



From Linkage to Gene Detection

Ben Hayes and Mike Goddard

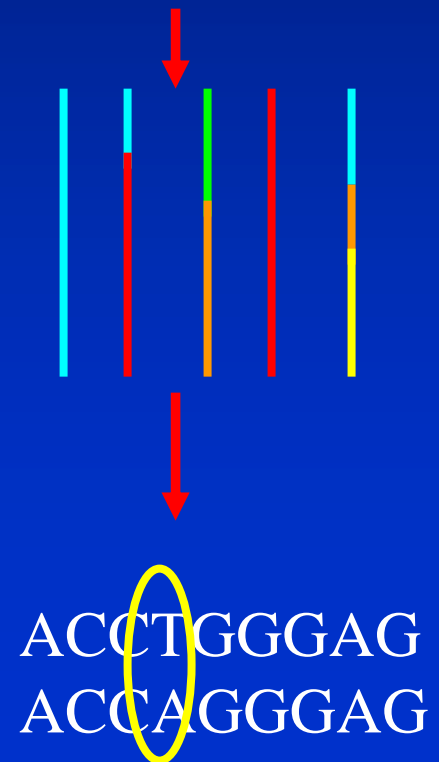
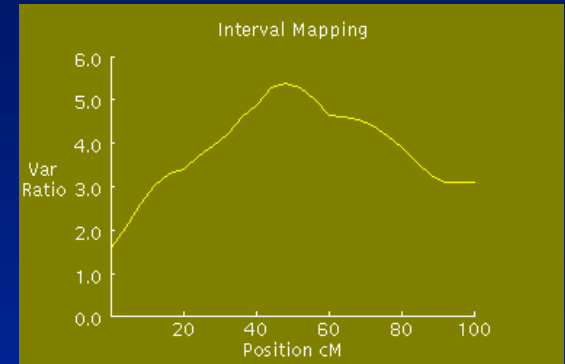


DGAT1 - A success story (Grisart et al. 2002)

1. Linkage mapping detects a QTL on bovine chromosome 14 with large effect on fat % (Georges et al 1995)

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 alanine. Gene patented. (Grisart et al. 2002)



Linkage disequilibrium (LD) mapping

- Definitions of LD
- Why does LD occur?
- Extent of LD in humans and livestock
- What type of markers are appropriate for detecting LD?
- LD mapping
 - building IBD matrix from marker haplotypes
 - variance component approach
- Combined LD-LA mapping

Linkage disequilibrium (LD) mapping

- Definitions of LD
- Why does LD occur?
- Extent of LD in humans and livestock
- What type of markers are appropriate for detecting LD?
- LD mapping
 - building IBD matrix from marker haplotypes
 - variance component approach
- Combined LD-LA mapping

Definitions of LD

- Classical definition:
 - Two markers A and B on the same chromosome
 - Alleles are
 - marker A A1, A2
 - marker B B1, B2
 - Possible haplotypes are A1_B1, A2_B1, A2_B1, A2_B2
 - if frequencies of alleles all =0.5, expect frequencies of each haplotypes to be 0.25
 - any departure from 0.25 is linkage disequilibrium, ie *genes not in random association.*

Definitions of LD

- In fact, LD is required for both linkage and linkage disequilibrium mapping
- Difference is
 - linkage analysis of QTL only considers the LD that exists within **families**
 - extends for 10s of cM
 - broken down after only a few generations
 - linkage disequilibrium analysis of QTL requires a marker allele or alleles of multiple markers to be in LD with a QTL allele across the whole **population**
 - association must have persisted across multiple generations to be a property of the population
 - so marker and QTL must be very closely linked

Definitions of LD

- Measuring the extent of LD (determines how dense markers need to be for LD mapping)

$$D = \text{freq}(A1_B1) * \text{freq}(A2_B2) - \text{freq}(A1_B2) * \text{freq}(A2_B1)$$

– highly dependent on allele frequencies

- not suitable for comparing LD at different sites

$$r^2 = D^2 / [\text{freq}(A1) * \text{freq}(A2) * \text{freq}(B1) * \text{freq}(B2)]$$

– very widely used, and equivalent for loci with multiple alleles exist

– But, only considers two loci at a time

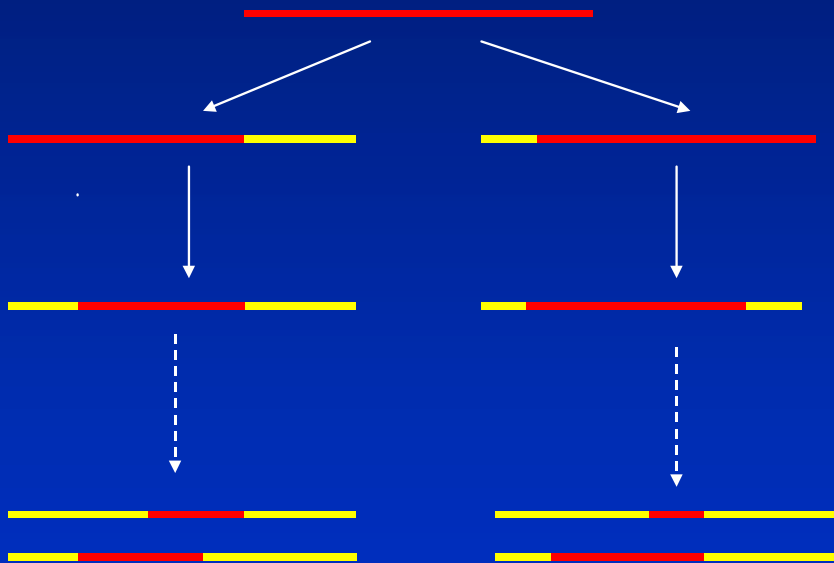
- cannot extract LD information available from multiple loci
- does not reflect linear nature of chromosomes (eg. recombination)
- not particularly intuitive with regards to the causes of LD

Definitions of LD

- Measures of LD
 - Chromosome segment homozygosity (CSH)
 - Multi locus measure of LD (Hayes et al. 2003)

