Joint association and linkage QTL mapping in (full) and half-sib families by regression

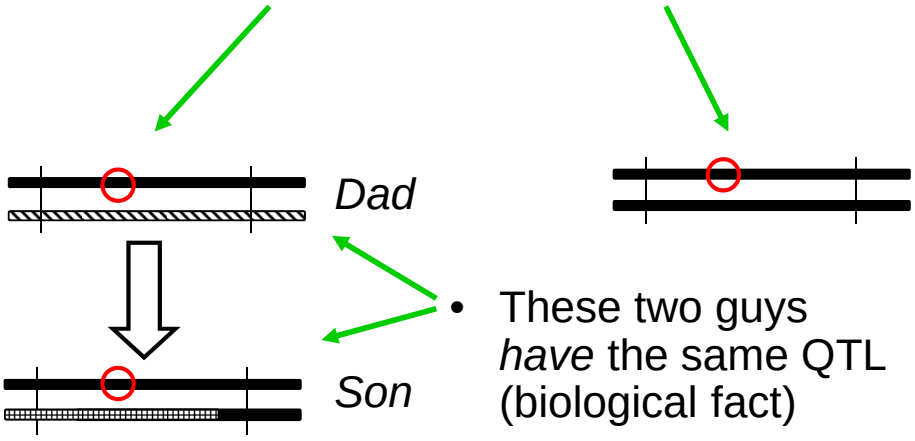
(Theory: Legarra A & Fernando R.L, GSE 2009)

Association mapping

- The principle of association mapping is that we can predict the QTL state from the marker state in a close marker
  - (at the population level, i.e., “linkage disequilibrium”)
- *Now, what do we do with related animals?*

Related animals

- These two guys *perhaps* have the same QTL (likely assumption)

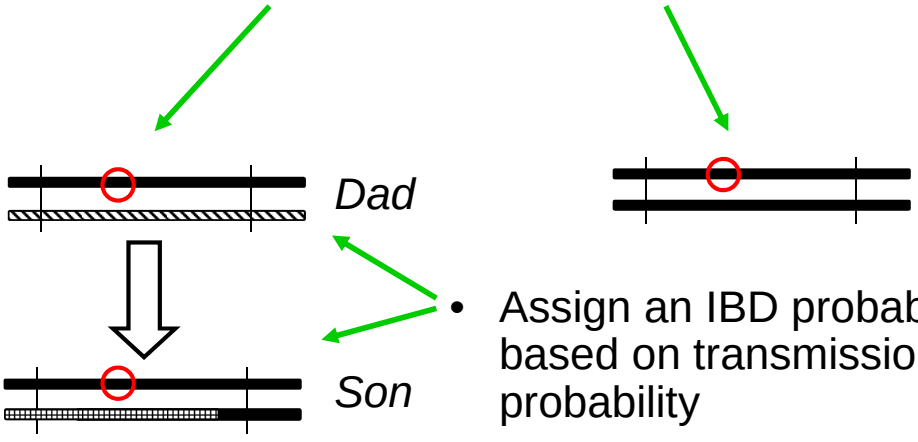


- These two guys *have* the same QTL (biological fact)

- We can accommodate these two informations

IBD method (Meuwissen et al. 2002)

- Assign an IBD probability based on population genetics



- Assign an IBD probability based on transmission probability

- Construct a mixed model

Inconvenients

- We lost nice properties of regression methods (speed, flexibility)
  - Can't do: Bootstrap CI, Permutation tests
- IBD matrix rather tricky
  - Need "bending" or "clustering"
- Big IBD matrix ($4n^2$)

Regression

- Regression is a crude but efficient alternative to mixture models
- Regression is based on conditional expectations
  - Expectation for the founders
  - Expectation for the offspring
- We reason for two-marker haplotypes but formulae are identical for any size of haplotypes
- I will make the presentation for half-sib families but full-sib families are an immediate extension
- There is also a mixed-model version for general pedigrees but it has never been programmed

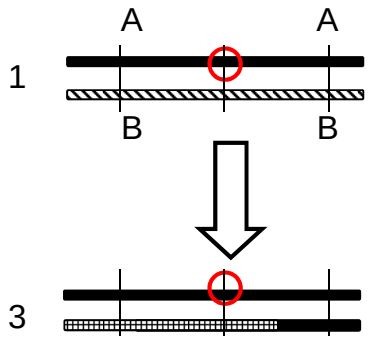
- **LDdecay → QTLMAP "LD"**
- LDLA → QTLMAP "LDLA"

Founders

- If there is strong LD (and a QTL segregating) we hope that QTL alleles carried by haplotypes "AA" or "BB" are very different
  - We can simply postulate an effect of AA and another of BB
- This is a regression of phenotype on haplotype state (haplotype association analysis)

$$y_1 = b_{AA} + b_{BB} + e_1$$
$$y_2 = 2b_{AA} + e_2$$

Non Founders



- The son 3 inherited the QTL at the "black chromosome" with probability *p*
- Or the QTL at "grey" with probability *1-p*

