The **second term** in Var(g) is the increased variance due to segregation in crossed individuals but note that it tends to zero **CONDITIONAL** on marker information if these are highly informative and closely spaced. Suppose we could isolate a set of F2 individuals whose genome origin could be known without error, its genetic variance would be exactly

$$\sum_{j=1}^{n_{loci}} \delta_j \, \sigma^2_{Aj} + (1 - \delta_j) \, \sigma^2_{Bj} \quad ; \delta = \begin{cases} 1 \text{ if A origin} \\ 0 \text{ if B origin} \end{cases}$$

Then

Prob. of allele being of origin A

$$\text{Var}(g_{\,i}) \cong \sum_{h=1}^{2} \sum_{j=1}^{n_{loci}} p^h_{ij} \, \sigma^2_{Aj} + (1 - p^h_{ij}) \, \sigma^2_{Bj}$$

Can be applied to a QTL, or a genome portion, chr, etc

$$\text{Cov}\,(g_i\,, g_{i'}) = \sum_{h=1}^{2} \sum_{j=1}^{n_{loci}} \left[\rho^h_{A(i,i')j} \, \sigma^2_{Aj} + \rho^h_{B(i,i')j} \, \sigma^2_{Bj} \right]$$

Prob. of alleles from both indivs. Being IBD and being of origin **A** (**B**)

Finally, as usual

ML estimates can be obtained maximizing

$$\ln L = -1/2\,[\text{Constant} + \log|\mathbf{V}| + (\mathbf{y}-\mathbf{X}\,\beta)'\,\mathbf{V}^{-1}\,(\mathbf{y} - \mathbf{X}\,\beta)].$$

NOTE: This is a linearized likelihood in the sense that it approximates a mixture by a multivariate normal $\mathbf{y} \sim N(\mathbf{X}\beta, \mathbf{V})$.

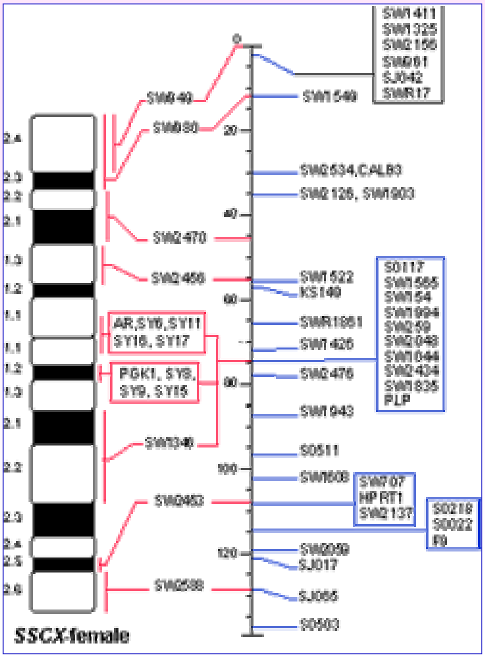
Examples:

Sex chromosome QTL in the IBMAP cross

Pérez-Enciso et al. (2002)

Whole genome analysis

Ponz et al. (2001)



Porcine X chr

Recall:

Different X/Y chr. lengths

Dosage compensation in mammals

Dosage compensation modelling

Dam origin allele

Dosage compensation parameter, ½ typically

males $y_M = \mu_M + g^2 + e$

females $y_F = \mu_F + \psi^1 g^1 + \psi^2 g^2 + d_{g1,g2} + e$

Sire origin allele

dominance

Dosage compensation modelling

$E(g_i)$

$$\Pr(g_i^2 \in A)\mu_{gA} + \Pr(g_i^2 \in B)\mu_{gB} \text{ male}$$

$$\sum_{h=1}^{2} \psi^h \, [\Pr(g_i^h \in A)\, \mu_{gA} + \Pr(g_i^h \in B)\, \mu_{gB}], \text{ females}$$

$\mathrm{Cov}(g_i, g_{i'}) =$

$$\Pr(g_i^2 \equiv g_{i'}^2 \in A)\sigma^2_{Ag} + \Pr(g_i^2 \equiv g_{i'}^2 \in B)\sigma^2_{Bg}, \text{ i,i' males}$$

Male female

$$\sum_{h=1}^{2} \psi h \, [\Pr(g_i^2 \equiv g_{i'}^h \in A)\, \sigma^2_{Ag} + \Pr(g_i^2 \equiv g_{i'}^h \in B)\, \sigma^2_{Bg}]$$

Female, female

$$\sum_{h=1}^{2} \sum_{h'=1}^{2} \psi h \, \psi h' \, [\Pr(g_i^h \equiv g_{i'}^{h'} \in A)\sigma^2_{Ag} + \Pr(g_i^h \equiv g_{i'}^{h'} \in B)\sigma^2_{Bg}]$$

