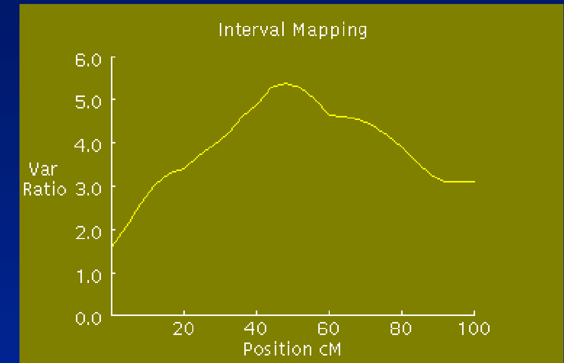
From Linkage to Gene Detection

Ben Hayes and Mike Goddard

Why detect QTL?

- Use markers linked to QTL in MAS
  - > genetic gain (esp. hard to select for traits)
- Use markers/ marker haplotypes in LD with QTL in MAS
  - >> genetic gain
- Find genetic mutation underlying QTL effect
  - patent = $$$$ (maybe)

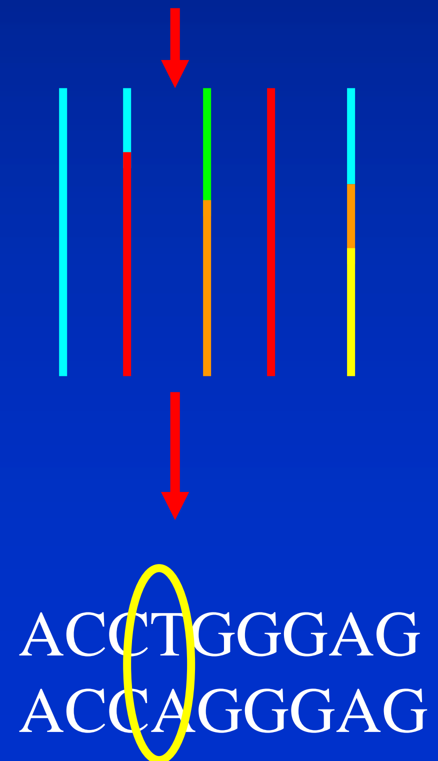
DGAT1 - A success story (Grisart et al. 2002)



1. Linkage mapping detects a QTL on bovine chromosome 14 with large effect on fat % (Georges et al1995)

2. Linkage disequilibrium mapping refines position of QTL (Riquet et al. 1999)

3. Selection of candidate genes. Sequencing reveals point mutation in candidate (DGAT1). This mutation found to be functional - substitution of lysine for analine. Gene patented. (Grisart et al. 2002)



Aim of course:

- *Provide you with a set of criteria for the design and analyses of successful QTL mapping experiments*

Optimising the design of linkage experiments to detect QTL

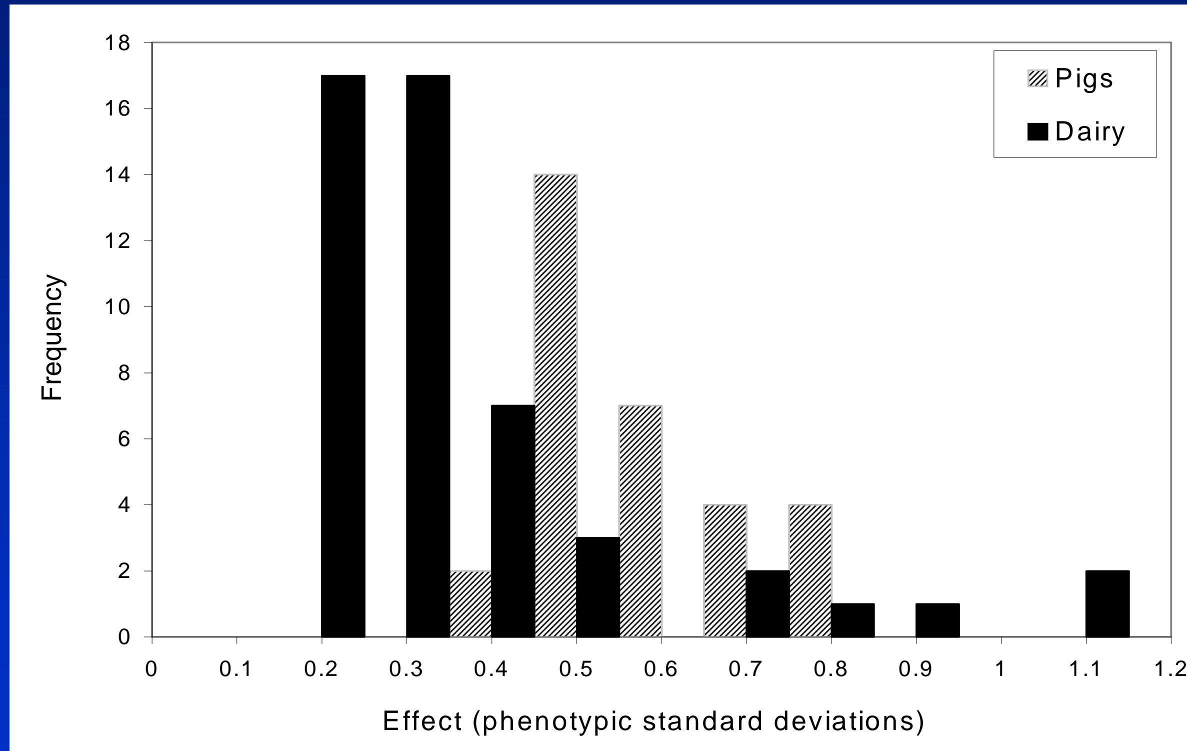
- Key parameters are:
  - distribution of QTL effects (how QTL are potentially detectable in a mapping experiment)
  - population structure
  - significance thresholds
  - precision of QTL mapping (width of confidence interval)
  - efficient genotyping strategies

