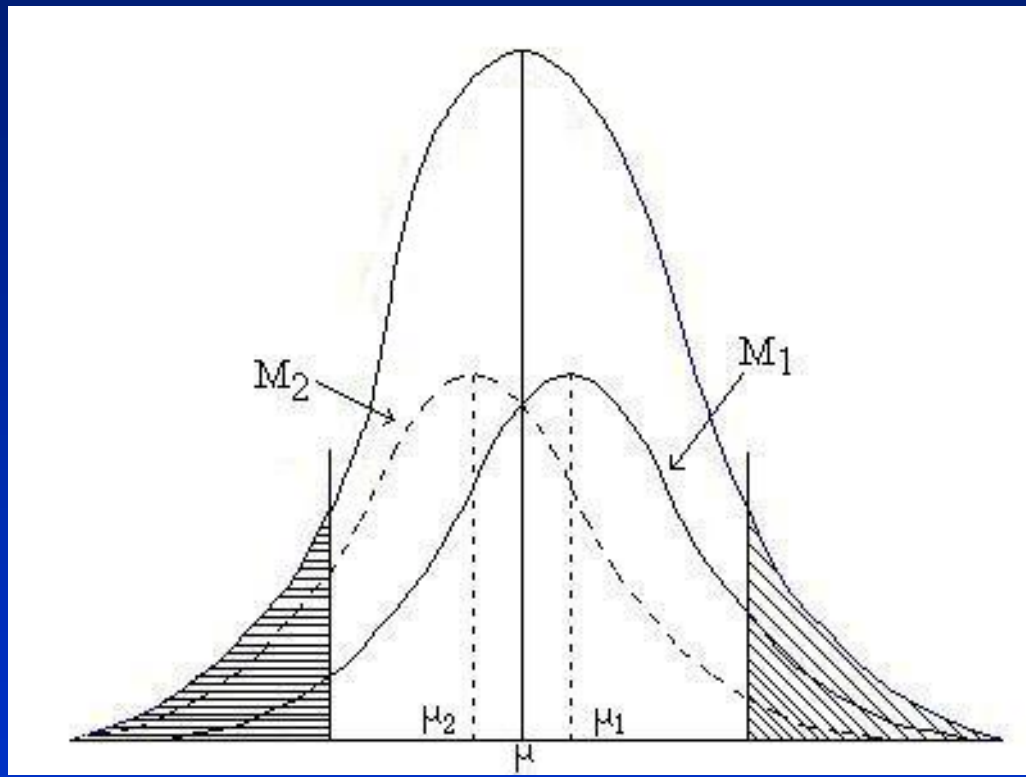
Optimising the design of linkage experiments to detect QTL

- Key parameters are:
  - distribution of QTL effects (how QTL are potentially detectable in a mapping experiment)
  - population structure
  - significance thresholds
  - precision of QTL mapping (width of confidence interval)
  - efficient genotyping strategies