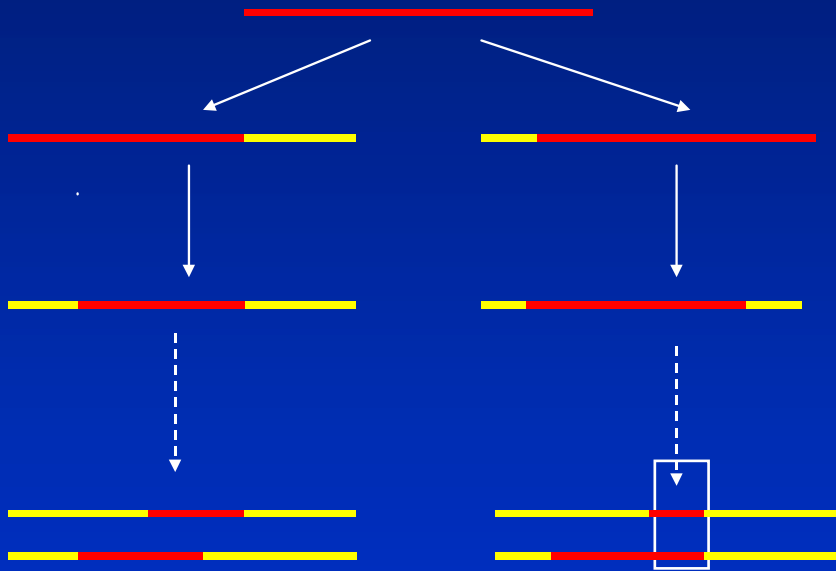
Definitions of LD

- A chunk of ancestral chromosome is conserved in the current population



Definitions of LD

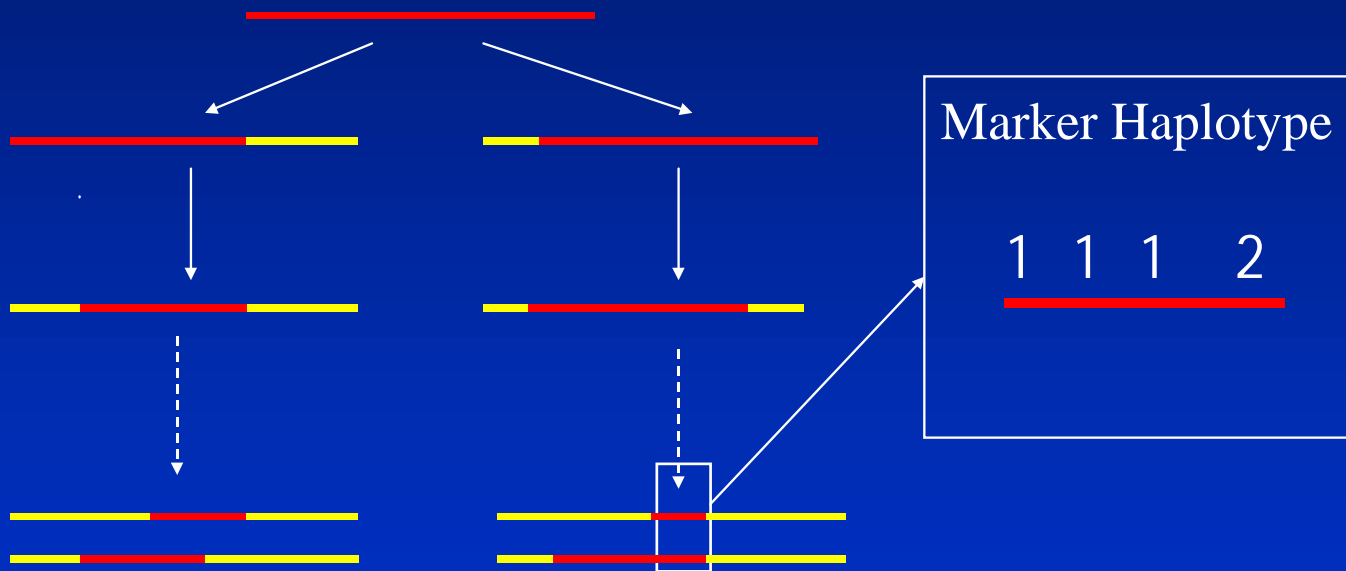
- A chunk of ancestral chromosome is conserved in the current population



- Chromosome segment homozygosity (CSH) = $\Pr(\text{Two chromosome segments randomly drawn from the population are derived from a common ancestor})$

Definitions of LD

- A chunk of ancestral chromosome is conserved in the current population



- Chromosome segment homozygosity (CSH) = $\Pr(\text{Two chromosome segments randomly drawn from the population are derived from a common ancestor})$

Definitions of LD

- Haplotype homozygosity = $CSH + \text{Identical chance}$
(and not IBD)
- For two loci
$$HH = CSH + (Hom_A - CSH)(Hom_B - CSH)/(1 - CSH)$$
- Derivation for multiple loci similar, but more complex

Linkage disequilibrium (LD) mapping

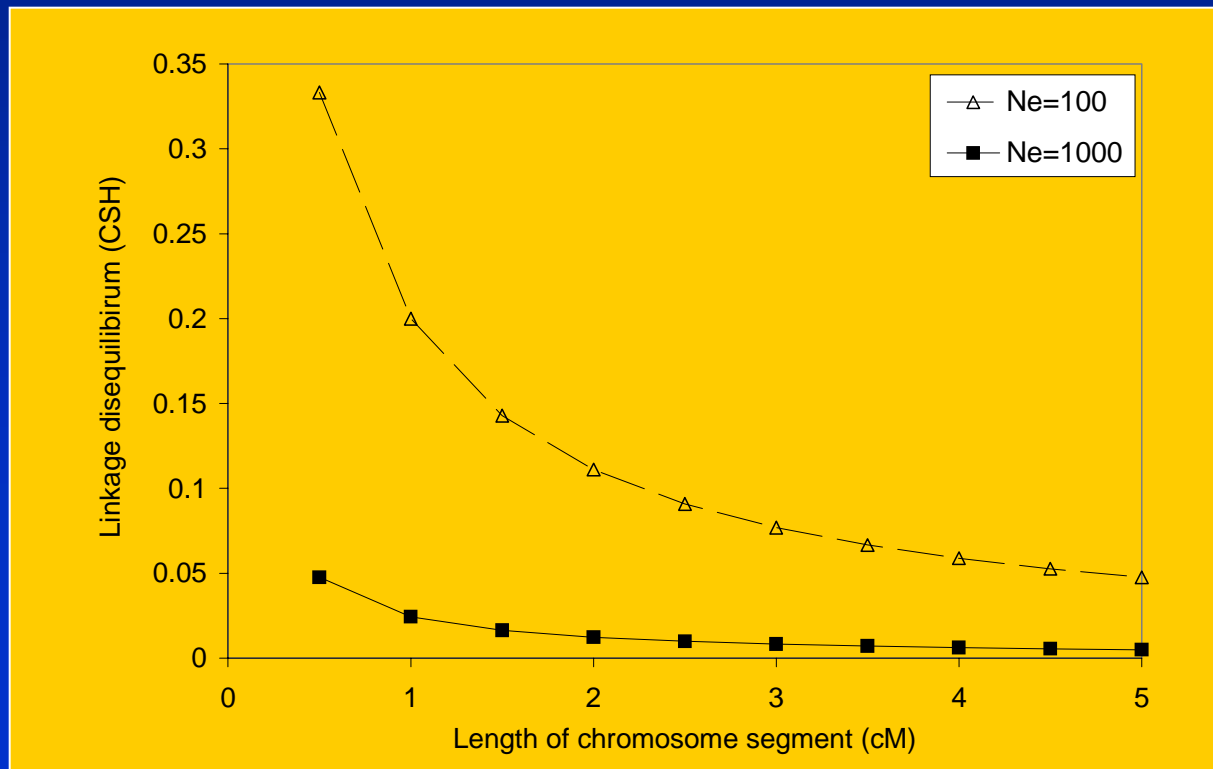
- Definitions of LD
- Why does LD occur?
- Extent of LD in humans and livestock
- What type of markers are appropriate for detecting LD?
- LD mapping
 - building IBD matrix from marker haplotypes
 - variance component approach
- Combined LD-LA mapping