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

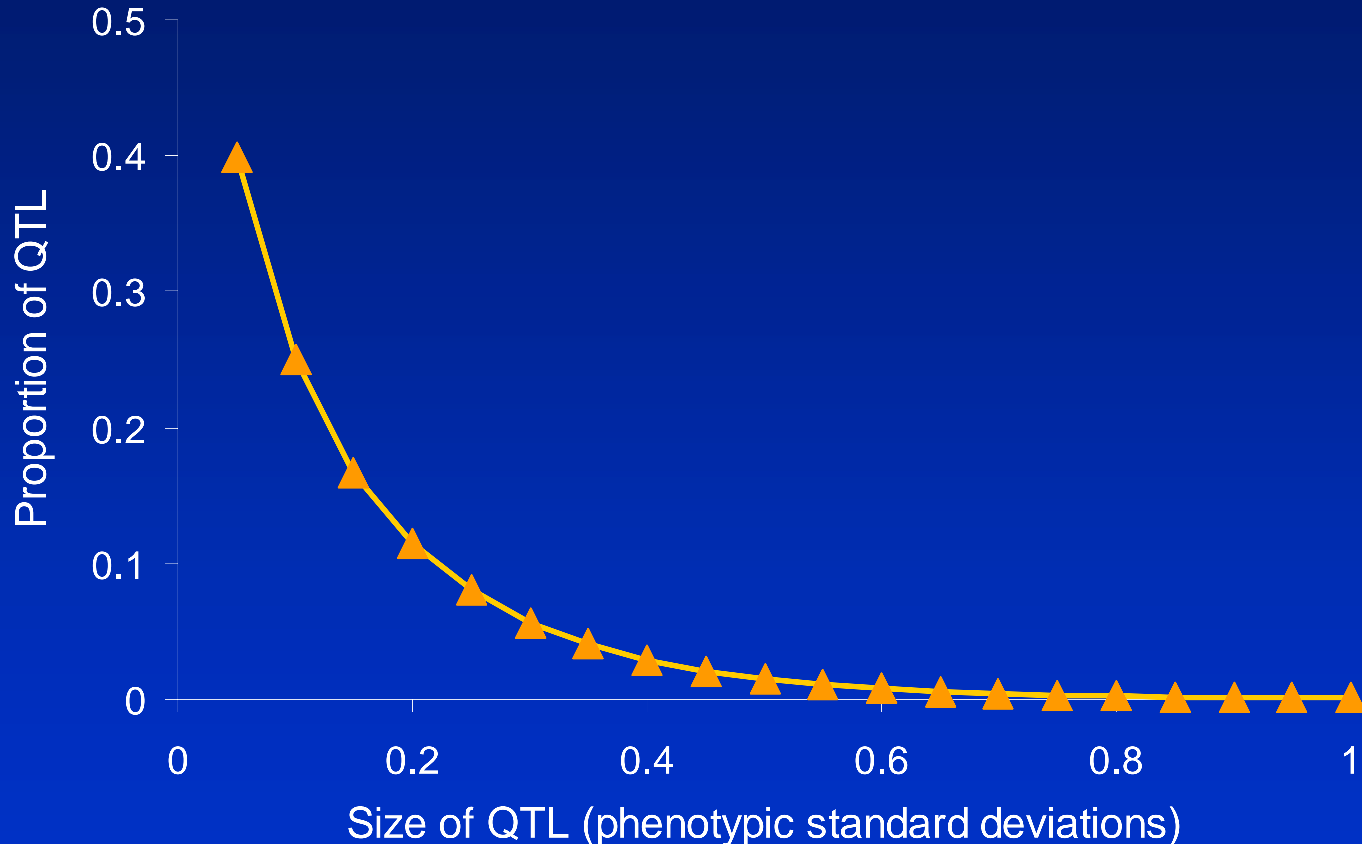
The distribution of QTL effects

- From results of QTL mapping experiments



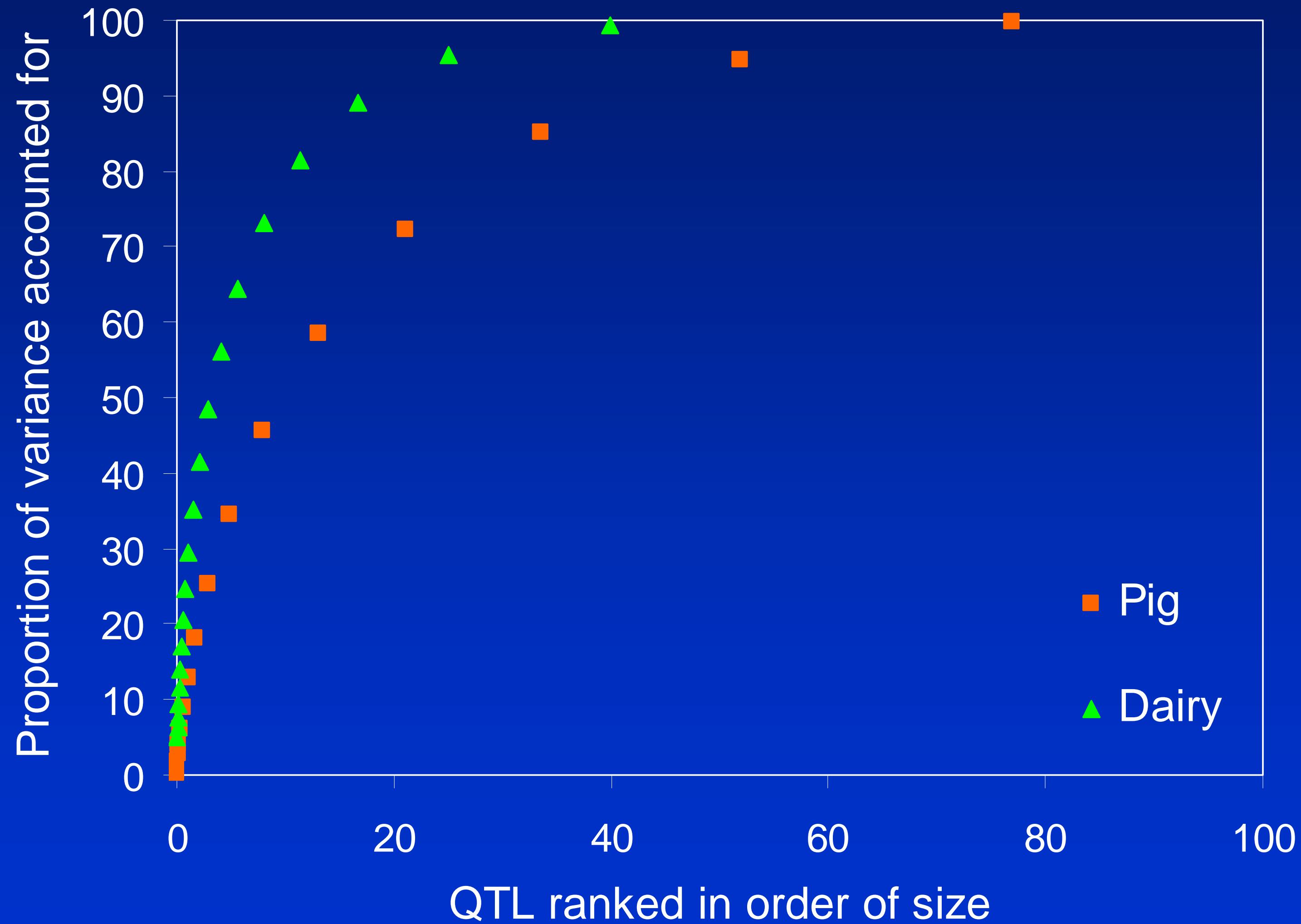
- Two problems
  - no small effects, effects estimated with error

The distribution of QTL effects



- Fit a gamma distribution