Selective genotyping

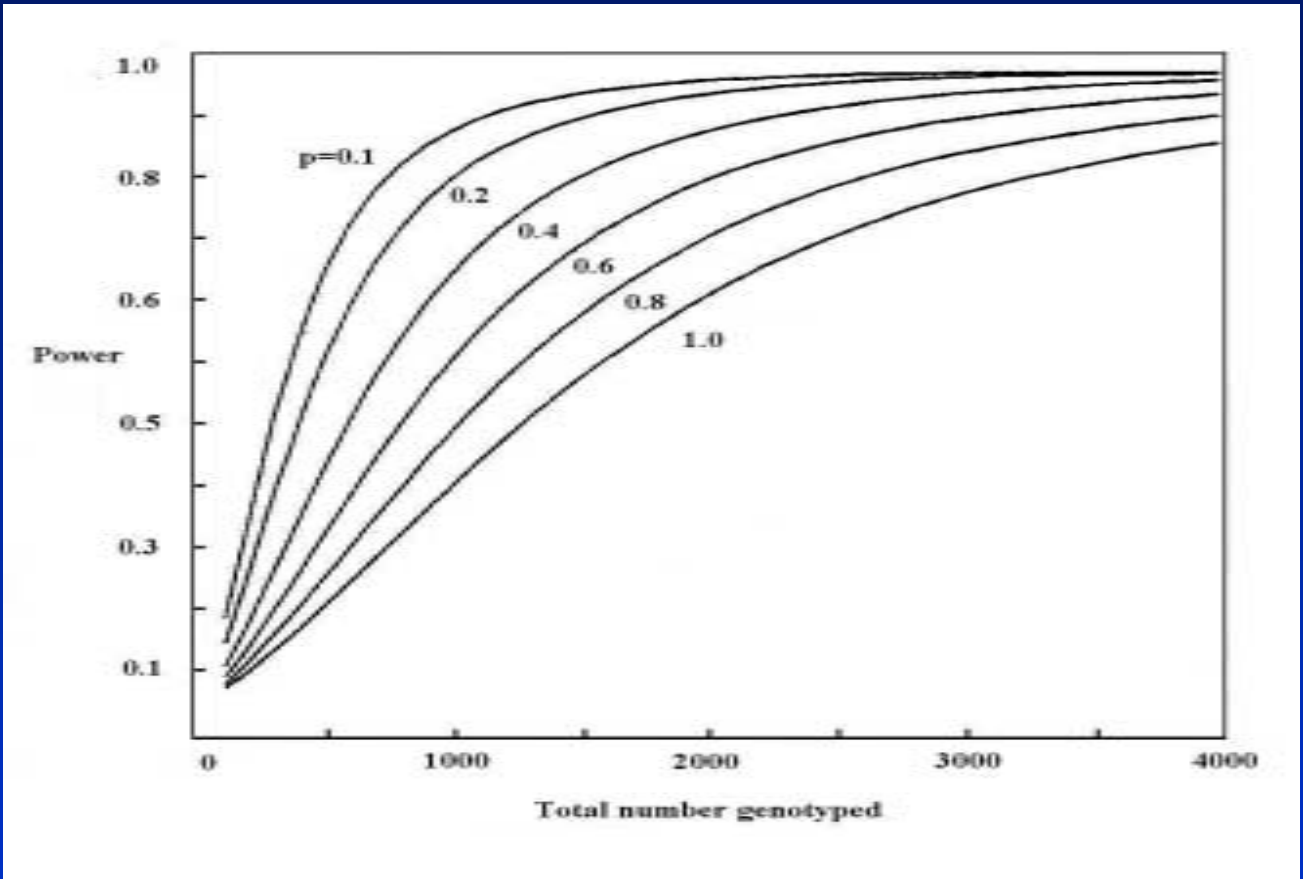
- Individuals most deviating from mean are most informative for linkage, as their QTL genotypes can be inferred from their phenotypes more clearly than progeny with average phenotypes



Selective genotyping

- In fact, not necessary to genotype more than 50% of a population to get maximum power from a design.
- Selective genotyping: only genotype progeny within a half sib family with extreme high and low phenotypes
- Either
  - increase the power for a given number of total genotypes,
  - or reduce cost for a total number of phenotypes (progeny)