Causes of LD

- Migration
 - LD artificially created in F2 designs
- Mutation
- Selection
- Small finite population size
 - generally implicated as the key cause of LD in livestock populations, where effective population size is small
 - LD due to crossbreeding (migration) is large when crossing inbred lines but small when crossing breeds that do not differ markedly in gene frequencies
 - disappears after only a limited number of generations

Causes of LD

- Predicting extent of LD with finite population size
- $E(r^2)$ and $E(\text{CSH}) = 1/(4Nc+1)$
 - N = effective population size
 - c = length of chromosome segment



Linkage disequilibrium (LD) mapping

- Definitions of LD
- Why does LD occur?
- **Extent of LD in humans and livestock**
- What type of markers are appropriate for detecting LD?
- LD mapping
 - building IBD matrix from marker haplotypes
 - variance component approach
- Combined LD-LA mapping

Extent of LD in humans and livestock

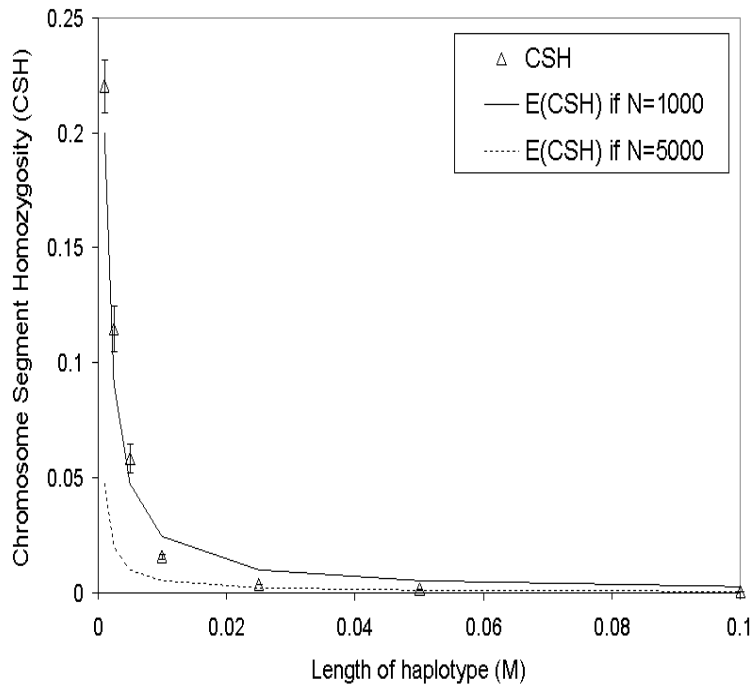
- Effective population size livestock ~ 100
- Effective population size humans ~ 10000
- If accept finite population size as cause of LD, LD livestock \gggg LD humans
- This is what is observed
 - Significant LD in humans $\sim 5\text{kb}$ (0.005cM)
 - depends on population
 - Significant LD in livestock $\sim 0.5\text{cM} - 10\text{cM}$
 - cattle and sheep

Extent of LD in humans and livestock

- Refining the pattern of LD
 - LD depends on both recent and historical recombinations
 - Both recent and historical population size will effect extent of LD
 - In livestock, current pop size \ll historical pop size
 - In humans, current pop size \gggg historical pop size
 - How does changing pop size affect the pattern of LD?
 - Simulations with increasing or decreasing pop size

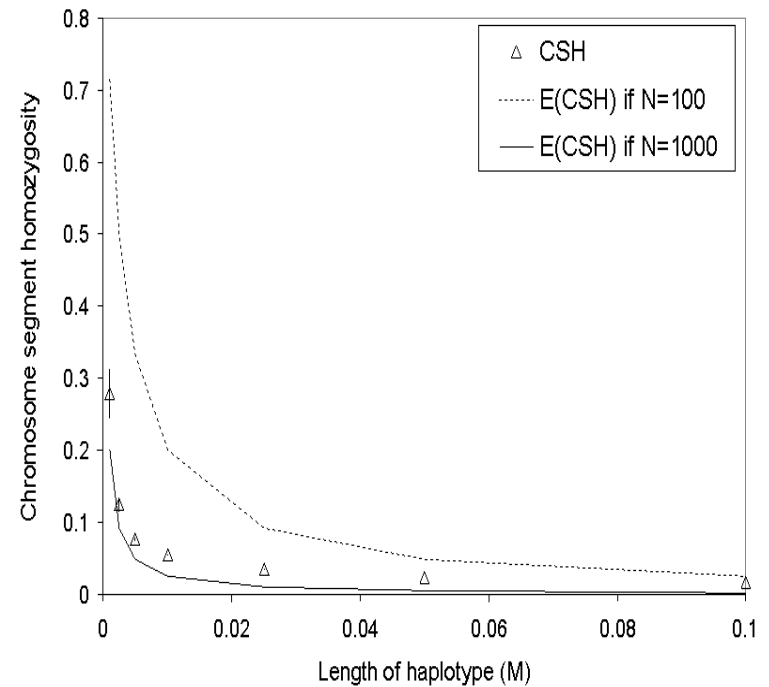
Extent of LD in humans and livestock

1000 to 5000



A

1000 to 100



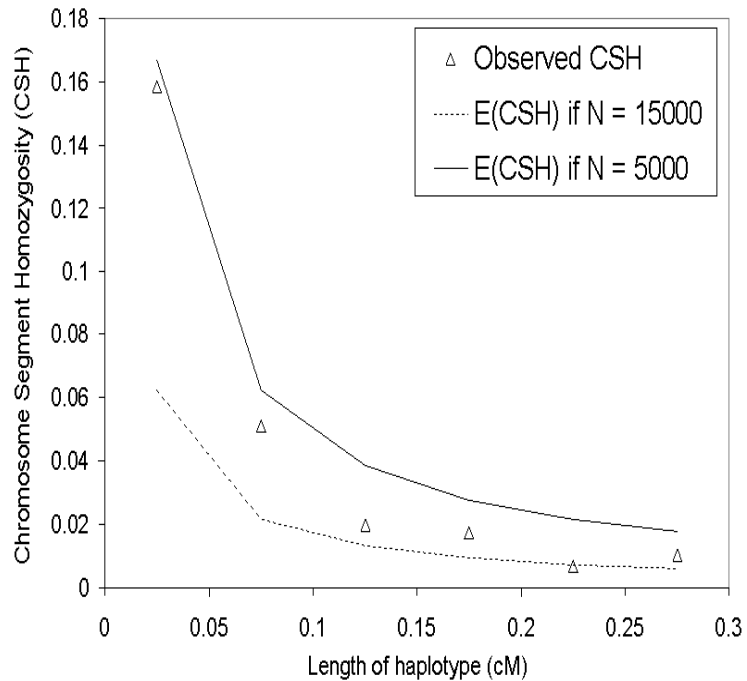
B

Extent of LD in humans and livestock

- Conclusion: LD at short distances depends on historical population size, LD at long distances reflects recent population size
- $E(LD) = 1/(4N_t c + 1)$
 - $t = 1/(2c)$ generations ago