- Many small QTL, few QTL of large effect.

The distribution of QTL effects



The distribution of QTL effects

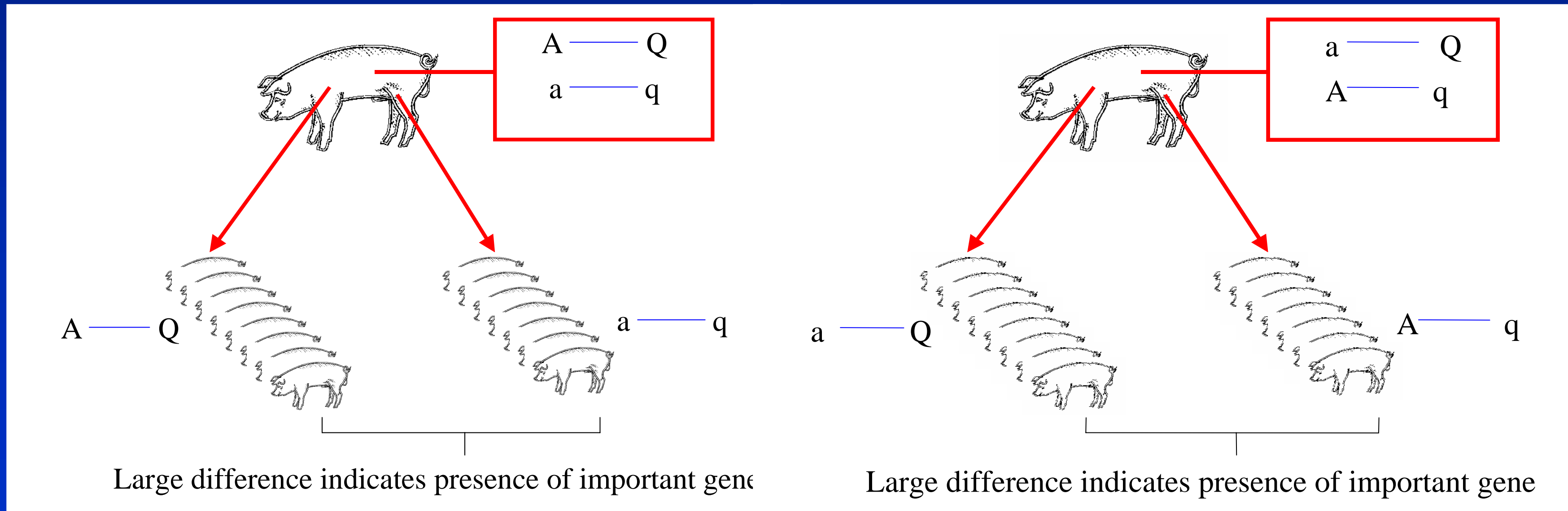
- Many small QTL, few of large effect, but...
- 5-10 large QTL explain the majority of the genetic variance
- Mapping experiment should be able to detect QTL as small as $0.2\sigma p$?