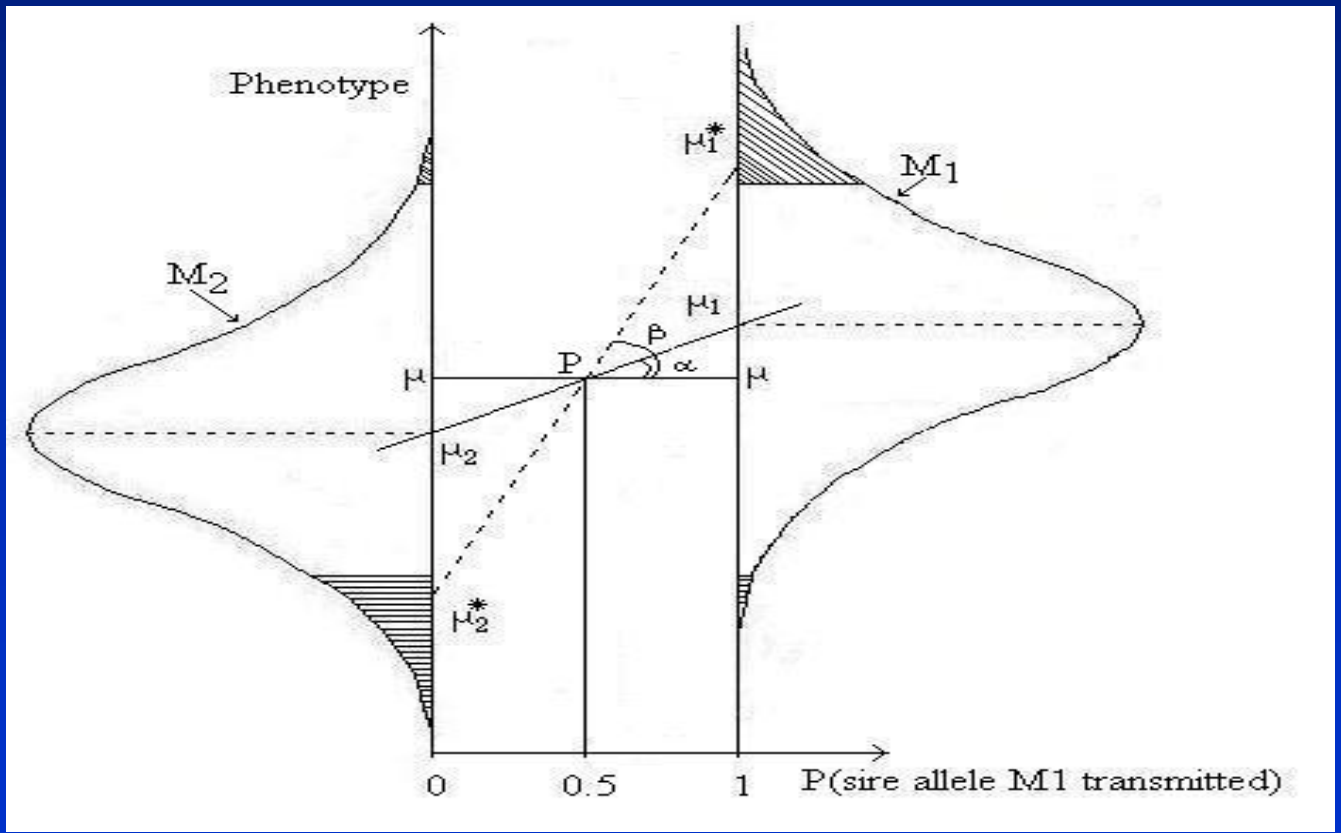
Selective genotyping



- But recommend selection be at least 10% in either tail, data may contain artefacts

Selective genotyping

- Problem: QTL variance is over-estimated with selective genotyping



- will erode advantage of subsequent MAS

Selective genotyping

- Solution: include pedigree and phenotypes of un-genotyped animals in a variance component analysis
  - assumes each every animal carries two unique QTL alleles
  - sire alleles A, B
  - progeny 1 sire allele A, progeny 2 sire allele A, progeny 3 sire allele B, progeny 4 is not genotyped
  - IBD (or G) matrix, assuming QTL at marker and tracing sire alleles only, is:

	P1	P2	P3	P4
P1	1			
P2	1	1		
P3	0	0	1	
P4	0.5	0.5	0.5	1

Selective genotyping

- Solution: include pedigree and phenotypes of un-genotyped animals in a variance component analysis