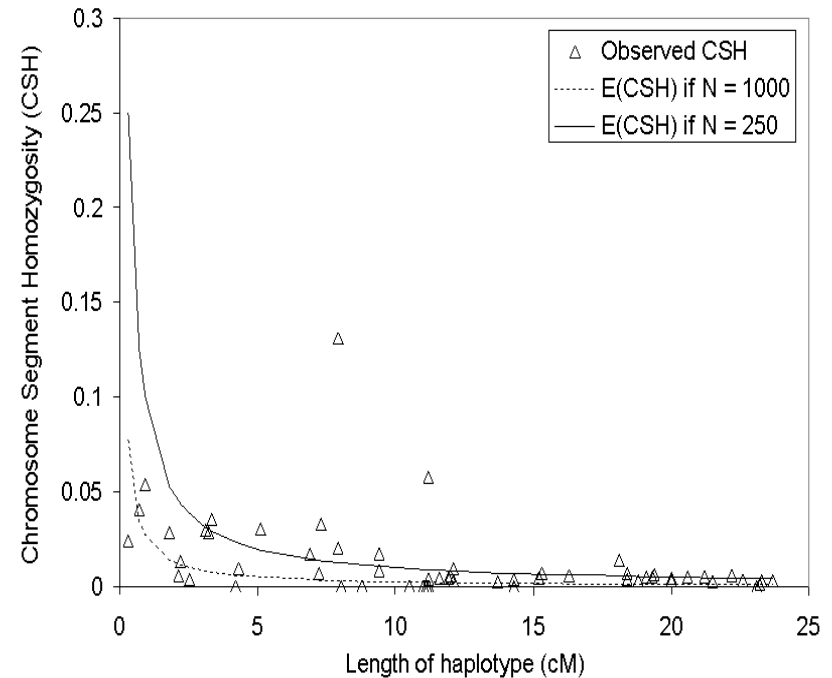
Extent of LD in humans and livestock

Humans



A

Holstein Friesians



B

Linkage disequilibrium (LD) mapping

- Definitions of LD
- Why does LD occur?
- Extent of LD in humans and livestock
- What type of markers are appropriate for detecting LD?
- LD mapping
 - building IBD matrix from marker haplotypes
 - variance component approach
- Combined LD-LA mapping

Markers for detecting LD

- Need to be very dense in humans, 1 marker ~ 0.005cM
- Livestock ~ 0.5-2 cM
 - (therefore lower precision if QTL detected)
- In linkage analysis, use microsatellites
 - highly polymorphic, very informative
 - are they dense enough for LD mapping?
 - 2141 microsats on the cattle map
 - average marker spacing 1.4cM
 - will have areas of higher/lower density

Markers for detecting LD

- Will probably need denser markers
- Alternative to microsats is single nucleotide polymorphisms (SNPs)
ACCCCTTG
ACCTTTG
- density in humans 1/kilobase (0.001cM)
- not as highly polymorphic, about 5 SNPs = 1 microsat
- have to find them yourself
- but can be functional mutations
- data set can contain both SNPs and microsats

Linkage disequilibrium (LD) mapping

- Definitions of LD
- Why does LD occur?
- Extent of LD in humans and livestock
- What type of markers are appropriate for detecting LD?
- LD mapping
 - building IBD matrix from marker haplotypes
 - variance component approach
- Combined LD-LA mapping

LD mapping

- Principle:
 - Existence of LD implies small segments of chromosome in population which are descended from the same common ancestor (IBD).
 - IBD chromosome segments will not only carry identical marker haplotypes; if there is a QTL within chromosome segment, IBD chromosome segments will also carry identical QTL alleles.
 - If two animals carry chromosomes which are likely to be IBD at a point on the chromosome carrying a QTL, then their phenotypes will be correlated.

LD mapping

- Building IBD matrix from marker haplotypes
 - Calculate the probability 2 chromosomes are IBD at putative QTL position based on marker haplotypes, store these probabilities in an IBD matrix (**G**).
 - If the correlation between animal's phenotypes is proportional to **G** there is evidence for a QTL at this position.
 - Sort genotype data into haplotypes first