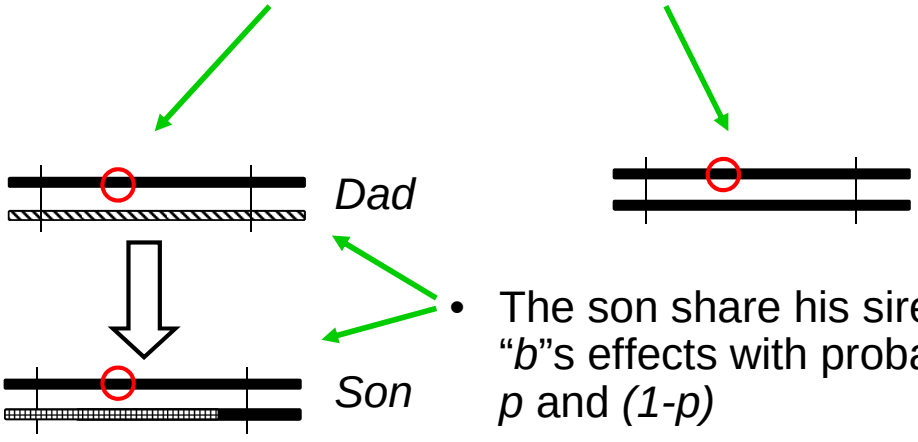
$$y_1 = b_{AA} + b_{BB} + e_1$$

$$y_3 = pb_{AA} + (1-p)b_{BB} + e_3$$

- The *p*'s use the linkage information and are computed using all available markers
- <u>We don't care</u> what haplotype son 3 is carrying, because we know the paternal QTL is one of his sire's

Regression method

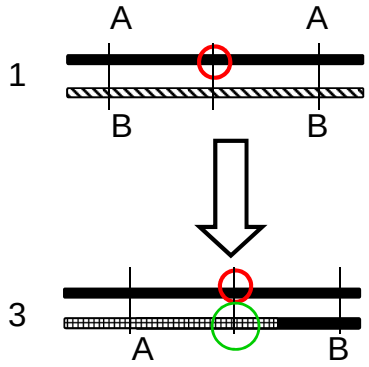
- These two guys share the same b_{BB} effect



- The son share his sire's "*b*"s effects with probability *p* and *(1-p)*

- "LDdecay" because it models the *decay* of *LD* in the founders through transmission probabilities *p*

Non Founders – maternal info



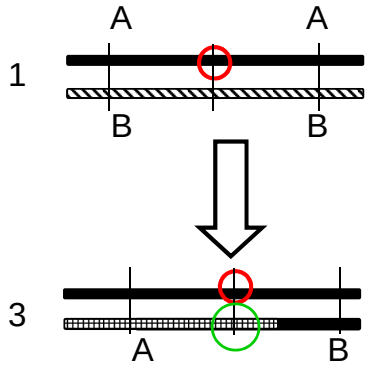
- The “barred” chromosome in 3 is the maternal one, and is carrier of haplotype AB
- Add its effect

$$y_1 = b_{AA} + b_{BB} + e_1$$

$$y_3 = pb_{AA} + (1-p)b_{BB} + b_{AB} + e_3$$

- The chromosome from the dam also uses populational LD
- Lots of information

Regressions



- Regression for the sire

$$y_1 = b_{AA} + b_{BB} + e_1$$

- Regression for the son

$$y_3 = pb_{AA} + (1-p)b_{BB} + b_{AB} + e_3$$

- We can’t mix regressions for founders and sons because residual variances are not the same unless *p* =1 or 0
- But use for half-sib designs is straightforward (equation for y_3)
  - Families are tied together through *b*’s

Example

Table 1: Pedigree and markers for the numerical example

animal	dam	sire	Maternal haplotype	Paternal haplotype
1	0	0	10	01
2	0	0	11	00
3	0	0	01	11
4	1	2	10	00
5	3	2	01	11
6	3	2	01	01
7	4	5	00	11
8	4	5	00	01

y

Recombination ⇨

- Regression model LDdecay $\boldsymbol{y} = \boldsymbol{Tb} + \boldsymbol{e}$

Recombination

$$\boldsymbol{y} = \begin{pmatrix} 0.98 & 0 & 1 & 0.02 \\ 0.02 & 1 & 0 & 0.98 \\ 0.50 & 1 & 0 & 0.50 \\ 1 & 0.02 & 0 & 0.98 \\ 1 & 0.98 & 0 & 0.02 \end{pmatrix} \begin{pmatrix} b_{00} \\ b_{01} \\ b_{10} \\ b_{11} \end{pmatrix} + \boldsymbol{e}$$