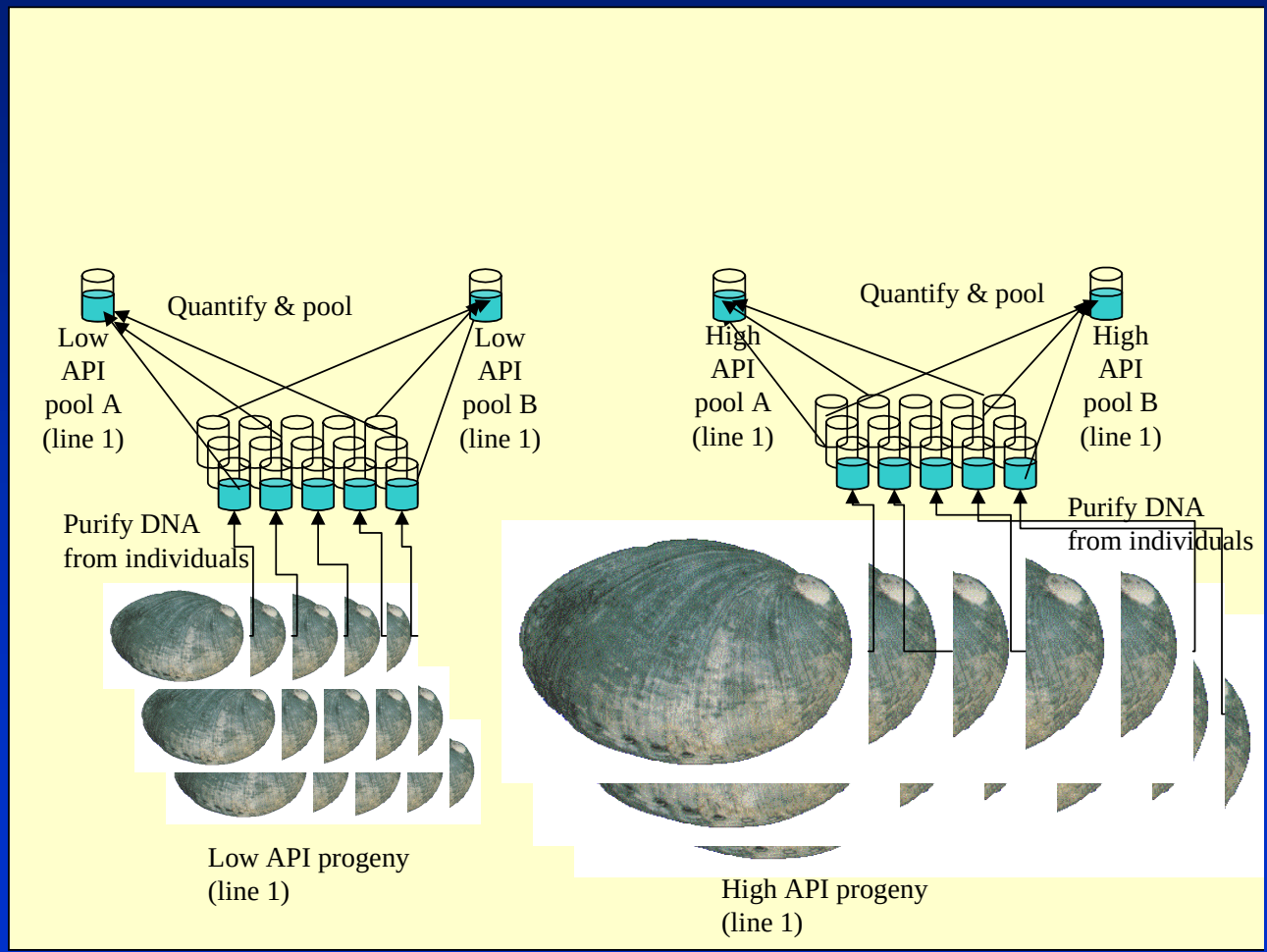
$$Y = \mu + Zu + Zv + e$$

  - **Y** = vector of phenotypes, μ = mean, **Z** a design matrix, **u** a vector of polygenic effects, **v** a vector of QTL allele effects, **e** a vector of random residuals, where
  - $u \sim (0, \mathbf{A}\sigma_u^2)$, $v \sim (0, \mathbf{G}\sigma_v^2)$, $e \sim \sim (0, \mathbf{I}\sigma_e^2)$
- QTLs as random regresses effect back towards zero

Strategy	QTL size
True	0.32
100% genotyped	0.30±0.02
20% genotyped	0.93±0.02
20% genotyped, ungenotyped animals included in the analysis	0.31±0.02