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

Optimising the number and size of half-sib families

- Half sib design…..



Optimising the number and size of half-sib families

- To detect a QTL, two criteria must be satisfied
  - Sire families must be large enough to distinguish the allele substitution effect from error (eg. detect QTL as small as $0.2\sigma p$)
  - at least one of the sires must be heterozygous for the QTL

Optimising the number and size of half-sib families

- To detect a QTL, two criteria must be satisfied
  - Sire families must be large enough to distinguish the allele substitution effect from error (eg. detect QTL as small as 0.2σp)

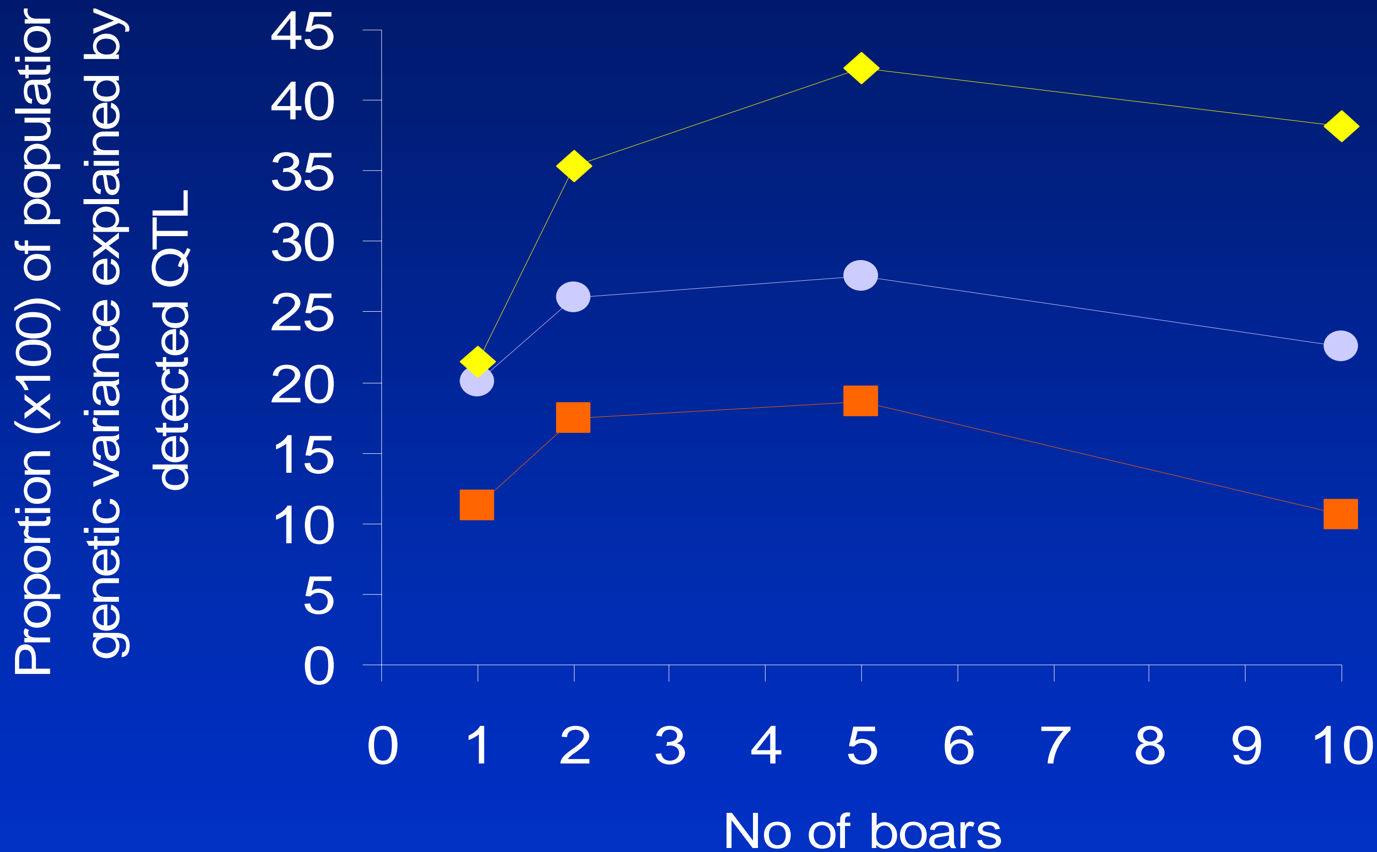
			Size of QTL effects[1]		
Sires	Progeny per sire	Total number of progeny	0.1	0.2	0.3
5	200	1000	0.03	0.18	0.50
	400	2000	0.07	0.44	0.80
	600	3000	0.12	0.64	0.90
	800	4000	0.18	0.76	0.94
	1000	5000	0.25	0.83	0.96
	2000	10000	0.55	0.95	0.97