Biometrical consequences of dosage compensation

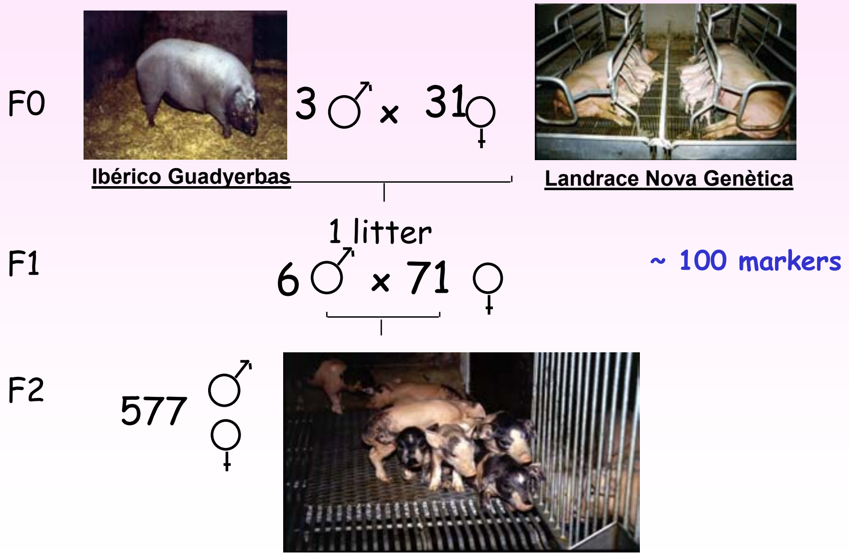
Provided that ψ = 1/2

$$\sigma^2_{gF} = \frac{1}{2}\sigma^2_{gM}$$

$$E[\mathrm{Cov}(FS_F)] = \frac{3}{4}\sigma^2_g; \quad [\frac{1}{2}\sigma^2_g < \mathrm{Cov} < \sigma^2_g]$$

$$E[\mathrm{Cov}(FS_{F,M})] = \frac{1}{4}\sigma^2_g; \quad [0 < \mathrm{Cov} < \frac{1}{2}\sigma^2_g]$$

IBMAP experimental protocol



F0

3 ♂ × 31 ♀

Ibérico Guadyerbas

Landrace Nova Genètica

1 litter

F1

6 ♂ × 71 ♀

~ 100 markers

F2

577 ♂ ♀

Traits analyzed

Carcass weight (**CW**)

Carcass length (**CL**),

pH at 24h post mortem (**pH**),

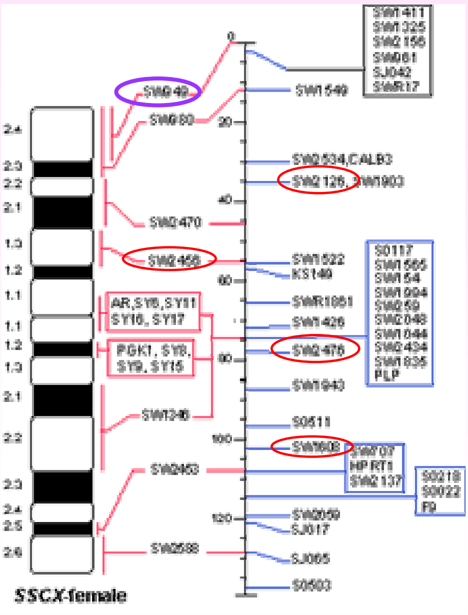
Minolta meat color components, **a***, **b***, and **L***

Haematin content (**Haem**)

Subcutaneous backfat thickness (**BFT**)

Longissimus muscle thickness (**LT**)

Intramuscular fat percentage (**IMF**)



Markers typed

Statistical analisys strategy

Step 1: 5 cM segments

Model – c1 : $\mathbf{y} = \mathbf{X}\,\beta + \mathbf{c}_s\,\mathbf{a}_s + \mathbf{u}_0 + \mathbf{e}$

Model – v1 : $\mathbf{y} = \mathbf{X}\,\beta + \mathbf{u}_s + \mathbf{u}_0 + \mathbf{e}$