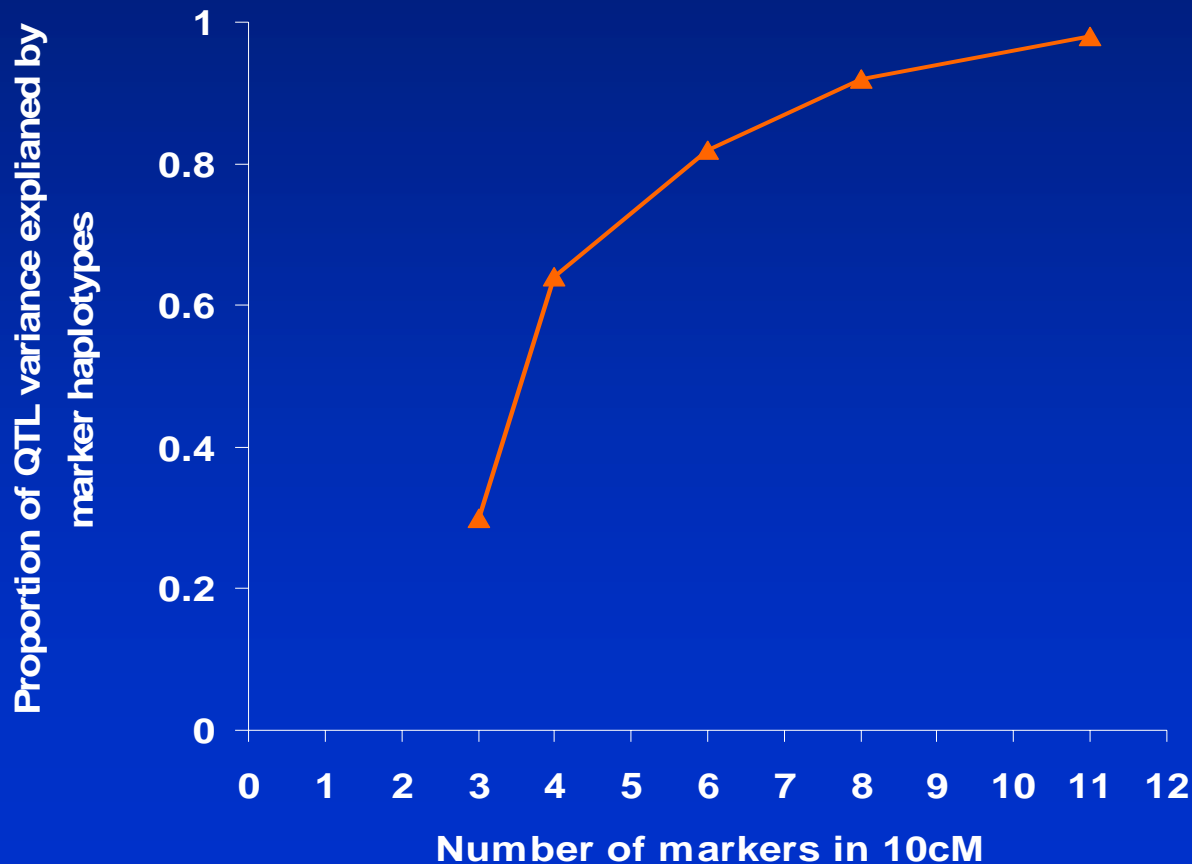
LD mapping

- Building IBD matrix from marker haplotypes
 - Consider three haplotypes drawn from population at random (P is putative QTL position)
 - A 112P112
 - B 212P112
 - C 222P222
 - $P(\text{IBD at QTL A,B}) > P(\text{IBD at QTL B,C})$, as longer identical haplotype

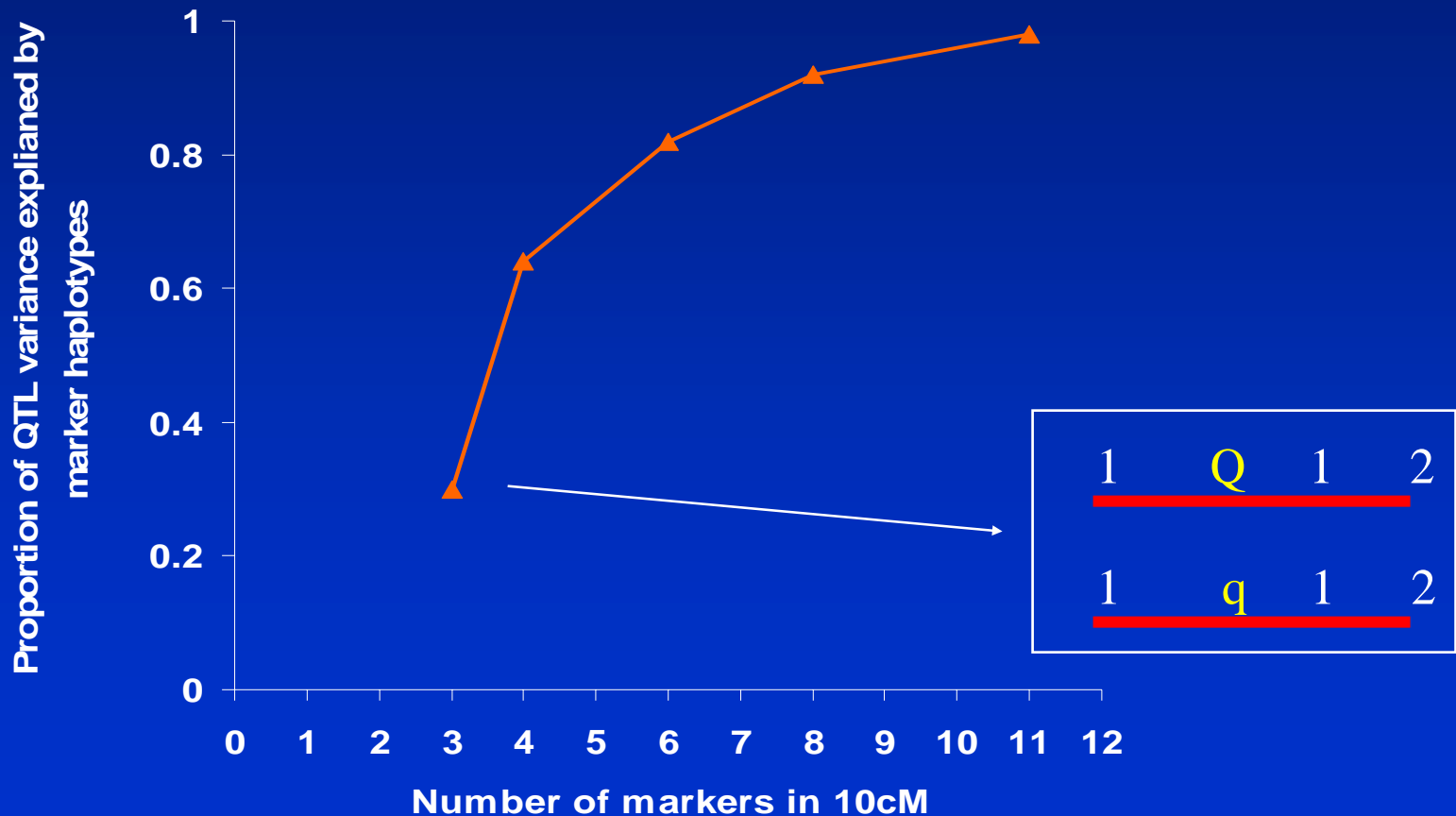
LD mapping

- Building IBD matrix from marker haplotypes
 - Parameters which determine IBD coefficients are effective population size, length of haplotype and number of markers in the haplotype