[1] Residual standard deviations

Optimising the number and size of half-sib families

- To detect a QTL, two criteria must be satisfied
  - Sire families must be large enough to distinguish the allele substitution effect from error (eg. detect QTL as small as $0.2\sigma p$)
  - at least one of the sires must be heterozygous for the QTL
- If we are have maximum total number of progeny, is it better to have
  - many sires and small sire families, or
  - few sires and large sire families?
  - criteria: the proportion of total genetic variance explained by detected QTL

Optimising the number and size of half-sib families



Proportion of genetic variance explained by detected QTL in genome scans with 1, 2, 5 or 10 boars and 500 (■), 1000 (O) or 2000 (♦) total progeny allocated to the mapping experiment.

Optimising the number and size of half-sib families

- Large half sib families are necessary
- With a set number of total progeny, using five sires appears to balance
  - having enough sires such that at least one is heterozygous for the QTL
  - sufficiently large families to distinguish QTL from error
- Other strategies
  - choose phenotype which more accurately reflects genotype
    - granddaughter design
  - fewer progeny needed

Optimising the design of linkage experiments to detect QTL

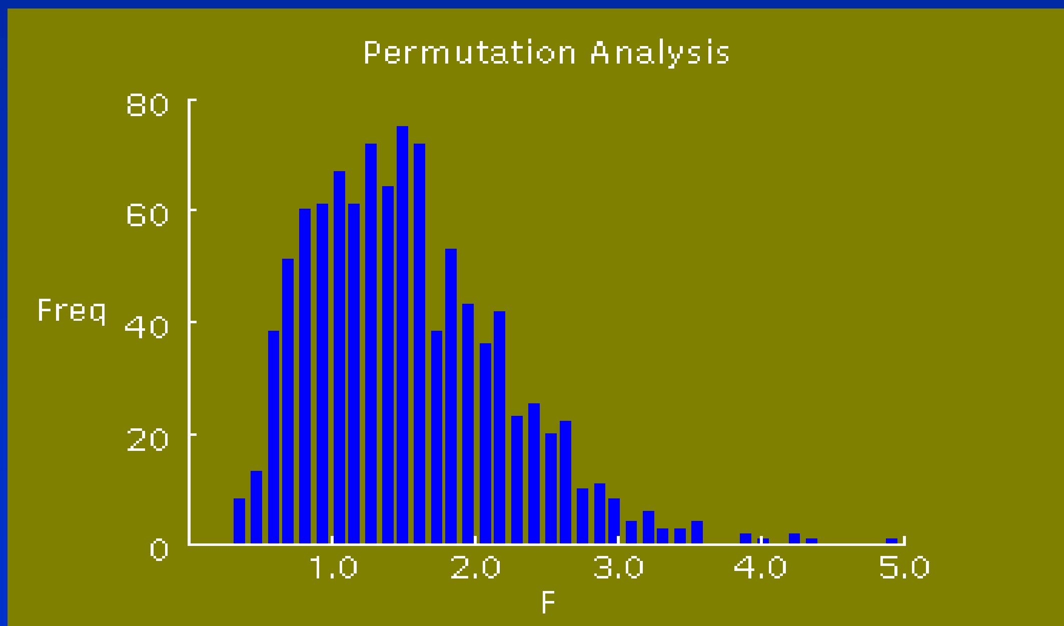
- Key parameters are:
  - distribution of QTL effects (how QTL are potentially detectable in a mapping experiment)
  - population structure
  - significance thresholds
  - precision of QTL mapping (width of confidence interval)
  - efficient genotyping strategies

Which significance threshold to use in a genome scan?

- Setting significance thresholds for QTL detection (multiple testing problem)
  - Many positions along genome analysed for presence of QTL
  - When these multiple tests are performed, ‘nominal’ significance levels of single test don’t correspond to the actual significance levels in whole experiment
  - Need more stringent thresholds
- Two approaches
  - Bonferoni correction
  - permutation testing (in half sib familes)

Which significance threshold to use in a genome scan?