Probability of individual *i* of having inherited the QTL in the founder haplotype *j*

Haplotypes in the founders

- LDdecay → QTLMAP "LD"
- **LDLA → QTLMAP "LDLA"**

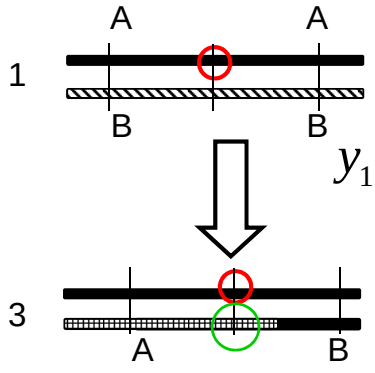
LDLA

- What if sires' QTLs are different from what we expect based on haplotype?
- There will be a difference:

$$QTL_1 = b_{AA} + v_1$$
$$QTL_2 = b_{BB} + v_2$$

- We can include this in the regression equations

Regressions



- Regression for the sire

$$y_1 = b_{AA} + v_1 + b_{BB} + v_2 + e_1$$

- Regression for the son

$$y_3 = pb_{AA} + pv_1 + (1-p)b_{BB} + (1-p)v_2 + b_{AB} + e_3$$

- We can't mix regressions for founders and sons because residual variances are not the same unless *p* =1 or 0
- But use for half-sib designs is straightforward (equation for y_3)
  - Families are tied together through *b*'s

Example

Table 1: Pedigree and markers for the numerical example

animal	dam	sire	Maternal haplotype	Paternal haplotype
1	0	0	10	01
2	0	0	11	00
3	0	0	01	11
4	1	2	10	00
5	3	2	01	11
6	3	2	01	01
7	4	5	00	11
8	4	5	00	01

y

- Regression model LDdecay $\boldsymbol{y} = \boldsymbol{T_b b} + \boldsymbol{T_v v} + \boldsymbol{e}$

$$\boldsymbol{y} = \begin{pmatrix} 0.98 & 0 & 1 & 0.02 & 0.02 & 0.98 & 0 & 0 \\ 0.02 & 1 & 0 & 0.98 & 0.98 & 0.02 & 0 & 0 \\ 0.50 & 1 & 0 & 0.50 & 0.50 & 0.50 & 0 & 0 \\ 1 & 0.02 & 0 & 0.98 & 0 & 0 & 0.02 & 0.98 \\ 1 & 0.98 & 0 & 0.02 & 0 & 0 & 0.98 & 0.02 \end{pmatrix} \begin{pmatrix} b_{00} \\ b_{01} \\ b_{10} \\ b_{11} \\ v_{2,1} \\ v_{2,2} \\ v_{5,1} \\ v_{5,2} \end{pmatrix} + \boldsymbol{e}$$

Haplotypes in the founders

Residual QTL effects in the founders

Probability of individual *i* of having inherited the QTL in the founder haplotype *j*

Probability of *i* of having inherited the "residual" QTL in the founder chromosome

Caveat LDLA

$$y_3 = \underbrace{pb_A + (1-p)b_B}_{\text{This is Haley-Knott LA regression and LD}} + \underbrace{b_B}_{\text{This is maternal LD}} + \underbrace{pv_{1,1} + (1-p)v_{1,2}}_{\text{This is "pure" LA Haley-Knott regression}} + e_3$$

"LDdecay"

Caveat LDLA

- LDLA allows detecting linkage in absence of LD
- But not without complications
- If
  - « *v* » effects are considered for sires and dams
  - *and* all founders are genotyped (including dams)
  - *and* founders have no record (e.g., full-sib designs)
  - then the LDLA regression is formally *equivalent* to the LA regression
- In this case it is better to use LDdecay

Extensions

- What if animals come from two populations?
  - E.g. Romane x Blackbelly BC
    - "BB" haplotype may not have the same effect in each breed
  - Need to define "within-breed" haplotype effects
  - C Moreno did it for QTLmap (e.g. Sallé et al. 2012)