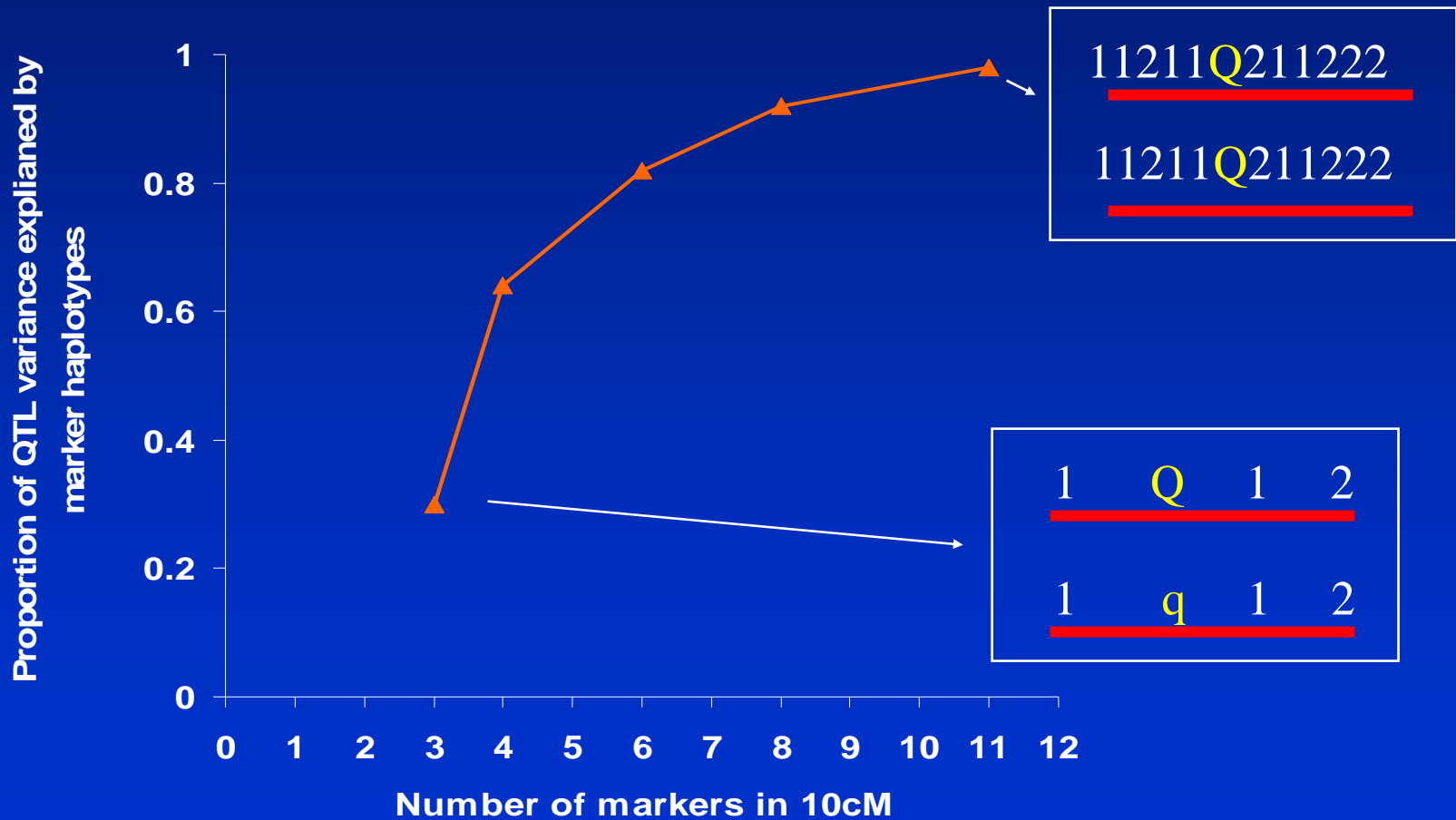
Proportion of QTL variance explained by surrounding markers



Proportion of QTL variance explained by surrounding markers



Proportion of QTL variance explained by surrounding markers



LD mapping

- Building IBD matrix from marker haplotypes
- Algorithm of Meuwissen and Goddard (2001)
 - deterministically predicts IBD coefficients between two marker haplotypes
 - takes into account
 - number of markers flanking QTL position which are identical by state
 - probability identical by chance \sim marker homozygosity
 - extent of LD based on length of haplotype, effective population size

LD mapping

- Building IBD matrix from marker haplotypes
 - An example
 - 6 markers in 10cM, putative QTL position in centre
 - Sample four haplotypes from the population
 - 112112, 112112, 122112, 222122
 - IBD matrix is:

	112112	112112	122112	222122
112112	1			
112112	0.82	1		
122112	0.63	0.63	1	
222122	0.49	0.49	0.56	1

LD mapping

- Variance component model

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

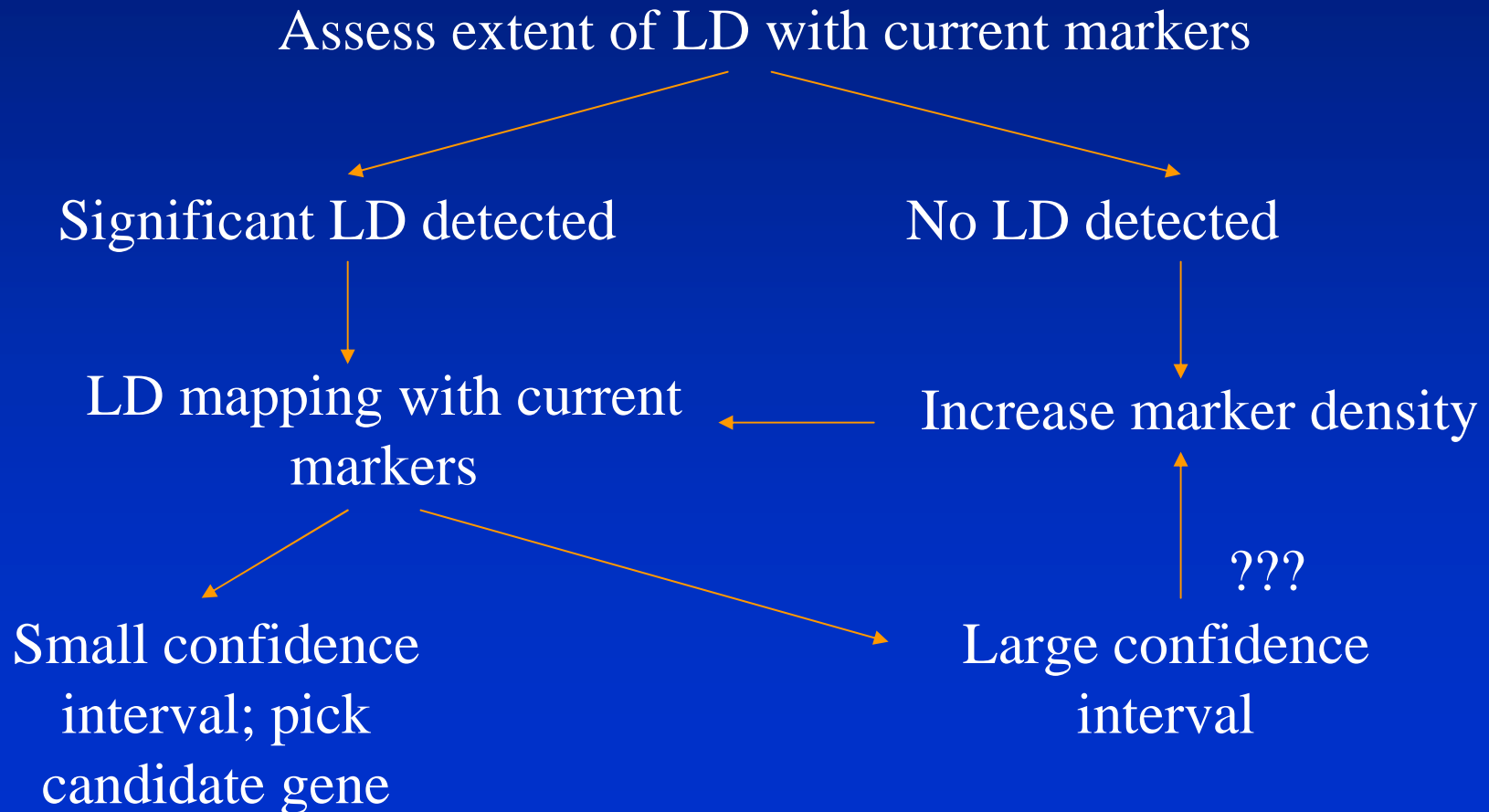
- Y = vector of phenotypes, μ = mean, X , Z and W are design matrices, b is a vector of fixed effects, u a vector of polygenic effects, v a vector of QTL allele effects, e a vector of random residuals, where
 - $u \sim (0, A\sigma_u^2)$, $v \sim (0, G\sigma_v^2)$, $e \sim (0, I\sigma_e^2)$
- For each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

LD mapping

- Accuracy of estimating QTL variance depends on
 - number of unique haplotypes sampled from the population
 - must be large enough to be representative of the population.
 - number of observations per unique haplotype
 - determines the accuracy of estimating the haplotype effects
 - If marker haplotypes are to be used in MAS, accuracy of estimating the effect of a unique haplotype determines amount of improvement in accuracy of selection as a result of using marker information.