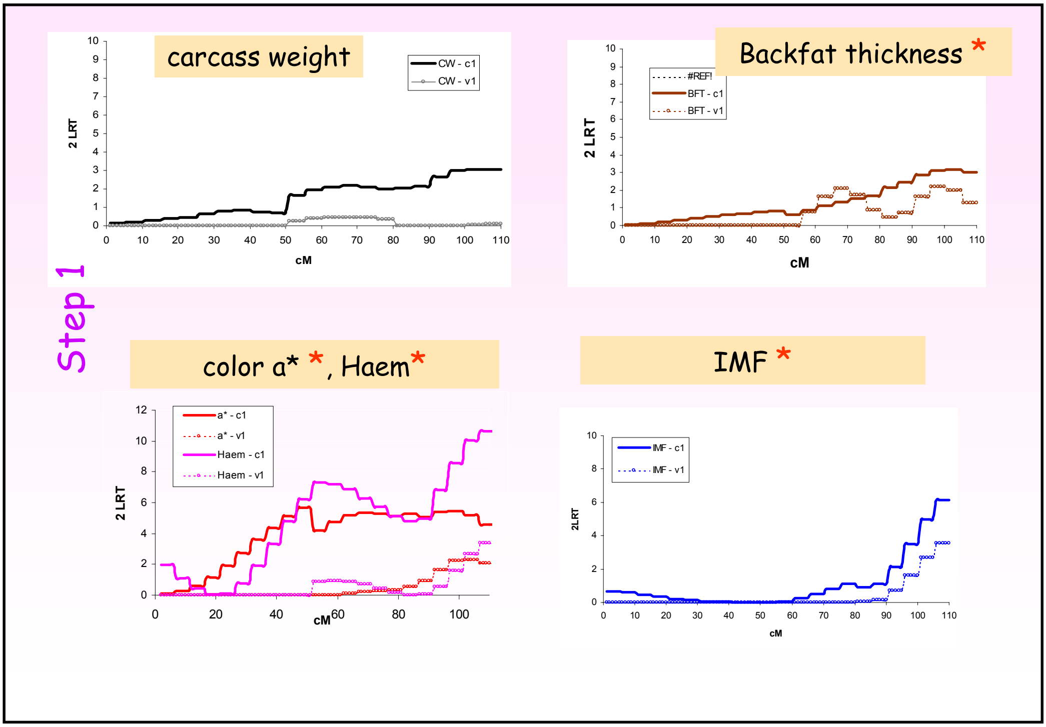
Step 2: 2cM segments

Model – c1 : $\mathbf{y} = \mathbf{X}\,\beta + \mathbf{c}_s\,\mathbf{a}_s + \mathbf{u}_0 + \mathbf{e}$

Model – c2 : $\mathbf{y} = \mathbf{X}\,\beta + \mathbf{c}_s\,a_s + \mathbf{c'}_s\,d_s + \mathbf{u}_0 + \mathbf{e}$

Model – m1 : $\mathbf{y} = \mathbf{X}\,\beta + \mathbf{c}_s\,a_s + \mathbf{u}_s + \mathbf{u}_0 + \mathbf{e}$

Model – m2 : $\mathbf{y} = \mathbf{X}\,\beta + \mathbf{c}_s\,a_s + \mathbf{u}_{sA} + \mathbf{u}_{sB} + \mathbf{u}_0 + \mathbf{e}$