Performance

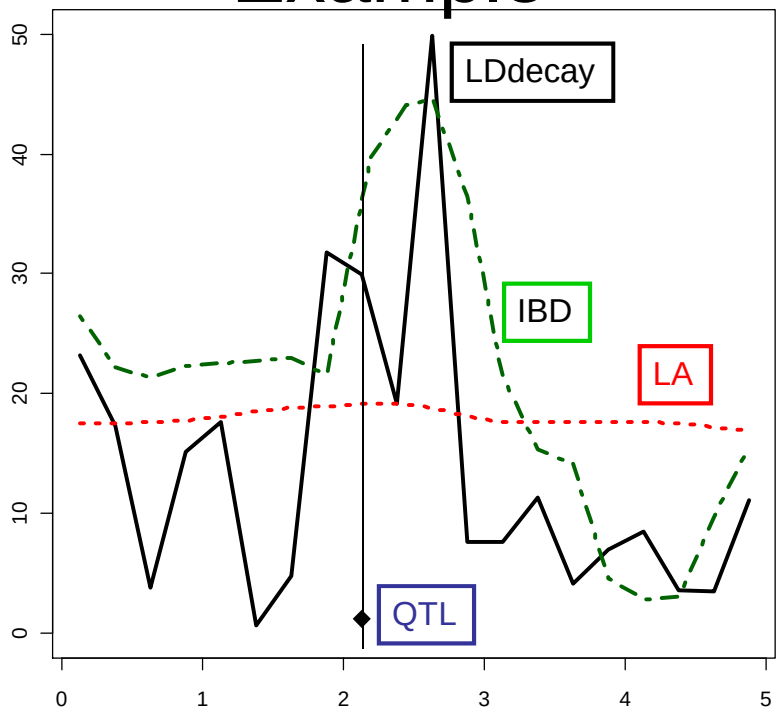
- Simulation under several scenarios
  - 15 x 20 half-sib families, one big QTL, drift, 21 markers
  - It depends on the scenario, but generally all methods (LDdecay, Meuwissen's IBD) perform similarly
  - IBD method is slightly biased towards the center because it uses all markers (but this is implementation dependent)

Performance

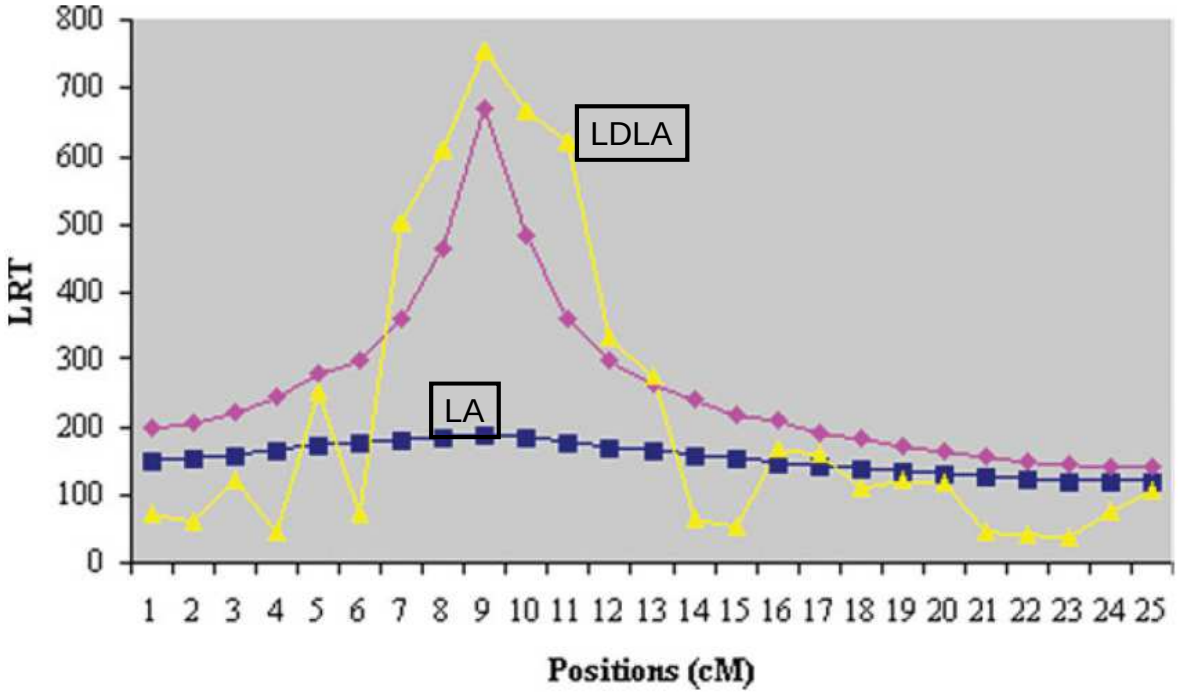
- 5 cM

Method	Bias	MSE	
LA	0.29	2.22	← Linkage alone is not accurate
LD decay	0.11	0.69	} Both are similar
IBD	0.34	0.78	

Example



Roldan et al. 2012



Conclusion & perspectives

- As good as other methods. Performance is more dependent on the scenario than on the method itself
  - See Roldan et al. (Gen Res) for a more extensive evaluation
- The method is very simple to implement provided phases and probabilities of transmission can be computed
  - This is easy with SNP chips
  - Allows bootstrap confidence intervals & Permutation tests