LD mapping

- Pathway for LD mapping



Linkage disequilibrium (LD) mapping

- Definitions of LD
- Why does LD occur?
- Extent of LD in humans and livestock
- What type of markers are appropriate for detecting LD?
- LD mapping
 - building IBD matrix from marker haplotypes
 - variance component approach
- Combined LD-LA mapping

Combined LD-LA mapping

- Extent of LD very variable
- LD can exist between loci on different chromosomes!!
- Combine LD and linkage information to filter spurious peaks

Combined LD-LA mapping

- Consider a half sib design
 - LD information from sire haplotypes, maternal haplotypes of progeny
 - Linkage information from paternal haplotypes of progeny

- IBD matrix:

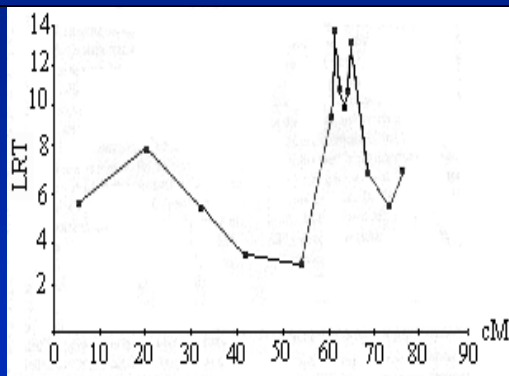
	SH	MHP	PHP
SH	[a]	[a]	[b]
MHP	[a]	[a]	[b]
PHP	[b]	[b]	[b]

- $a = \text{LD}$ (Meuwissen and Goddard 2001)
- $b = \text{linkage}$

Combined LD-LA mapping

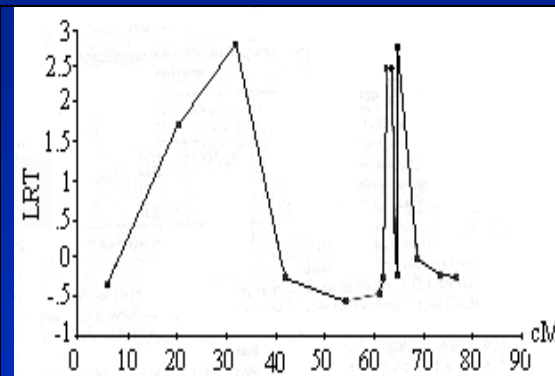
- Example of twinning QTL in Norwegian dairy cattle (Meuwissen et al. 2002)

LD



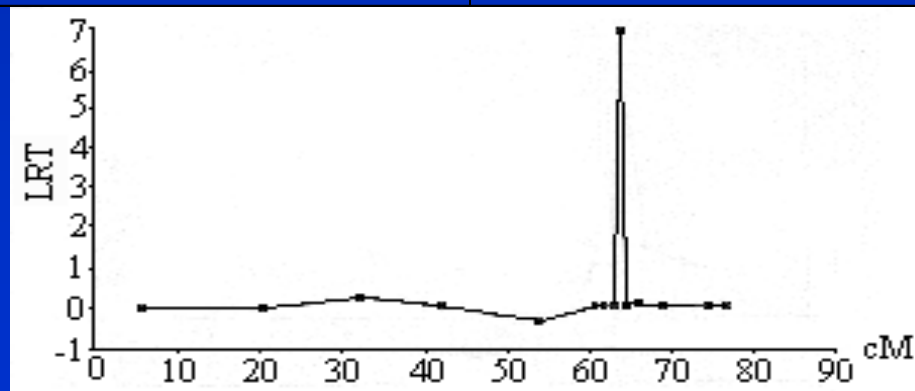
A

LA



B

LD-LA

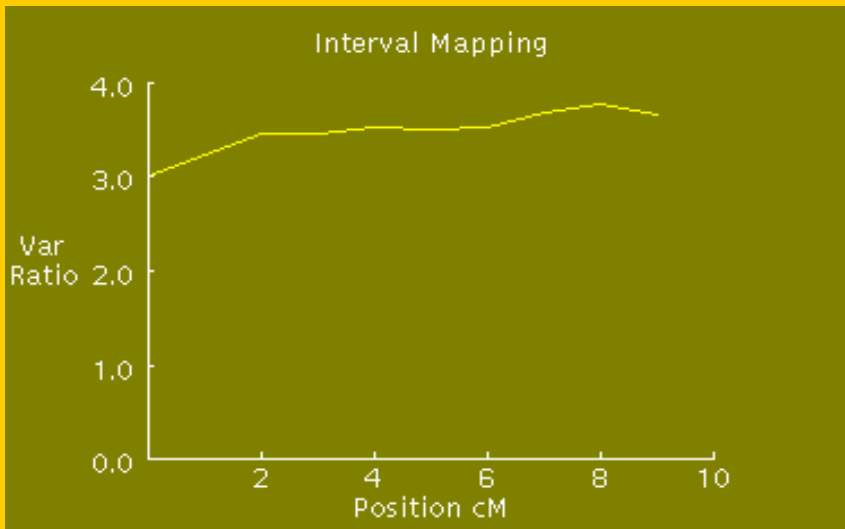


C

Combined LD-LA mapping

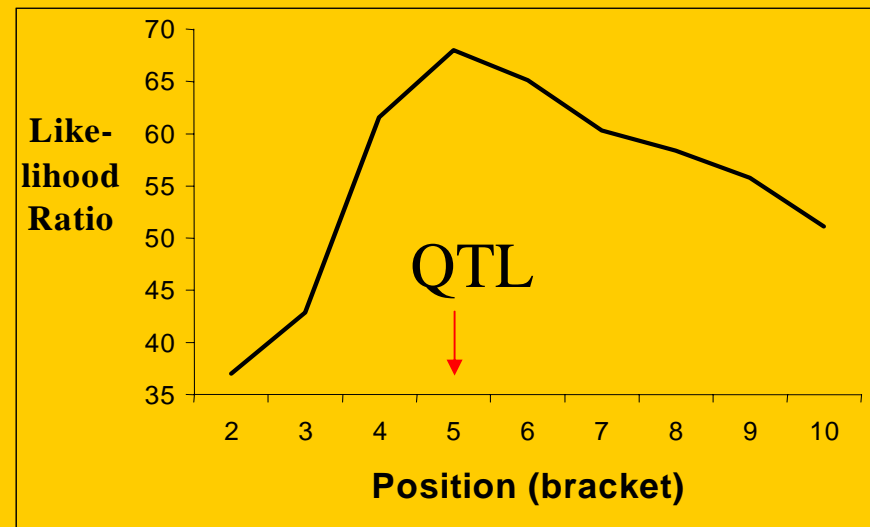
- How much information does LD add to the analysis?
 - Depends on marker spacing