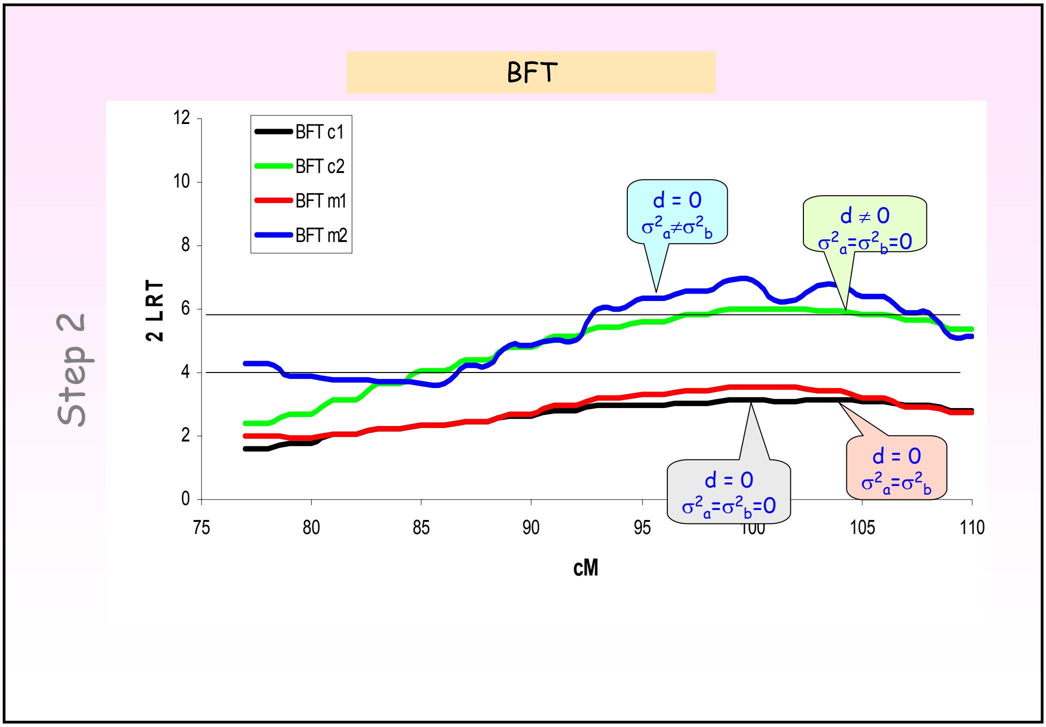
Step 1

carcass weight

Backfat thickness *

color a* *, Haem*

IMF *



Step 2

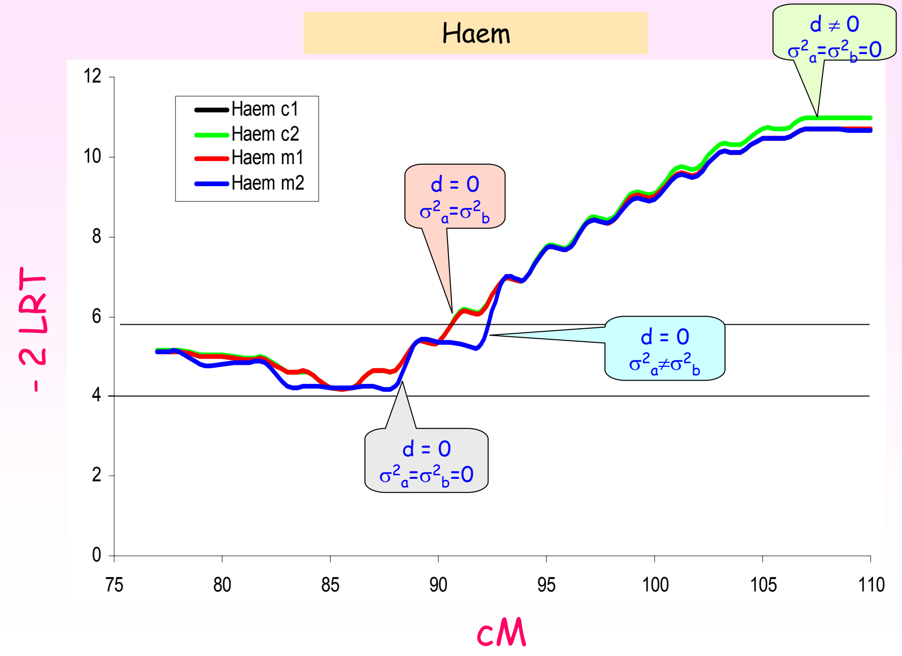
BFT

d = 0
$\sigma^2_a \neq \sigma^2_b$

d ≠ 0
$\sigma^2_a = \sigma^2_b = 0$

d = 0
$\sigma^2_a = \sigma^2_b = 0$

d = 0
$\sigma^2_a = \sigma^2_b$



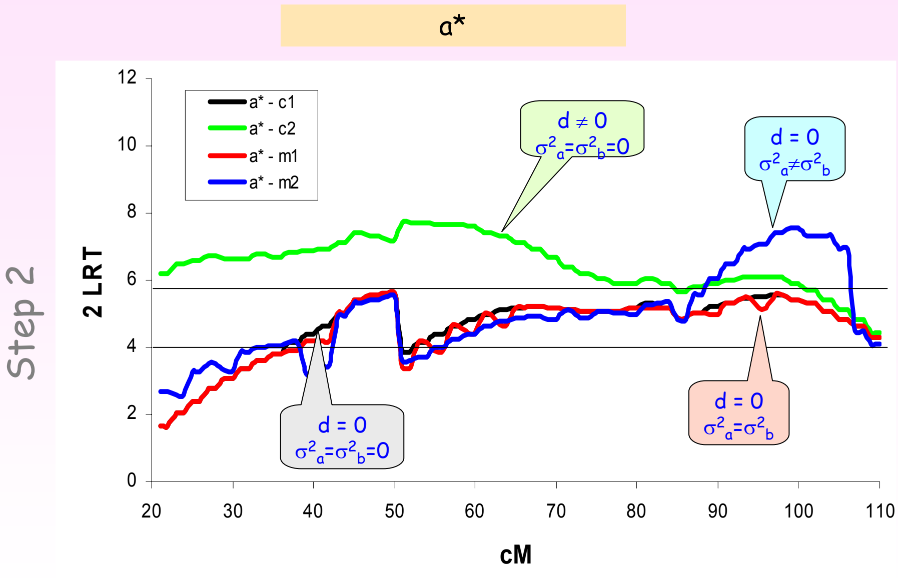
Haem

d ≠ 0 $\sigma^2_a=\sigma^2_b=0$

d = 0 $\sigma^2_a=\sigma^2_b$

d = 0 $\sigma^2_a\neq\sigma^2_b$

d = 0 $\sigma^2_a=\sigma^2_b=0$



Step 2

a*

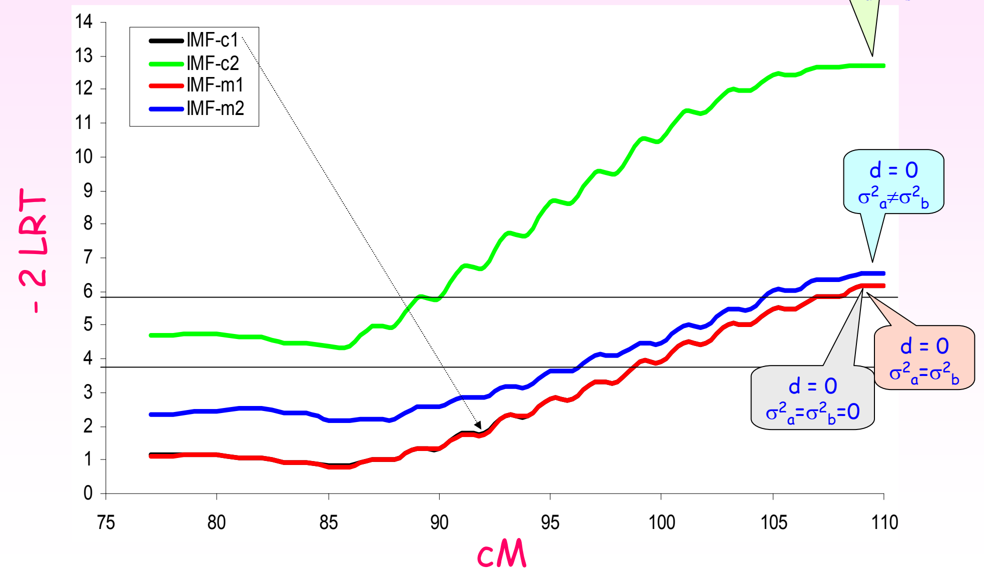
d ≠ 0 $\sigma^2_a=\sigma^2_b=0$

d = 0 $\sigma^2_a\neq\sigma^2_b$

d = 0 $\sigma^2_a=\sigma^2_b=0$

d = 0 $\sigma^2_a=\sigma^2_b$

Intramuscular fat %



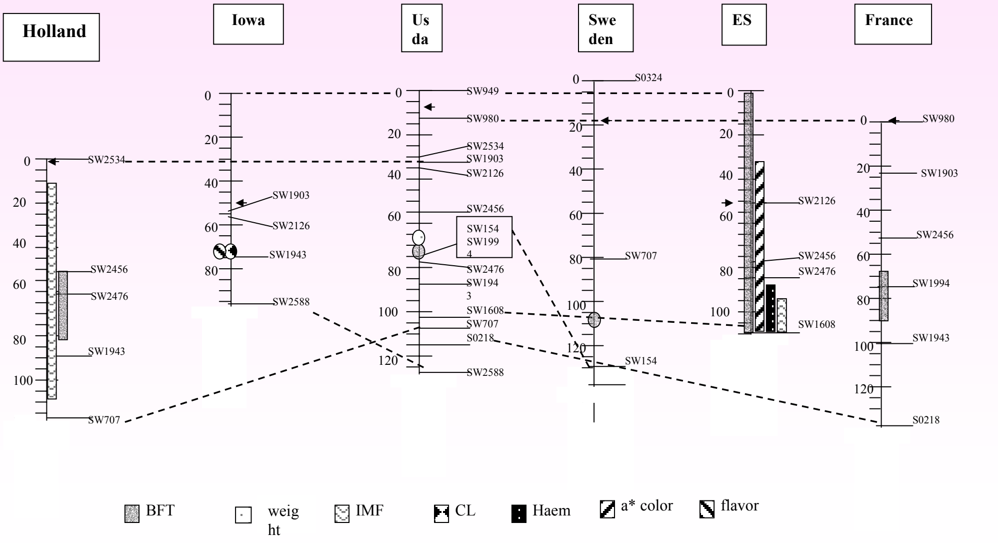
$d \neq 0$ $\sigma^2_a=\sigma^2_b=0$

$d = 0$ $\sigma^2_a\neq\sigma^2_b$

$d = 0$ $\sigma^2_a=\sigma^2_b$

$d = 0$ $\sigma^2_a=\sigma^2_b=0$

Comparative X-QTL mapping



Conclusions

A variety of genetic actions are revealed:

- Strong overdominance for IMF
- Additivity for Haem
- Alleles fixed in Iberian, not necessarily so in Landrace

Results for a* are difficult to interpret (2 QTL?)

Evidence for BFT not conclusive

All QTL experimients coincide in the most promising region

Example 2

Sex chromosome QTL in IBMAP cross

Pérez-Enciso et al. (2002)

Whole genome analysis: wool traits in sheep

Ponz et al. (2001)

Whole genome analysis

Recall...

a genome scan is not the only possible nor feasible strategy, common sense and statistical theory dictates that we should consider jointly all sources of variation.

This seems specially important if we want to discover epistatic relations.

Whole genome analysis: Segment Mapping

(Pérez-Enciso & Varona, 2000)

The **segment mapping** approach consists of dividing the region of interest (e.g., the whole genome) in a series of segments, bounded by arbitrary positions, and trying to obtain the most 'reasonable' partition.