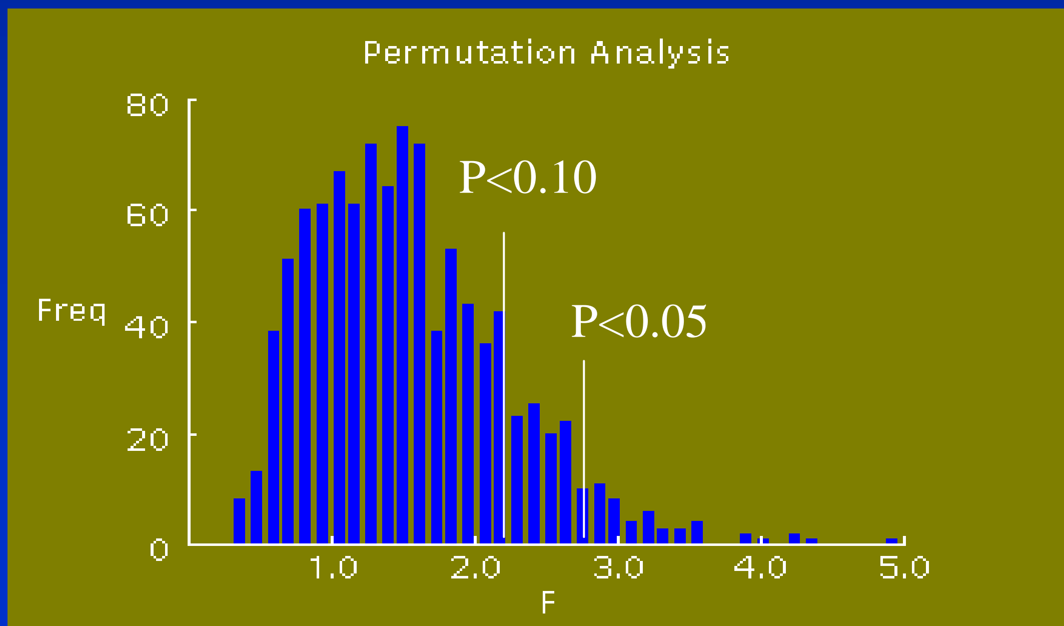
- Permutation testing
  - remove the link between phenotypes and genotypes
  - in practise, shuffle phenotypes across genotypes within half sib families, run QTL mapping experiment on new data
  - do this enough times to create distribution of the test statistic



-has the advantage of being an empirical distribution based on your data

Which significance threshold to use in a genome scan?

- Permutation testing
  - remove the link between phenotypes and genotypes
  - in practise, shuffle phenotypes across genotypes within half sib families, run QTL mapping experiment on new data
  - do this enough times to create distribution of the test statistic



-has the advantage of being an empirical distribution based on your data

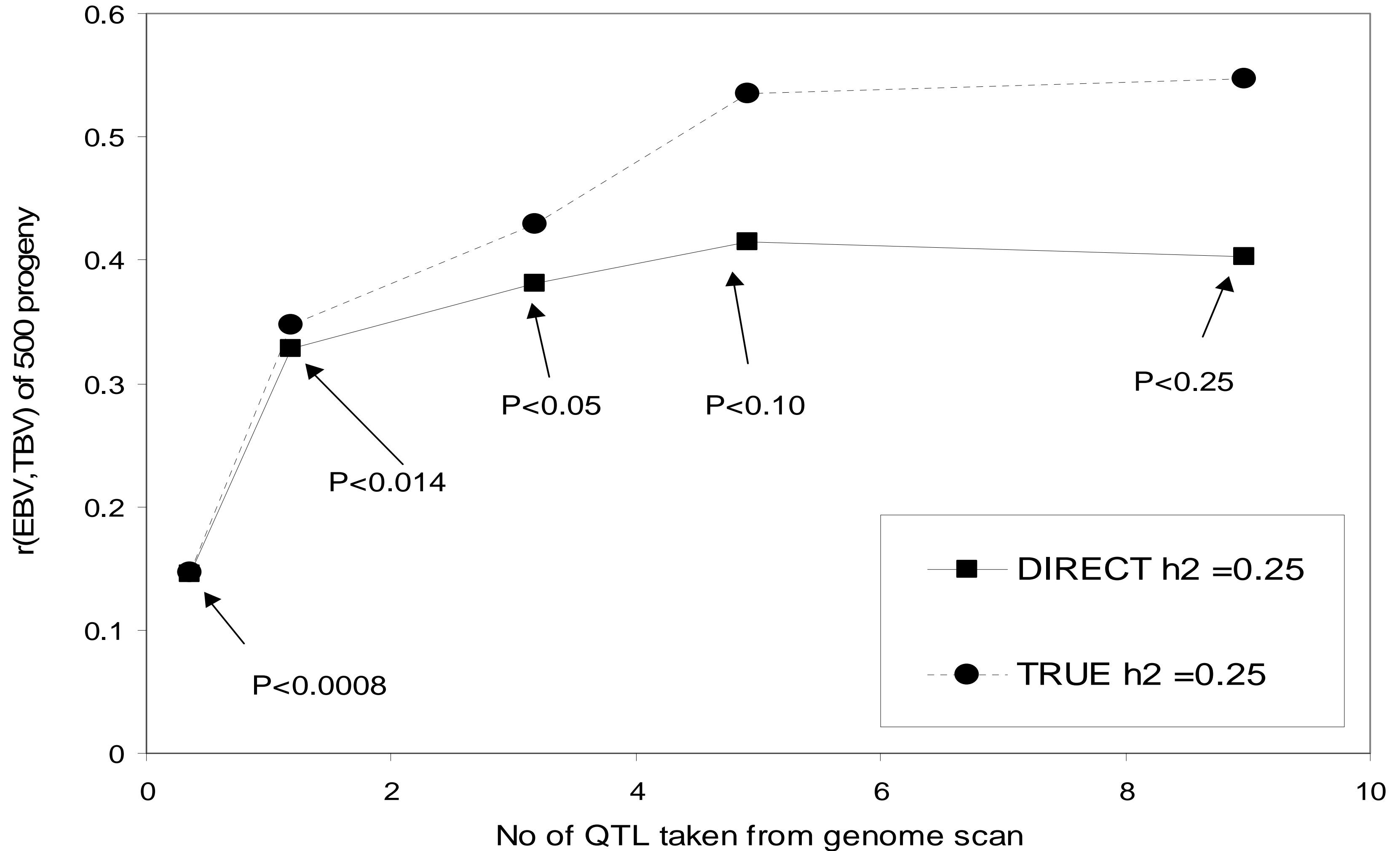
Which significance threshold to use in a genome scan?