LA



A

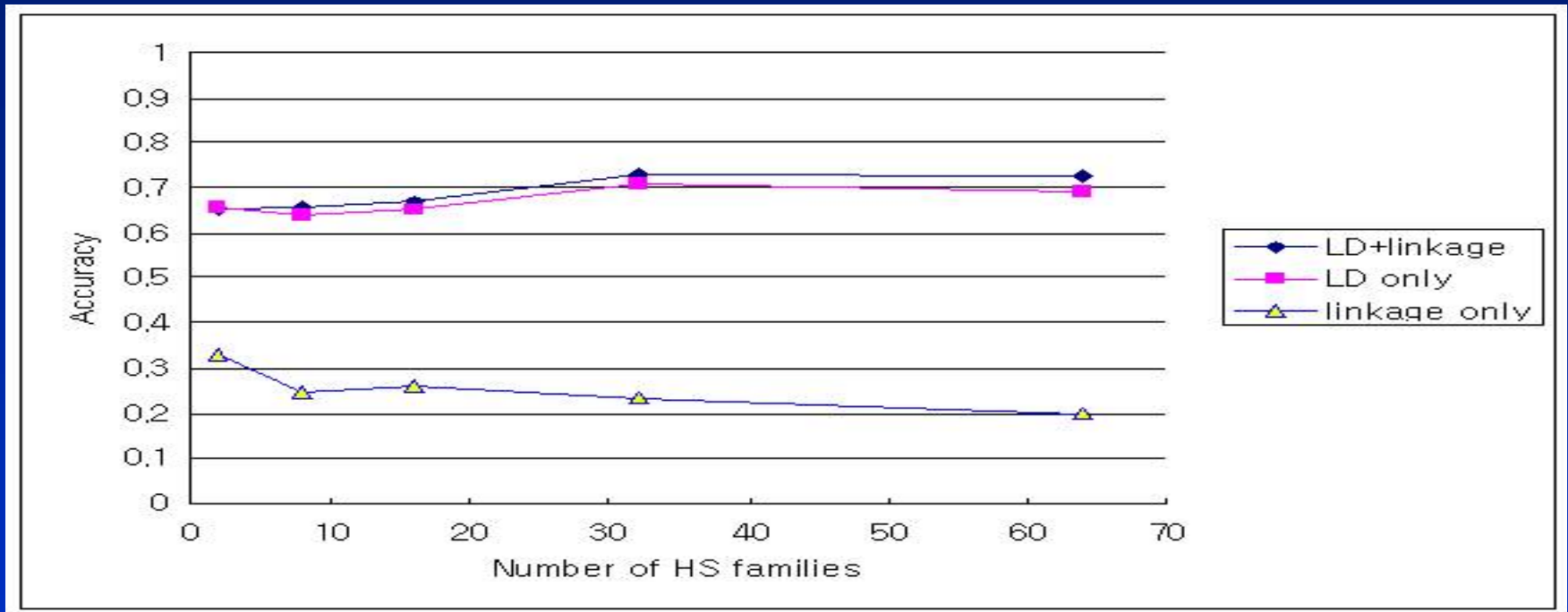
LD-LA



B

Combined LD-LA mapping

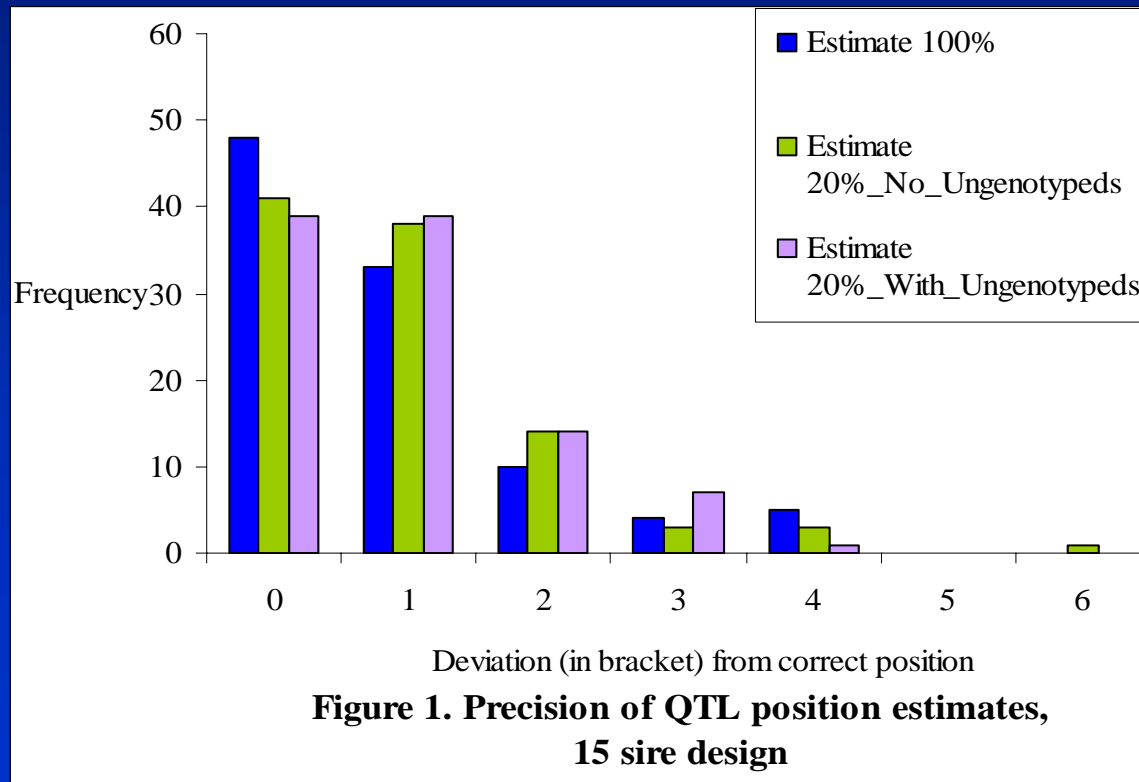
- Can we use half-sib families for LD analysis?



- Yes

Combined LD-LA mapping

- Can we use half-sib families for LD analysis?
 - + selective genotyping ?



- Yes

Combined LD-LA mapping

- LDLA analysis + selective genotyping =
Cheap? experiment able to position QTL with
high degree of precision