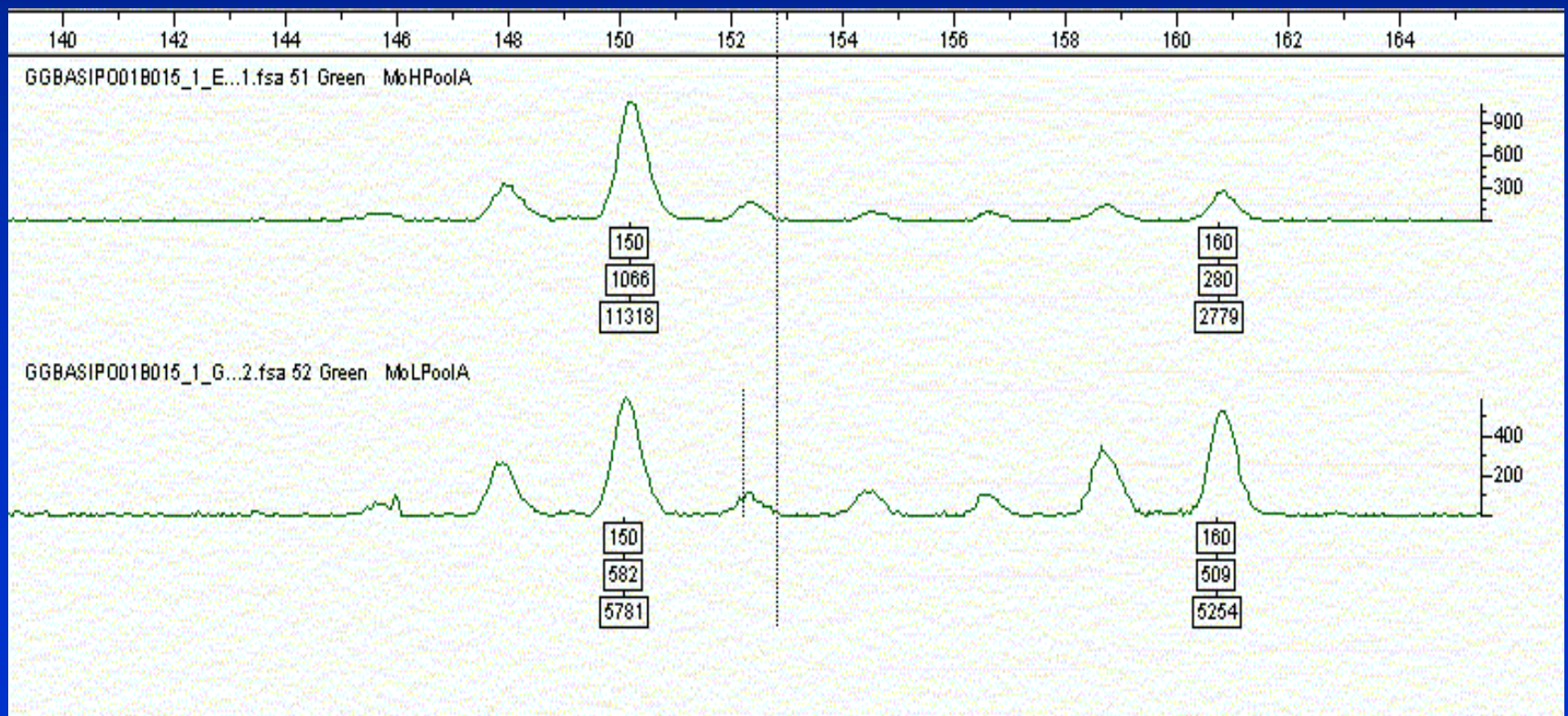
Selective DNA pooling (Darvisi and Soller 1994)

