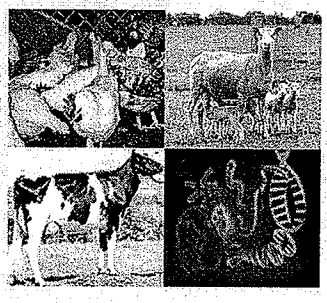
Quantitative trait loci (QTL) mapping



Sang Hong Lee

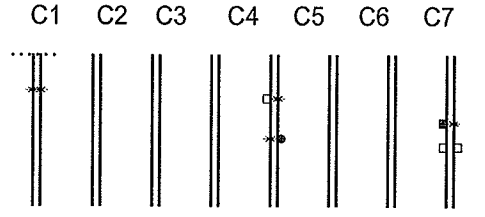
UNE

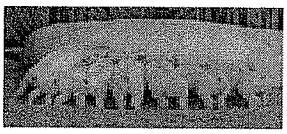
IAEA, Korea, April, 2006

What is QT and QTL?

Chromosome ->

C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 C11

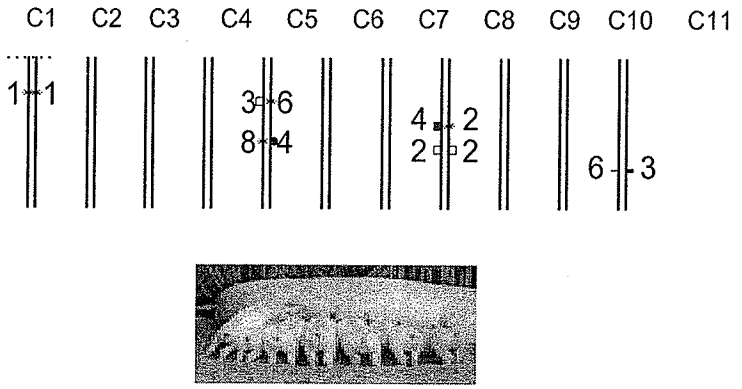




QT – phenotype influenced by multiple genes (litter size, weight etc.)
QTL – a gene affecting the phenotypic variation in continuously varying traits

QTL mapping

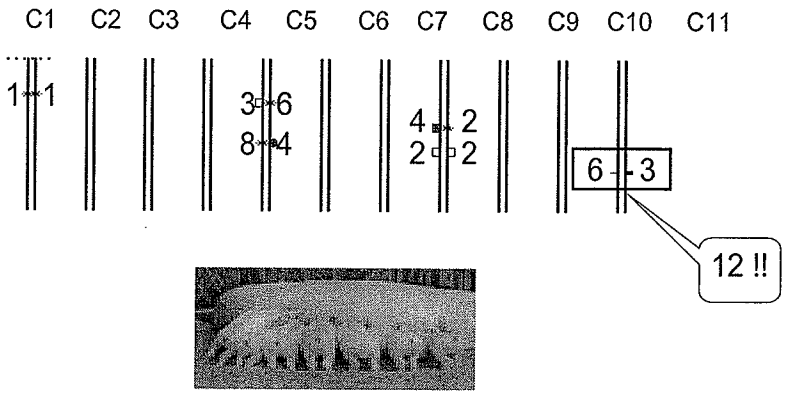
What is QT and QTL?



Each QTL has additive genetic effect

QTL mapping

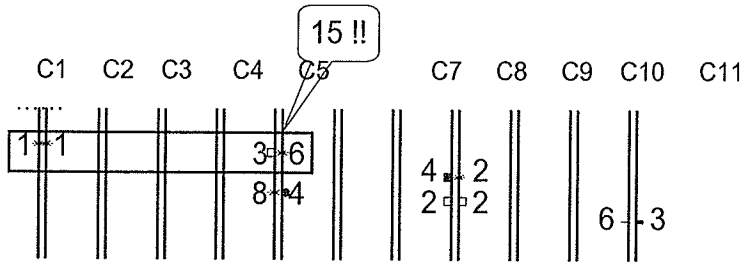
What is QT and QTL?

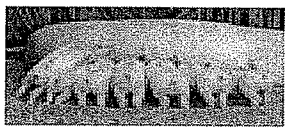


Intra-locus interaction (dom. or reces.) in a QTL

QTL mapping

What is QT and QTL?

