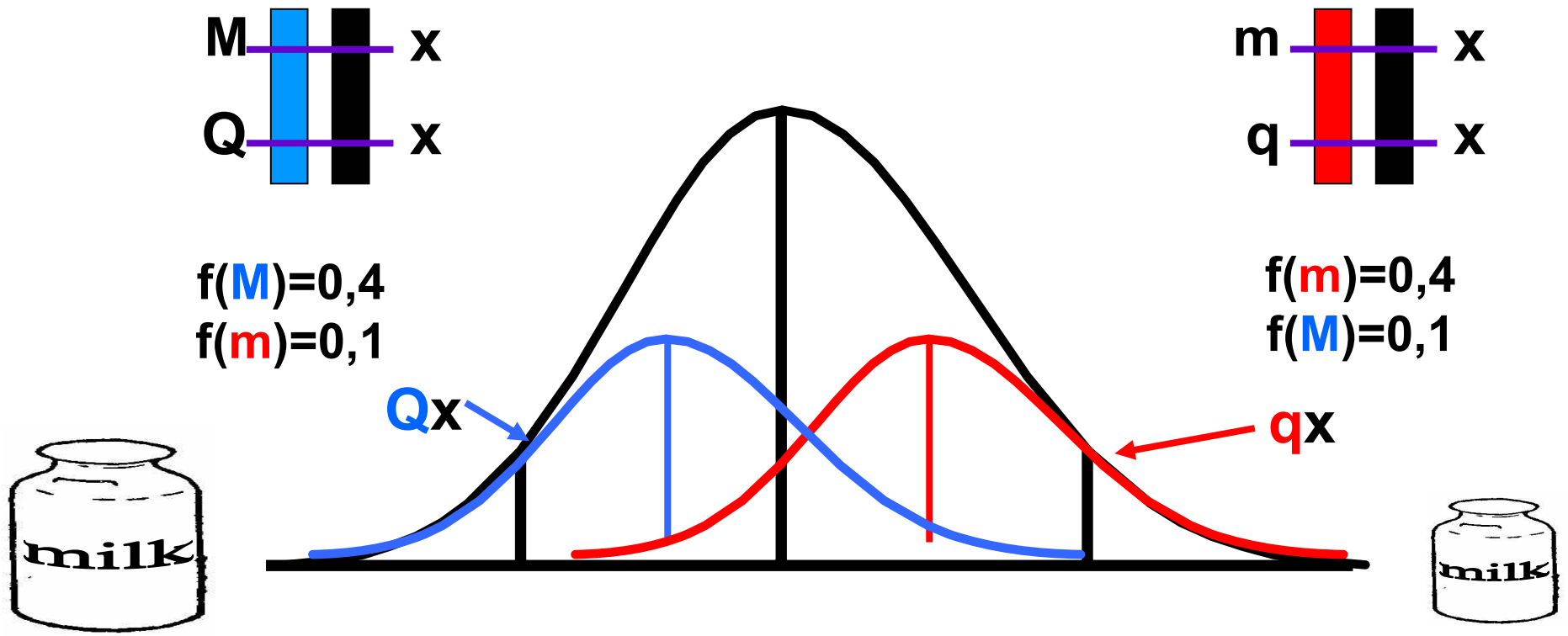
University of Natural Resources
and Applied Life Sciences, Vienna
Department of Sustainable Agricultural
Systems

approximate interval mapping for selective DNA pooling

M. Dolezal[1,2], H. Schwarzenbacher[1,2]
J. Sölkner[1], P. M. Visscher[2]

1. University of Natural Resources and Applied Life Sciences Vienna
2. University of Edinburgh, Institute of Cell, Animal and Population Biology

selective DNA pooling



10 % best daughters of a sire

M m

10 % worst daughters of a sire

single marker - single sire test

- significance of marker m_i within family
  - for sires heterozygous at m_i

$$Z_{Di} = \frac{D_i}{SE(D_i)}$$

D_i sire marker allele frequency difference between high an low pool

single marker – across sires

- **significance of marker m_i over n heterozygous sires**

$$TS = \sum_{j=1}^{n} Z_{ij}^2 \sim \chi^2 \text{ (n)}$$

n....number of sires heterozygous at m_i

single marker – across sires

- for each marker another number of sires is heterozygous

  thus

- TS more or less strongly affected by the number and QTL status of the specific sires that are heterozygous.