- Pool DNA of high and low phenotype animals



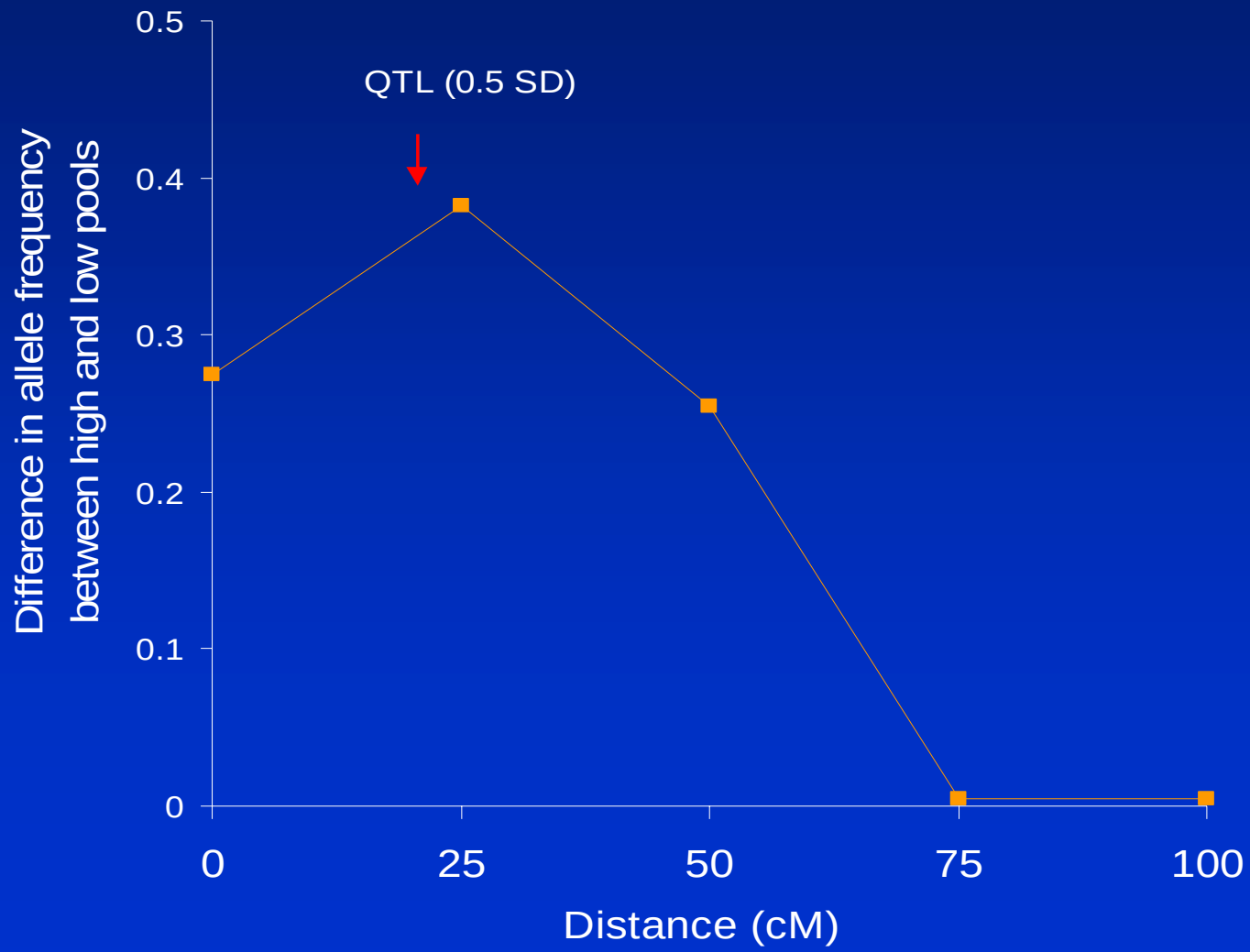
Selective DNA pooling

- Determine linkage by distribution of sire alleles between pools of DNA of high and low phenotypes
- For marker BMS12, sire 1 150--Q
  160--q



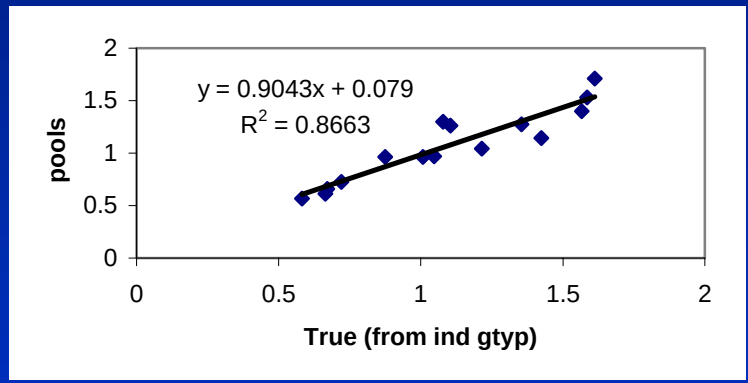
Selective DNA pooling

- Genome scan…..



Selective DNA pooling

- Three difficulties with DNA pooling
  - Very accurate quantification of amount of DNA from each animal required (kits available?) to estimate allele frequency differences with any precision



  - With microsatellite markers, estimates of allele frequencies confounded by stutter bands, but correction procedures have been devised
  - Only has power to detect QTL for the trait on which the pools were based

Selective DNA pooling

- Has been used to detect QTL affecting protein% in milk from Israeli-Holstein Friesian cattle (Lipkin et al 1998)
- Accessed 80.6% and 48.3% of power available from selective and full genotyping, respectively
- Statistical power of 45 600 of individual genotypings obtained from 328 pool genotypings (5 significant effects were detected)
- “The DNA pooling methodology can make genome wide mapping of QTL accessible to moderately sized breeding organisations”
- Need good people in the lab though!