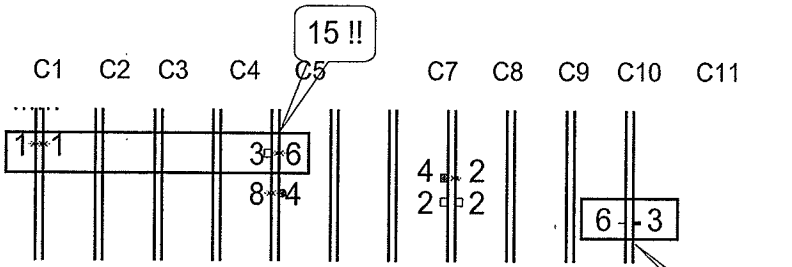


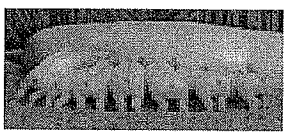


Inter-locus interaction (epistasis) between QTLs

QTL mapping

What is QT and QTL?



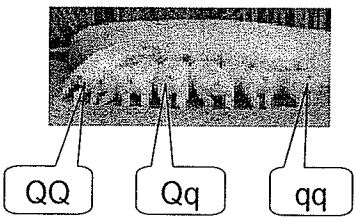


y = additive QTL effect + interaction within and between QTL + Env.
QT phenotype (y) is variable for each individual ~ N(0, Vp)

QTL mapping

What is QT and QTL?

- QTL is a part of factors determining phenotypes
  - Animals having similar QTL genotypes have similar phenotypes
- QTL genotypes are not directly observed
  - QTL genotype probability can be estimate using genetic marker
  - The probability is different across chromosome
  - Regression of phenotypes on QTL genotype probability > QTL position



QTL mapping

QTL genotype probability given a marker

Q q x q q
M m m m

If recombination rate between two loci = 0.1

Q q q q
M m m m

90 %

Q q q q
m m M m

10 %

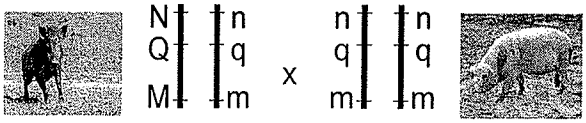
Prob. of having inherited paternal QTL alleles given the marker genotype,

Pr (QM) = Pr (qm) = 0.45, and Pr (Qm) = Pr (qM) = 0.05
Pr (Q|M) = 0.9 and Pr (Q|m) = 0.1

when using a single marker
hard to distinguish QTL effect and its position

QTL mapping

QTL gen. prob. given two flanking markers



If recombination rate between NQ=0.1, QM=0.3 and NM=0.34

Pr (NM) = Pr (nm) = (1-0.34)/2 = 0.33
Pr (Nm) = Pr (nM) = 0.34/2 = 0.17

Pr (NQM) = (1-0.1)(1-0.3)/2 = 0.315
Pr (NQm) = (1-0.1)(0.3)/2 = 0.135

Pr (nQm) = (0.1)(0.3)/2 = 0.015
Pr (nQM) = (0.1)(1-0.3)/2 = 0.035

Pr (Q|NM) = Pr(NQM)/Pr(NM) = 0.955
Pr (Q|Nm) = Pr(NQm)/Pr(Nm) = 0.795

Pr (Q|nm) = Pr(nQm)/Pr(nm) = 0.045
Pr (Q|nM) = Pr(nQM)/Pr(nM) = 0.205

QTL mapping

QTL mapping methods

Using phenotypes and QTL genotype prob.

- Regression method
  - regression of phenotypes on QTL genotype prob.
  - simple, faster
- Likelihood method
  - maximizing density function, i.e. pr(Q|y)
- Variance component method
  - random effects model (IBD probabilities etc.)
  - flexible for complex pedigree, multi allelic QTL
- Bayesian approach

QTL mapping

QTL mapping methods

Regression method

$$y = \mu + \alpha.x + e$$

phenotypes

QTL genotype probabilities

The residual sum of squares (SSE) can be obtained

x is different across chromosome

A region having the lowest SSE is the most likely position

$$approx.LR = n \ln\left(\frac{total\ sum\ of\ squares}{residual\ sum\ of\ squares}\right)$$