- No distinction between a single QTL or n-QTLs within a segment.
- We are interested in quantifying the contribution to genetic variance of each segment rather than in estimating accurately the position.
- Generalization over classical approaches.
- No 'hierarchies' between segments.

Ponz et al., 2001

Synthetic sheep breed INRA401 = Romanov x Berrichon du cher.

Wool characteristics: staple length, mean fiber diameter, coefficient of variation of fiber diameter.

30 rams, 690 ewes and 1109 phenotyped offspring.

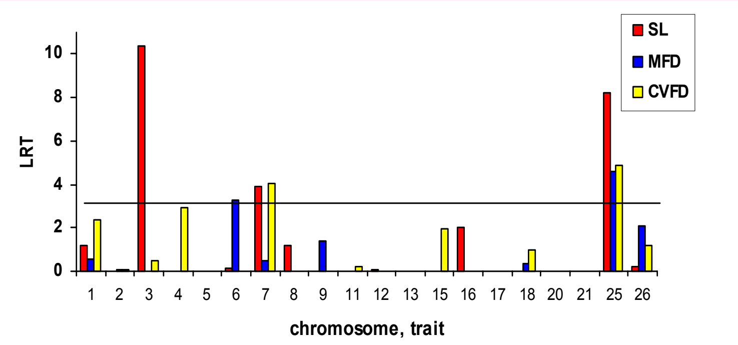
Sparse genotyping, 40 microsatellites distributed in 20 chromosomes out of 26 in the sheep genome.

$$\mathbf{y} = \mathbf{X}\,\beta + \Sigma_{s=0}\,\mathbf{g}_s + \mathbf{e}$$

- Which fraction of the genetic variance is explained by typed markers?
- Which is the most reasonable course of action to take?

Analysis steps

1. One initial segment per chromosome; model M(0, chr) vs. M(0) for each chr. in turn.
2. For those significant chrs. ($P < 0.05$), split chrs. in subsegments a, b, c... delimited by each consecutive marker or by half the chr. if only two markers. Choose the combination with maximum likelihood, e.g., max from M(0,4), M(0, 4a), M(0,4b).
3. Assess the fraction of total genetic variance explained by selected segments and assess whether all variance is explained by these segments by comparing model M(0, i, j, k ...) vs. M(i, j, k, ...).



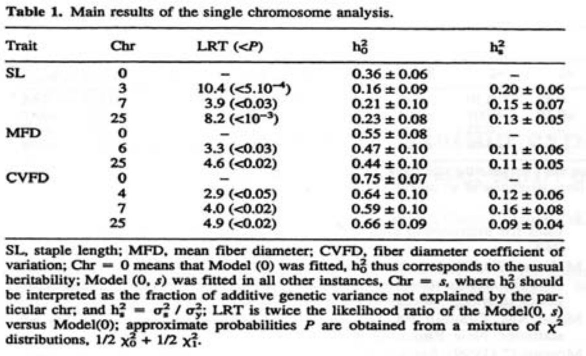


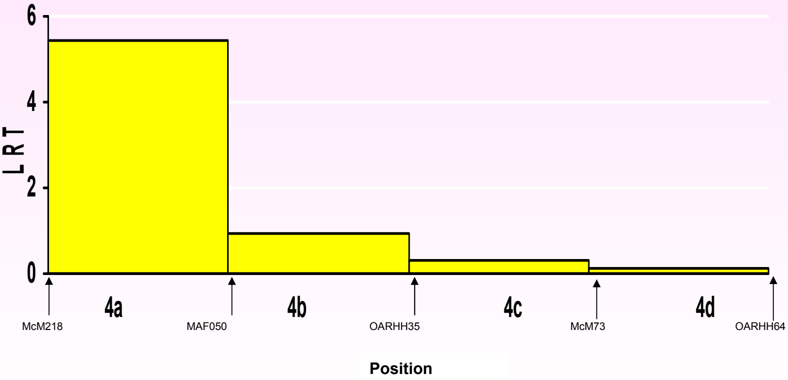
Table 1. Main results of the single chromosome analysis.

Trait	Chr	LRT (<*P*)	h_0^2	h_s^2
SL	0	–	0.36 ± 0.06	–
	3	10.4 (<5.10^{-4})	0.16 ± 0.09	0.20 ± 0.06
	7	3.9 (<0.03)	0.21 ± 0.10	0.15 ± 0.07
	25	8.2 (<10^{-3})	0.23 ± 0.08	0.13 ± 0.05
MFD	0	–	0.55 ± 0.08	–
	6	3.3 (<0.03)	0.47 ± 0.10	0.11 ± 0.06
	25	4.6 (<0.02)	0.44 ± 0.10	0.11 ± 0.05
CVFD	0	–	0.75 ± 0.07	–
	4	2.9 (<0.05)	0.64 ± 0.10	0.12 ± 0.06
	7	4.0 (<0.02)	0.59 ± 0.10	0.16 ± 0.08
	25	4.9 (<0.02)	0.66 ± 0.09	0.09 ± 0.04