approximate interval mapping

- an approximate multiple marker method
- predict test statistics (T_l)
- from observed teststatistics (T_i)
- simple selection index analogy
  - relationship T_l and T_i
  - function of genetic distance recombinationsrate between these 2 points

approximate interval mapping

$$T_i = Z^2_{Di}$$

- observed teststatistic T_i at position i
- prediction of teststatistic T_l at position l

  $E(T_l) \sim (1-2r)^2 T_i$
- under H_0 variance of T_i

  $var(T_i) \sim \chi^2_{(1)} = 2$

approximate interval mapping

- **covariance between predicted T_l and observed teststatistic T_i**

$$\text{cov}(T_l, T_i) \approx (1-2r)^2 \text{ var}(T_i) = 2(1-2r)^2$$

$$\text{cov}(E(T_l), T_i) \approx (1-2r)^2 \text{ var}(T_i) = 2(1-2r)^2$$

- **combination of above equations allows multipoint prediction for each position on the chromosome**

$$E(T_l) = \mathbf{b't}$$

approximate interval mapping

$$E(T_l) = b't$$

- **t .....vector of all observered T_i**
- **b .....solution to $V^{-1}c$**
- **V = var(t)..... variance-covariance-Matrix of T_i**
  - **2 on diagonal**
  - **$(1-2r_{il})^2$ ....offdiagonal elements**
- **c .....vector of covariances between predicted T_l and observed T_i**

simulation

M1	M2	M3	QTL	M4	M5	M6
0	20	40	50/55	60	80	100

- 10 half sib families - 2000 progeny
- 6 evenly spaced markers on a 100 cM chromosome
- sire markers ≠ markers of dams
- crossovers - Haldane mapping function
- a single biallellic QTL with population frequency 0.5 at
  - 50 or 55 cM
  - allele substitution effect of 0.25 σ_p
- overall heritability of the trait 0.25

simulation

$y_{ij} = \mu$

simulation

$$y_{ij} = \mu + g_{QTL_{ij}}$$

simulation

$$y_{ij} = \mu + g_{QTL_{ij}} + \tfrac{1}{2} g_{sire_i} + \tfrac{1}{2} g_{dam_{ij}}$$

simulation

$$y_{ij} = \mu + g_{QTL_{ij}} + \tfrac{1}{2} g_{sire_i} + \tfrac{1}{2} g_{dam_{ij}} + g_{M_{ij}}$$

simulation

$$y_{ij} = \mu + g_{QTL_{ij}} + \tfrac{1}{2} g_{sire_i} + \tfrac{1}{2} g_{dam_{ij}} + g_{M_{ij}} + \varepsilon_{ij}$$