QTL mapping

QTL mapping methods

Maximum Likelihood method

$$y = \mu + \alpha.x + e$$

$$L(\mu_Q, \mu_q, \sigma \mid y) = \prod_{i=1}^{n}\left[\Pr(Q)\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(y_i-\mu_Q)^2}{2\sigma^2}} + \Pr(q)\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(y-\mu_q)^2}{2\sigma^2}}\right]$$

Normal distribution probability density function (pdf)

x is different across chromosome

A region having the highest pdf is the most likely position

$$LR = -2\ln\left(\frac{Likelihood_{\ reduced}}{Likelihood_{\ full}}\right)$$

QTL mapping

QTL mapping methods

Variance component approach

β: Fixed effect
u: Polygenic effects
q: QTL effects
A: NRM
G: GRM

$$y = X\beta + Z_1 u + Z_2 q + e$$

$$V = Z_1 A Z_1' \sigma_u^2 + Z_2 G Z_2' \sigma_q^2 + R$$

$$\log L(y \mid Xb, \sigma_q^2, \sigma_u^2, \sigma_e^2) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|V| - \frac{1}{2}(y - Xb)'V^{-1}(y - Xb)$$

G is different across chromosome

Region having the highest log *L* is the most likely position

$$LR = -2\left(\frac{\log L_{reduced}}{\log L_{full}}\right)$$

QTL mapping

Design of QTL mapping

Design establishes an ideal pattern of marker and QTL genotypes to maximize mapping efficiency

Experimental population

- It is possible to obtain pure inbred lines
- Backcross (backcrossing F1 to one of parental lines)
- F2 (progeny from F1 cross)

  - Gamete frequencies and progeny genotypes at markers and QTL can be systematically obtained in backcross or F2 design
  - Possible to estimate additive and non-additive QTL effects as well (intra-locus and inter-locus interaction model)

QTL mapping

Design of QTL mapping

Design establishes an ideal pattern of marker and QTL genotypes to maximize mapping efficiency

Outbred population

- It is difficult to obtain inbred lines
- Half sib design is widely used
- Grand daughter design reduce residuals
  - Dominance or recessive is not estimable in half sib designs
- General pedigree
  - Complex pedigree can give extra information
  - It can be possible to estimate dominance

Alternative models

Multiple QTL model

Implementing other effects

- Dominance
- Recessive
- Epistasis

Multi trait QTL model

Test statistics

Significance threshold
- Permutation test
- Bayesian factor

Confidence interval
- Bootstrapping
- Posterior QTL density

Marker information content
- Degree of marker information

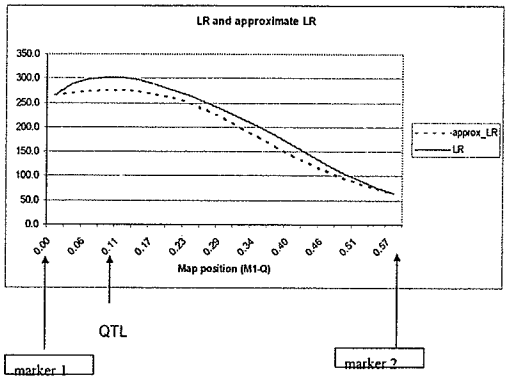
Other issues

High genotyping cost
- Selective genotyping
- DNA pooling

High false discovery rate (> 5% of number of tests)
- Increase significance level

Genotyping and pedigree error
- Population should be newly designed

Example (QTLdet.xls)

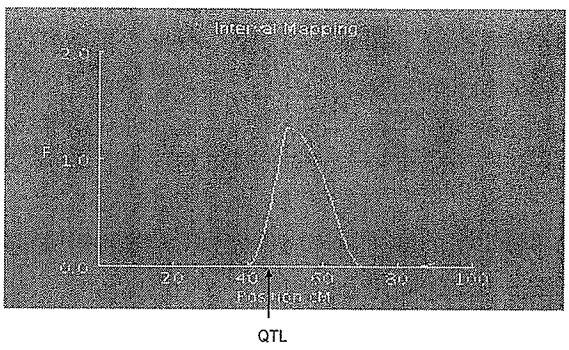


One half sib family (n=100)

Regression method (approx. LR) is equivalent to likelihood method (LR)