- How many QTL to take from the genome scan to MAS?
- Can set different significance thresholds
- If we reduce the stringency of the significance threshold
  - detect more QTL
    - explain a greater proportion of the genetic variance
  - but more of these will be ‘false positives’
    - over-estimate the genetic variance
    - erode the advantage of subsequent MAS

Which significance threshold to use in a genome scan?

- A small experiment:
  1. simulate a population pigs with markers and QTL segregating
  2. Select a sire, breed 200 progeny from him
  3. Perform a QTL mapping experiment, with significance thresholds (permutation testing)
     - experiment wise (0.05 fp/genome scan) $P < 0.0008$ for single test
     - chromosome wise (0.05 fp/chromosome) $P < 0.014$ for single test
     - point wise 0.05, 0.10 and 0.25
  4. For each significance level, take the QTL detected and use them in MAS in a different group of 500 of the sires progeny
  5. Which threshold maximises the accuracy of subsequent MAS?

Which significance threshold to use in a genome scan?



Which significance threshold to use in a genome scan?

- False discovery rate (Weller 1998)
  - The expected proportion of QTL that exceed the significance threshold that are in fact false positives
  - $FDR = mP_{max}/n$
    - m = number of positions tested
    - P_{max} = largest P value of the significant QTL
    - n = number of QTL which exceed significance threshold

P value	QTL detected	False discovery rate (FDR)
0.0008	0.35	0.04
0.014	1.3	0.20
0.05	3.2	0.24
0.1	4.9	0.34
0.25	9	0.58

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

Precision of QTL mapping with linkage

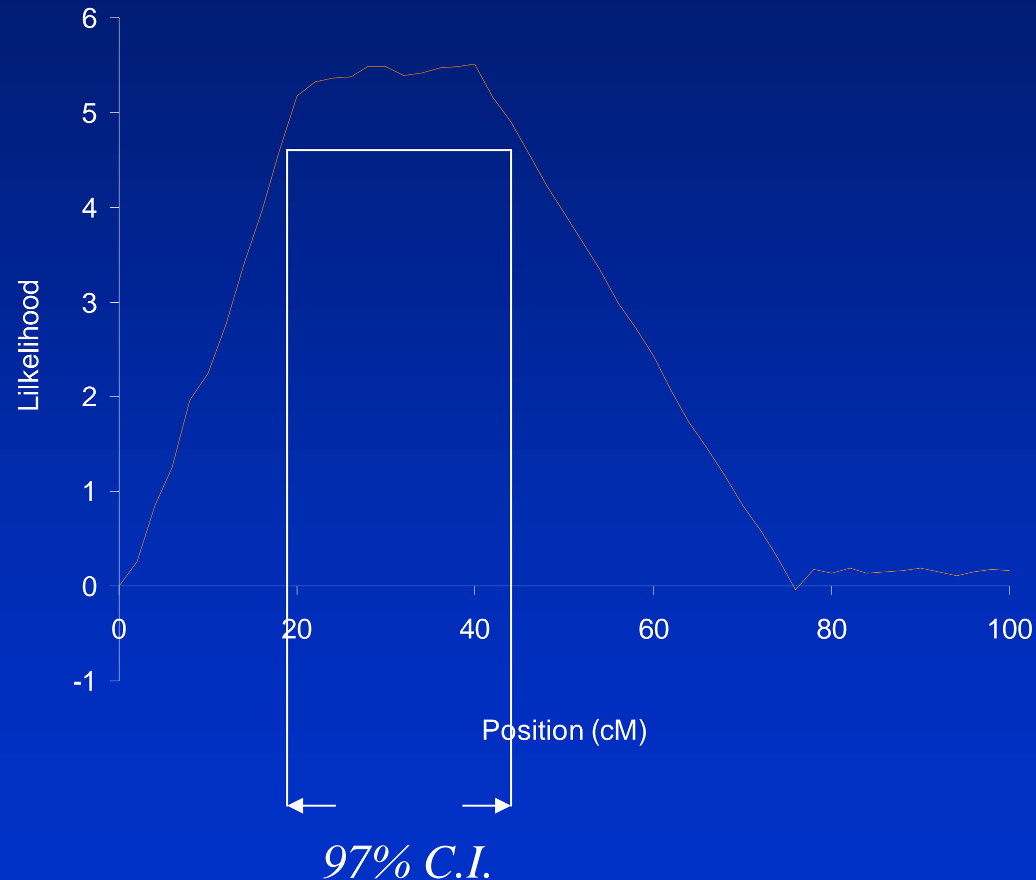
- Width of confidence interval determines
  - how many candidate genes must be investigated, or
  - size of chromosome segment to be saturated with dense markers for LD mapping
- Three approaches
  - deterministic prediction
  - Likelihood drop off
  - Boot-strapping