Linkage mapping in complex pedigrees

- In some species it is difficult or expensive to create large half-sib families or line crosses (eg. humans).
- An alternative is to use linkage information from existing pedigree (genotype existing animals)
  - potentially a large number of recombination events can be accessed
  - In practise, the large number of missing genotypes can reduce the power of complex pedigrees for QTL mapping

Linkage mapping in complex pedigrees

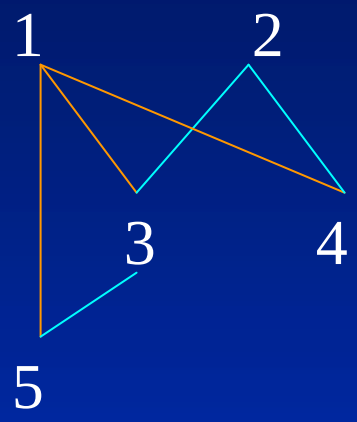
- A two stage approach for linkage mapping in complex pedigrees
  1. For each putative QTL position, calculate QTL (co) variance matrix. Also called the **IBD** or **G** matrix, has elements G_{ij}=Prob(QTL alleles i and j are identical by descent or IBD)
  2. For each position considered in step 1, construct a linear model to estimate QTL variances and other parameters, test for presence of QTL

Linkage mapping in complex pedigrees

- Calculating the IBD matrix
  - Dimensions (2 x number of animals) x (2 x number of animals), 2 QTL alleles for each animal
  - If marker information was complete, would contain 0s and 1s only.
  - The more marker genotypes are missing, the more the IBD matrix looks like the **A** matrix

Linkage mapping in complex pedigrees

Id	Sire	Dam	Marker 1 Allele 1	Marker 1 Allele 2
1	0	0	A	B
2	0	0	C	D
3	1	2	A	C
4	1	2	B	D
5	1	3	A	C

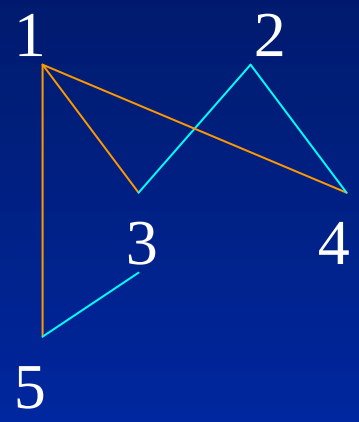


	Sire 1	Dam 2	Prog 3	Prog 4	Prog 5
Sire 1	1				
	0	1			
Dam 2	0	0	1		
	0	0	0	1	
Prog 3	1	0	0	0	1
	0	0	1	0	0
Prog 4	0	1	0	0	0
	0	0	0	1	0
Prog 5	1	0	0	0	1
	0	0	1	0	0

						1				
					1		0			
				1		0		0		
			1		0		0		0	
		1		0		0		0		0
	1		0		0		0		0	
1		0		0		0		0		0
	0		0		0		0		0	
0		0		0		0		0		0

Linkage mapping in complex pedigrees

Id	Sire	Dam	Marker 1 Allele 1	Marker 1 Allele 2
1	0	0	A	B
2	0	0	C	D
3	1	2	0	0
4	1	2	B	D
5	1	3	0	0



	Sire 1		Dam 2		Prog 3		Prog 4		Prog 5	
Sire 1	1									
	0	1								
Dam 2	0	0	1							
	0	0	0	1						
Prog 3	0.5	0.5	0	0	1					
	0	0	0.25	0.25	0	1				
Prog 4	0	1	0	0	0.5	0.5	1			
	0	0	0	1	0.5	0.5	0	1		
Prog 5	0.5	0.5	0	0	0.5	0	0	0	1	
	0.25	0.25	0.25	0.25	0.5	0.25	0	0.25	0.25	1

Linkage mapping in complex pedigrees

- Variance component model for estimation of QTL parameters

$$y_i = u_i + v_i^p + v_i^m + e_i$$

- y_i=phenotype of animal i
- u_i=polygenic effect of animal i
- v^p_i=effect of paternal allele for animal i
- v^m_i=effect of maternal allele for animal i