QTL mapping

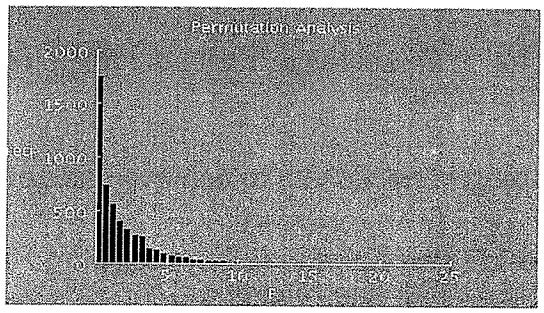
Example (QTLexpress)



10 half sib families each with 10 progeny

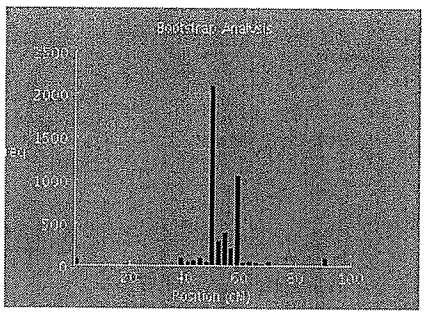
QTL mapping

Example (5000 Permutation)



Significance threshold: 6.33 (P = 0.05)
10.23 (P = 0.01)

Example (5000 Bootstrapping)



When analysis is based on linkage information, confidence interval of QTL region is wide (~30 cM)

Fine mapping of QTL

Ultimate goal in Quantitative Trait Loci mapping is to find actual genes

Fine mapping is necessary to achieve this goal

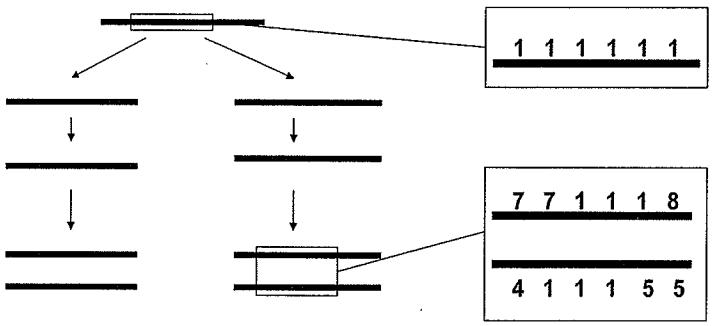
Combined LD and linkage (LDL) mapping is a promising tool for fine-mapping

* confidential interval of positioning QTL < 1cM

(Meuwissen and Goddard, 2000; 2002, Fanir et al.,2002; Grisart et al., 2002; Lee and Van der werf, 2004)

QTL mapping

Why LD ?



Recombination fraction (c) can be very small

c = 1 / 2 * no.generations

QTL mapping

And, why Linkage (pedigree information)?

To reconstruct haplotypes for genotyped animals

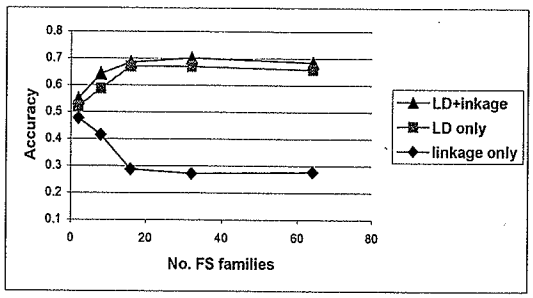
To estimate correlation between two haplotypes being transmitted from a common ancestor within pedigree i.e. identity by descent (IBD) prob.

To tract information of recombination within pedigree

- Information comes from only within recorded pedigree

QTL mapping

Designs for fine mapping of QTL



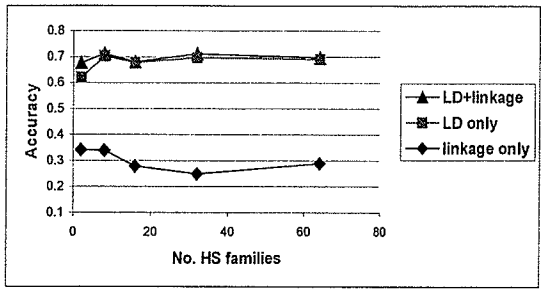
Best design for full sib families:

64 families each with 2 individuals in LDL mapping

2 families each with 64 individuals in linkage mapping

QTL mapping

Designs for fine mapping of QTL



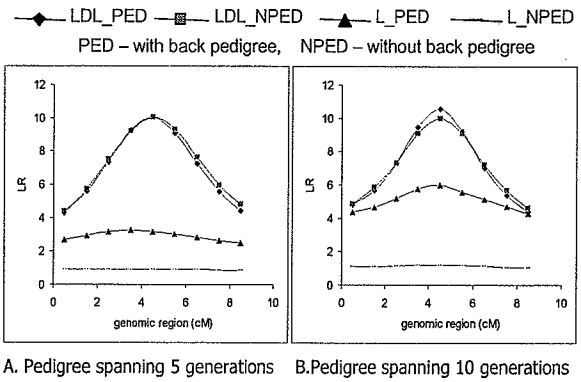
Best design for half sib families:

Not much different across designs in LDL mapping
- Common half sib designs can be used for fine-mapping

2~4 families each with 32~64 individuals in linkage mapping

QTL mapping

Use of back pedigree information



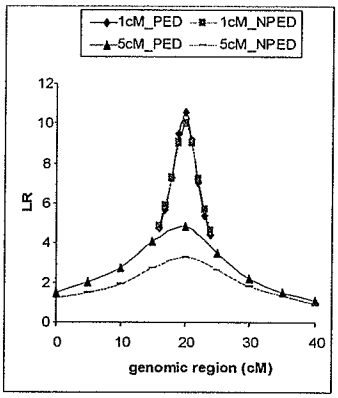
A. Pedigree spanning 5 generations B.Pedigree spanning 10 generations

LDL mapping perform much better than linkage mapping

Back pedigree information is important in linkage mapping, however is not in LDL mapping

QTL mapping

Levels of LD due to marker density



With Ne = 100,

Marker spacing of 1 cM
LD = 0.2
PED = NPED

Marker spacing of 5 cM
LD = 0.05
PED > NPED

Back pedigree information is important when LD is low, however is not when LD is high

QTL mapping

Practical session

We will try

- Simulating pedigree and genotypic and phenotypic data
- Estimating QTL position and effects

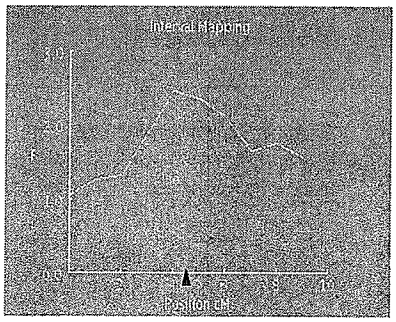
QTL mapping practice
- with a half sib family with two flanking markers
- with many half sib families with many markers
- with complex pedigree

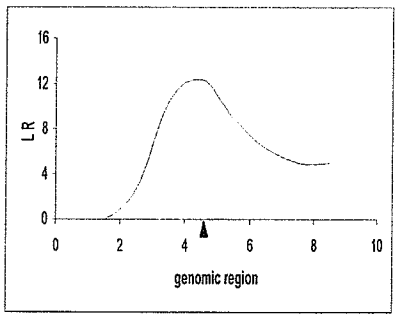
Compare LDL mapping and linkage mapping with varying marker density

Obtain significance threshold and confidence interval using
- Permutation test
- Bootstrapping
- Bayesian approach

QTL mapping

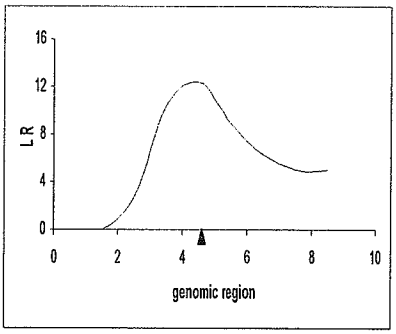
QTLexpress vs LDL

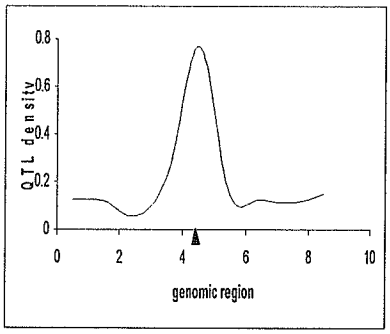




Power is higher in LDL mapping than linkage mapping

Likelihood vs Bayesian





Bayesian approach is more precise than likelihood approach
The Bayesian is also useful for multiple QTL model