SL, staple length; MFD, mean fiber diameter; CVFD, fiber diameter coefficient of variation; Chr = 0 means that Model (0) was fitted, h_0^2 thus corresponds to the usual heritability; Model (0, *s*) was fitted in all other instances, Chr = *s*, where h_0^2 should be interpreted as the fraction of additive genetic variance not explained by the particular chr; and $h_s^2 = \sigma_s^2 / \sigma_y^2$; LRT is twice the likelihood ratio of the Model(0, *s*) versus Model(0); approximate probabilities *P* are obtained from a mixture of χ^2 distributions, 1/2 χ_0^2 + 1/2 χ_1^2.

Single chromosome step

Chromosome dissection step: CV diameter, chr. 4



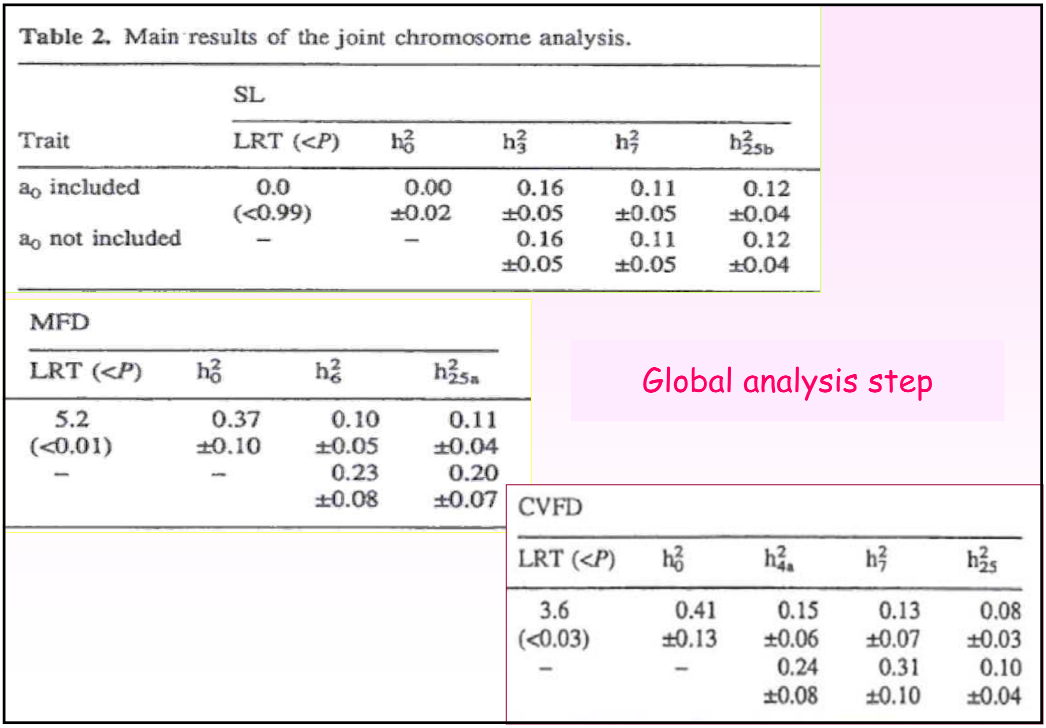


Table 2. Main results of the joint chromosome analysis.

Trait	SL LRT ($<P$)	h_0^2	h_3^2	h_7^2	h_{25b}^2
a_0 included	0.0 (<0.99)	0.00 ±0.02	0.16 ±0.05	0.11 ±0.05	0.12 ±0.04
a_0 not included	–	–	0.16 ±0.05	0.11 ±0.05	0.12 ±0.04

Trait	MFD LRT ($<P$)	h_0^2	h_6^2	h_{25a}^2
a_0 included	5.2 (<0.01)	0.37 ±0.10	0.10 ±0.05	0.11 ±0.04
a_0 not included	–	–	0.23 ±0.08	0.20 ±0.07

Trait	CVFD LRT ($<P$)	h_0^2	h_{4a}^2	h_7^2	h_{25}^2
a_0 included	3.6 (<0.03)	0.41 ±0.13	0.15 ±0.06	0.13 ±0.07	0.08 ±0.03
a_0 not included	–	–	0.24 ±0.08	0.31 ±0.10	0.10 ±0.04

Global analysis step

We concluded ...

1. All genetic variance for staple length is explained, we should pursue fine mapping.
2. About 60% of total genetic variance is not explained with current typing in mean fiber diameter: pursue genotyping other regions.
3. Results for CV of fiber diameter are intermediate for this trait, about 50% of variance explained: course of action is less obvious.