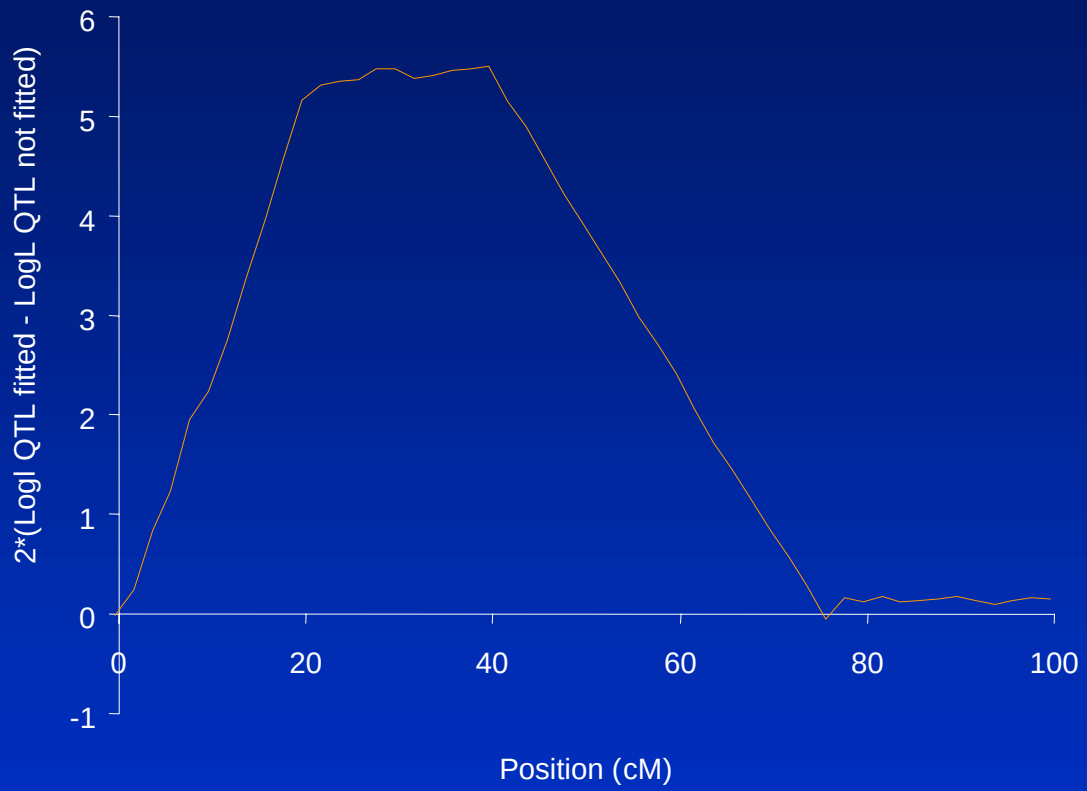
Linkage mapping in complex pedigrees

- Model

$$Y = \mu + Xb + Zu + Zv + e$$

  - **Y** = vector of phenotypes, μ = mean, **X**, **Z** and **W** are design matrices, **b** is a vector of fixed effects, **u** a vector of polygenic effects, **v** a vector of QTL allele effects, **e** a vector of random residuals, where
  - $u \sim (0, \mathbf{A}\sigma_u^2)$, $v \sim (0, \mathbf{G}\sigma_v^2)$, $e \sim \sim (0, \mathbf{I}\sigma_e^2)$
- For each putative QTL position compare LogL from above model and animal model only
  - $Y = \mu + Xb + Zu + e$

Linkage mapping in complex pedigrees



- Under null hypothesis of no QTL
  - 2*(LogL QTL fitted - LogL QTL not fitted) is distributed as a $\chi^2_{1,2\alpha}$ where α is the desired significance level (eg. at α=0.1 is 2.71)

Linkage mapping in complex pedigrees

- Advantages/disadvantages of complex pedigrees
  - can use existing animals
  - inferring missing genotypes can be complicated, alleles tracked over multiple generations
    - difficult in livestock pedigrees, where inbreeding and marriage loops are common
    - Simulation based methods (MCMC) most often used
    - UNE tools for segregation analysis?
  - Considerable advantage is that marker assisted breeding values (MEBV) are produced from the analysis
    - select from the current generation of candidates

Take home messages for today

- 10 or so QTL explain majority of total genetic variance
- Need experiments that can detect QTL =>0.2σp
- Make the half-sib families >> large!!!!
  - otherwise a waste of time
- Use efficient genotyping strategies to increase the power and decrease the cost of your experiment
  - selective genotyping
  - DNA pooling
  - complex pedigrees
- Work closely with people in the lab!