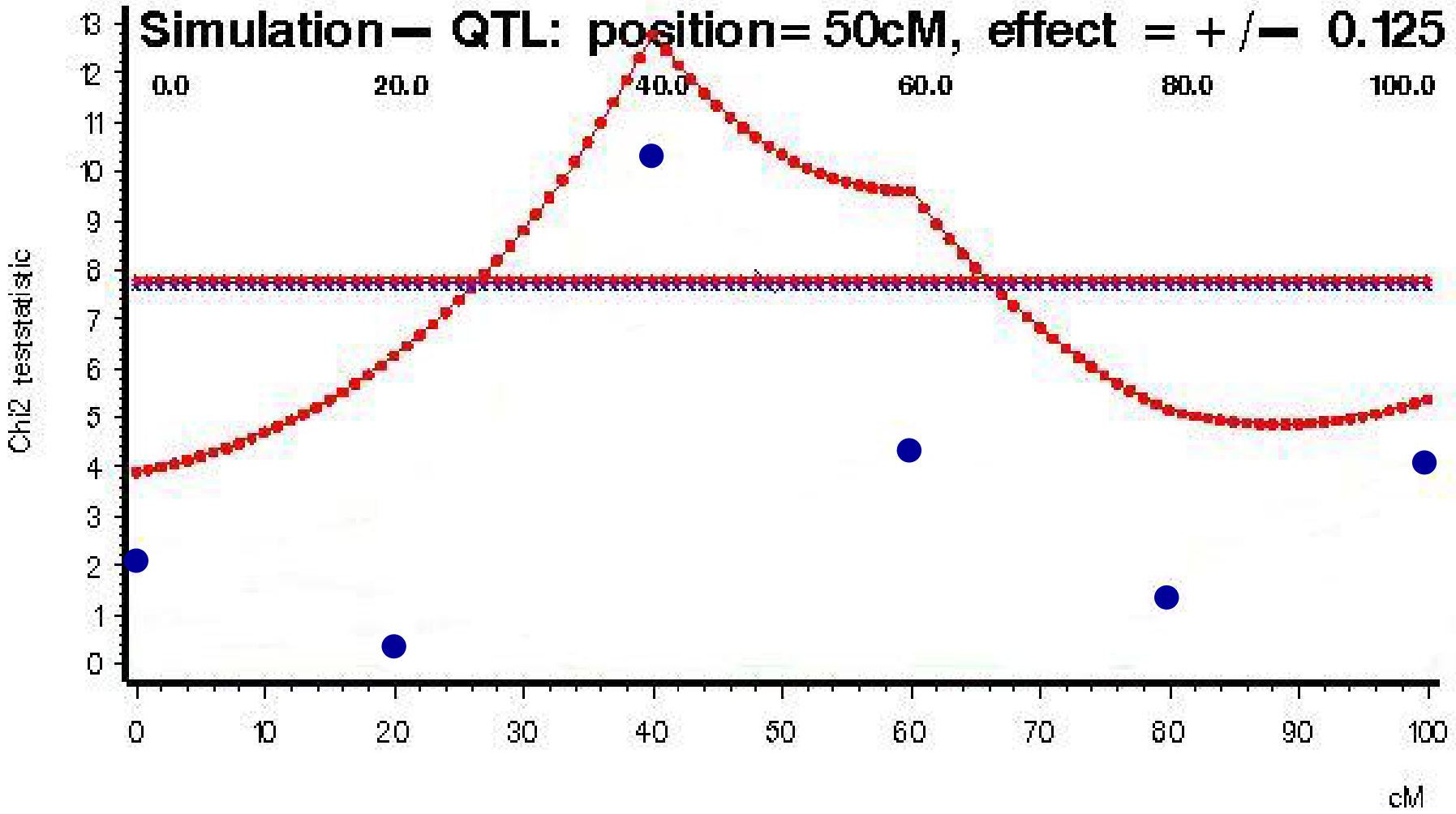
simulation

$$y_{ij} = \mu + g_{QTL_{ij}} + \tfrac{1}{2} g_{sire_i} + \tfrac{1}{2} g_{dam_{ij}} + g_{M_{ij}} + \varepsilon_{ij}$$

- sDNAp with 10% best + worst offspring V_T=0.0014
- significance thresholds based on 10.000 runs under H_0
- 3000 runs for each setting
- marker informativity:
  - all 6 markers fully informative
  - on average 50% informative markers
  - on average 30% informative markers

results

- **comparing AIM to single marker across sire analysis for sDNAp**
- **power**
  - **QTL detected**
    - either marker on the left or marker to the right significant
  - **Interval containing the QTL correctly identified**
    - both flanking markers significant
- **bias of QTL location estimate**


Simulation— QTL: position= 50cM, effect = +/— 0.125
0.0
20.0
40.0
60.0
80.0
100.0
Chi2 teststatistic
0
1
2
3
4
5
6
7
8
9
10
11
12
13
0
10
20
30
40
50
60
70
80
90
100
cM

power

marker informativity of sire	Interval identified SMM	Interval identified AIM	QTL identified SMM	QTL identified AIM
100%	94.0%	99.7%	99.7%	100.0%
50%				
30%				

power

marker informativity of sire	Interval identified SMM	Interval identified AIM	QTL identified SMM	QTL identified AIM
100%				
50%	86.7%	99.0%	98.3%	99.3%
30%				

power

marker informativity of sire	Interval identified SMM	Interval identified AIM	QTL identified SMM	QTL identified AIM
100%				
50%				
30%	54.3%	87.7%	90.3%	94.7%

power

marker informativity of sire	Interval identified SMM	Interval identified AIM	QTL identified SMM	QTL identified AIM
100%	94.0%	99.7%	99.7%	100.0%
50%	86.7%	99.0%	98.3%	99.3%
30%	54.3%	87.7%	90.3%	94.7%

bias of QTL locus estimate

	σ	μ	σ	μ
	SMM		AIM	
100%	0.0	5.0	0.0	5.0
50%	5.0	8.6	4.4	7.5
30%	6.6	9.8	6.4	9.3

conclusions

- Power of AIM is higher.
- Advantage of AIM increases with decreasing marker informativity of the sire.
- Bias smaller with AIM.
- Validity of AIM is constrained to 1 QTL models.
- In practice $(1–2r)^2$ drops off very rapidly with distance.
- Thus 2 QTL, well separated on the same chromosome will exert little confounding effect on their respective observed or predicted test statistics.

outlook

- different parameter settings to be tested by simulation
  - multi QTL models
  - QTL not centered in middle of chromosome
  - smaller family size
  - different QTL sizes

thank you for your attention