Day 1. Fine mapping and analysis of complex pedigrees

1. Combining linkage and linkage disequilibrium information
2. Analysis of crosses between outbred lines
3. QxPak software

QxPak

By M. Pérez-Enciso & I. Misztal

a versatile package for QTL & genetical genomics

Main features

- Multitrait
- Multi QTL
- Different models per trait
- Any number of chromosomes can be analyzed jointly
- Missing observations
- (approximate) Dealing with missing markers
- Flexible QTL modelling
- QTL x other effect (say sex) interaction
- Linkage vs Epistasis tests
- Friendly input file
- Can also be used efficiently for infinitesimal model analyses
- All individuals are included in the analyses

Four grand options

1. Classical REML/ML analyses
2. QTL studies
3. Genetical genomics
4. SNP association studies

Input file

ML_option *
Multitrait_option *
Datafile *
Outfile *
Markerfile *
Haplotypefile *
Number_of_inds
Number_of_qtl
Number_of_effects
Number_of_chromosomes
Marker_positions
Number_of_traits *
Number_of_MCMC_iterations *
Scan_step *
QTL
Effect
Trait
Initial_res_var *
Initial_gen_var *
Test *

QTL modelling

Qtl types defined are:
fix_a: additive fixed effect
fix_d: dominant fixed effect
fix_ad: add+dom fixed effect
snp_a: additive fixed effect (SNP)
snp_d: dominant fixed effect (SNP)
snp_ad: add+dom fixed effect (SNP)
ran_1: additive random effect (common variance to all breeds)
ran_2: additive random effects (different variance per breed)
mix_1a: mixed effect (fix_a + ran_1)
mix_1d: mixed effect (fix_d + ran_1)
mix_1ad: mixed effect (fix_ad + ran_1)
mix_2a: mixed effect (fix_a + ran_2)
mix_2d: mixed effect (fix_d + ran_2)
mix_2ad: mixed effect (fix_ad + ran_2)

Day_1_take_home_message

1. Fine mapping is a risky and very labor intensive task.
2. The main difficulty with complex pedigrees lies in computing IBD probabilities, MCMC methods are the sole means but they are not the panacea.
3. Similarly, Bayesian statistics is very attractive and helpful but does not solve the main problem.
4. Much work remains to be done to combine LD and linkage methods. Assessing the QTL genotype correctly is paramount.
5. A genome scan is not the only possible strategy in a QTL analysis.

Literature

Blasco A (2001) The Bayesian controversy in animal breeding. J Anim Sci 79: 2023-46

Hayes, B., Visscher, P.M., McPartlan, H.C., & Goddard, M.E. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effctive population size. Genome Research 13:635-643.

Liu,J.S., Sabatti,C., Teng,J., Keats,B.J. & Risch,N. Bayesian analysis of haplotypes for linkage disequilibrium mapping. Genome Res. 11, 1716-24 (2001).

Meuwissen, T. H., & Goddard, M. E. (2001). Prediction of identity by descent probabilities from marker-haplotypes. *Genet Sel Evol* **33,** 605-634.

Meuwissen, T. H., Karlsen, A., Lien, S., Olsaker, I., & Goddard, M. E. (2002). Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* **161,** 373-379.

Morris, A. P., Whitaker, J. C., & Balding, D. J. (2000). Bayesian fine-scale mapping of disease loci, by Hidden Markov Models. *Am. J. Hum. Genet.* **67,** 155-169.

Morris, A. P., Whittaker, J. C., & Balding, D. J. (2002). Fine-Scale Mapping of Disease Loci via Shattered Coalescent Modeling of Genealogies. *Am J Hum Genet* **70,** 686-707.

Perez-Enciso, M., Clop, A., Folch, J. M., Sanchez, A., Oliver, M. A., Ovilo, C., Barragan, C., Varona, L., & Noguera, J. L. (2002). Exploring Alternative Models for Sex-Linked Quantitative Trait Loci in Outbred Populations. Application to an iberian x landrace pig intercross. *Genetics* **161,** 1625-1632.

Pérez-Enciso, M., & Varona, L. (2000). Quantitative trait loci mapping in F2 crosses between outbred lines. *Genetics* **155,** 391-405.

Pérez-Enciso, M. (2003) Fine mapping of complex trait genes combining pedigree and linkage disequilibrium information: A unified Bayesian framework. Genetics, 163: 1497-510.

Ponz, R., Moreno, C., Allain, D., Elsen, J. M., Lantier, F., Lantier, I., Brunel, J. C., & Pérez-Enciso, M. (2001). Assessment of genetic variation explained by markers for wool traits in sheep via a segment mapping approach. *Mamm Genome* **12,** 569-572.

Sorensen D, Gianola D (2002) Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics. Springer Verlag, New York

Uimari, P., and M. J. Sillanpää, 2001 Bayesian oligogenic analysis of quantitative and qualitative traits in general pedigrees. Genet. Epidem. **21:** 224-242.

Yi, N., & Xu, S. (2001). Bayesian mapping of quantitative trait loci under complicated mating designs. *Genetics* **157,** 1759-1771.