Precision of QTL mapping with linkage

- Deterministic prediction (Darvasi and Soller 1997)
  - 95%C.I. = $L/(kNa^2)$
    - L = length of genome
    - k = number of informative parents per individual (1 for half sib designs, 2 for F2 designs)
    - N = number of individuals genotyped
    - a = allele substitution effect of QTL (residual standard devs).
  - Eg. L=3000, k=1, N=1000, a =0.5, 95%C.I=12cM
  - assumes very dense markers

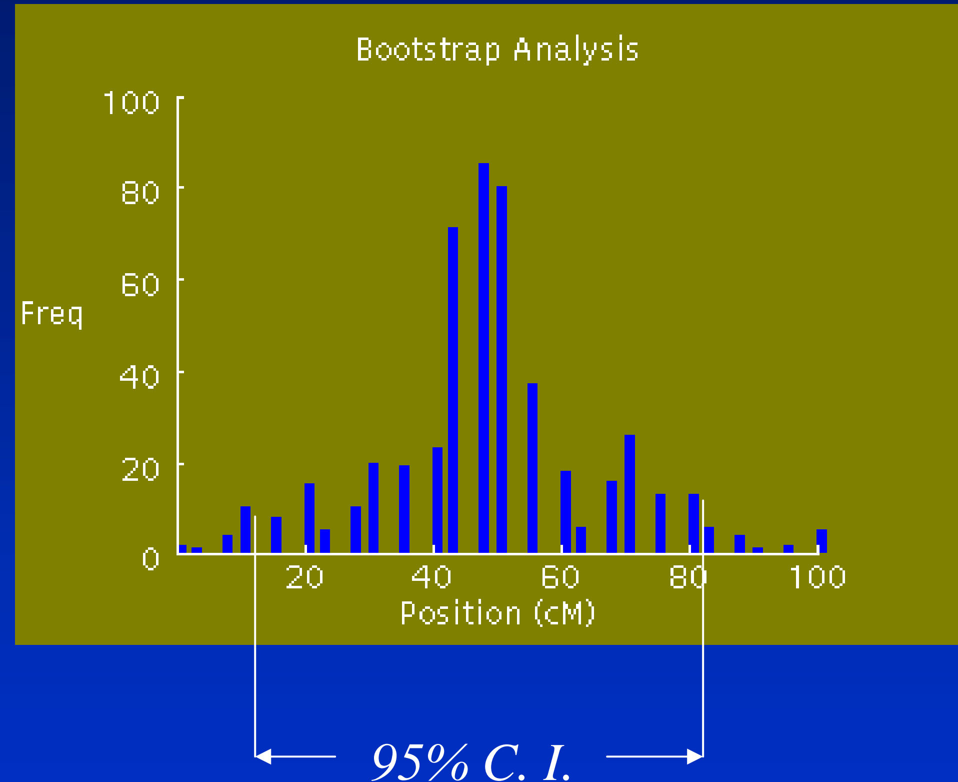
Precision of QTL mapping with linkage

- Likelihood drop off (Lander and Botstein (1989)
  - perform mapping experiment, location with the highest likelihood (LOD) is most likely putative QTL position
  - Calculate C.I. by moving sideways till LOD drops by one unit, width of interval = 96.8% C.I.



Precision of QTL mapping with linkage

- Boot-strapping (Visscher et al. 1996)
  - For data on N individuals, create boot-strap sample by sampling with replacement N individuals
  - For n bootstrap samples, perform QTL mapping experiment
  - Empirical 95% confidence intervals of QTL position are determined by ordering n estimates and taking 2.5 and 97.5 percentiles.



Precision of QTL mapping with linkage

- 95% Confidence intervals for QTL location from linkage mapping are very wide
  - often entire chromosome
  - enormous families required to generate sufficient recombinants to accurately position QTL
  - Alternatives?
  - increase power using efficient genotyping strategies
  - use historical recombinants (LD mapping)