Mixed models for QTLxE in multi-parent populations (AMPRIL, NAM, diallel)

Fred van Eeuwijk &

Martin Boer, Joao Paulo, Hans Jansen, Marcos Malosetti, Marco Bink

MAGIC, Cambridge, 13 June 2013

Mixed model QTL mapping

- Phenotype =
  - | Replicate +
  - | Incomplete block within replicate +
  - | Genotype +
  - | error

- QTL mapping:
  - | Genotype =
    - QTLs +
    - residual genotype

- QTLs are found by ‘regression’ on marker information (conditional QTL genotype probabilities given marker genotypes and map/ IBD)

QTL mixed model for multiple environments (QTLxE)

- Phenotype =
  - Environment mean +
  - Environment specific QTLs +
  - Environment specific residual genetic effect +
  - Environment specific error

$$\underline{P}_{ij} = \mu + E_j + \underline{G}_i + \underline{GE}_{ij} + \underline{\varepsilon}_{ij}$$

$$\underline{P}_{ij} = \mu + E_j + \sum x_i a + \underline{G}_i + \sum x_i a_j + \underline{GE}_{ij} + \underline{\varepsilon}_{ij}$$

$$\underline{P}_{ij} = \mu_j + \sum x_i a_j + \underline{G}_{ij} + \underline{\varepsilon}_{ij}$$

- Mean

$$\underline{\mu}_{ij} = \mu_j + \sum x_i a_j$$

- VCOV (genetic part)
  - Residual genetic variation should allow environment specific variances and correlations

$$VCOV(\underline{G}_{ij}) = \begin{bmatrix} \sigma_1^2 & & & & \\ \sigma_{21} & \sigma_2^2 & & & \\ . & . & . & & \\ . & . & . & . & \\ \sigma_{J1} & \sigma_{J2} & . & . & \sigma_J^2 \end{bmatrix}$$

QTL mixed model for multiple traits (genetic correlations)

- Phenotype =
  - Trait mean +
  - Trait specific QTLs +
  - Trait specific residual genetic effect +
  - Trait specific error

$$\underline{P}_{it} = \mu_t + \sum x_i a_t + \underline{G}_{it} + \underline{\varepsilon}_{it}$$

- VCOV for residual genetic variation between traits should allow for trait specific variances and correlations

$$VCOV(\underline{G}_{it}) = \begin{bmatrix} \sigma_1^2 & & & & \\ \sigma_{21} & \sigma_2^2 & & & \\ . & . & . & & \\ . & . & . & . & \\ \sigma_{T1} & \sigma_{T2} & . & . & \sigma_T^2 \end{bmatrix}$$

Multiple traits (genetic correlations) in multiple environments (QTLxE)

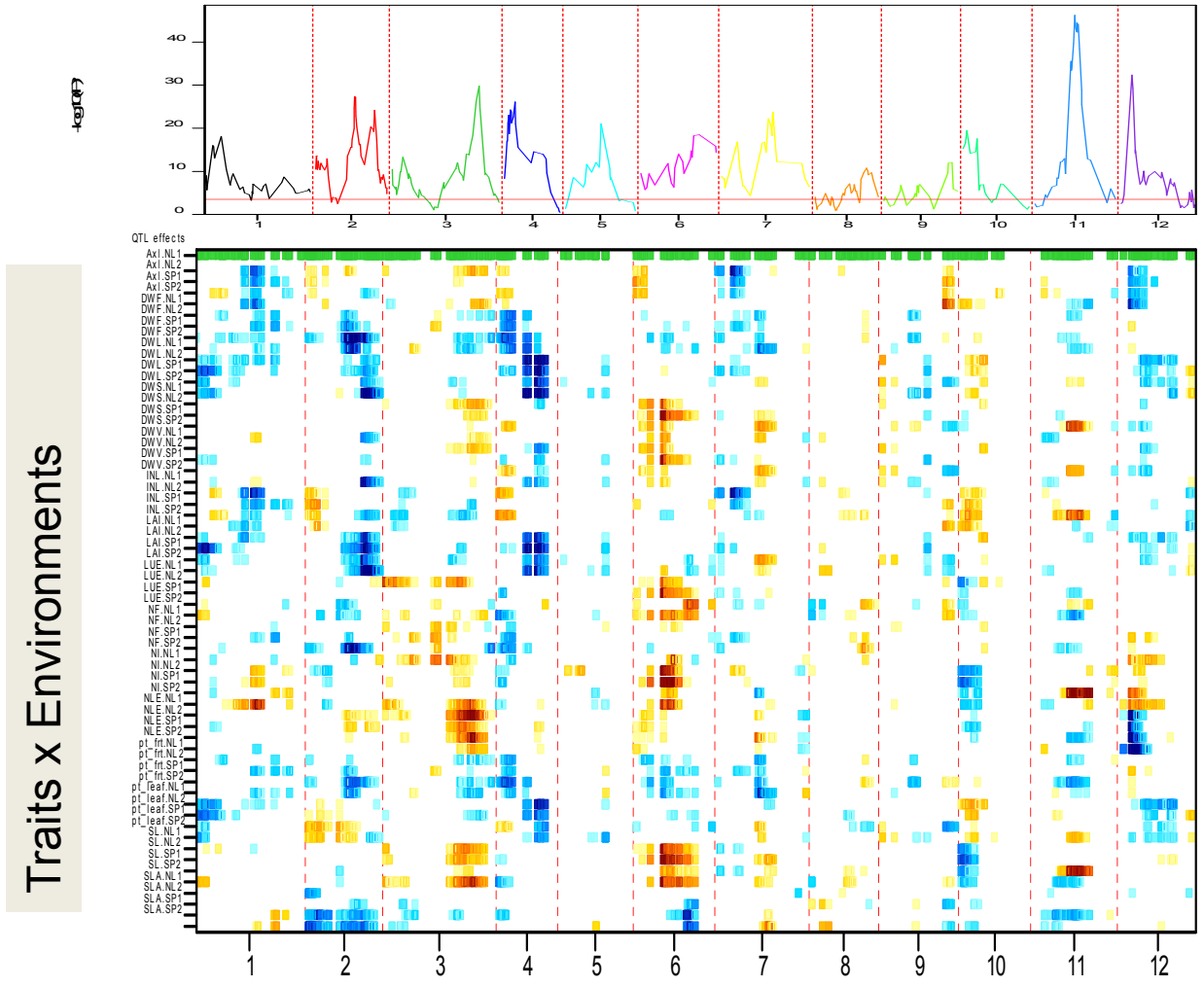
- Phenotype
  - Trait-environment mean +
  - Trait-environment specific QTLs +
  - Trait-environment specific residual genetic effect +
  - Trait-environment specific error

$$\underline{P}_{ijt} = \mu_{jt} + \sum x_i a_{jt} + \underline{G}_{ijt} + \underline{\varepsilon}_{ijt}$$

- VCOV
  - Genetic variances differ across trait-environment combinations
  - Genetic correlations between traits can be environment specific
  - Genetic correlations between environments can be trait specific

$$VCOV(\underline{G}_{ijt}) = VCOV(\underline{G}_{Traits}) \otimes VCOV(\underline{G}_{Envs})$$

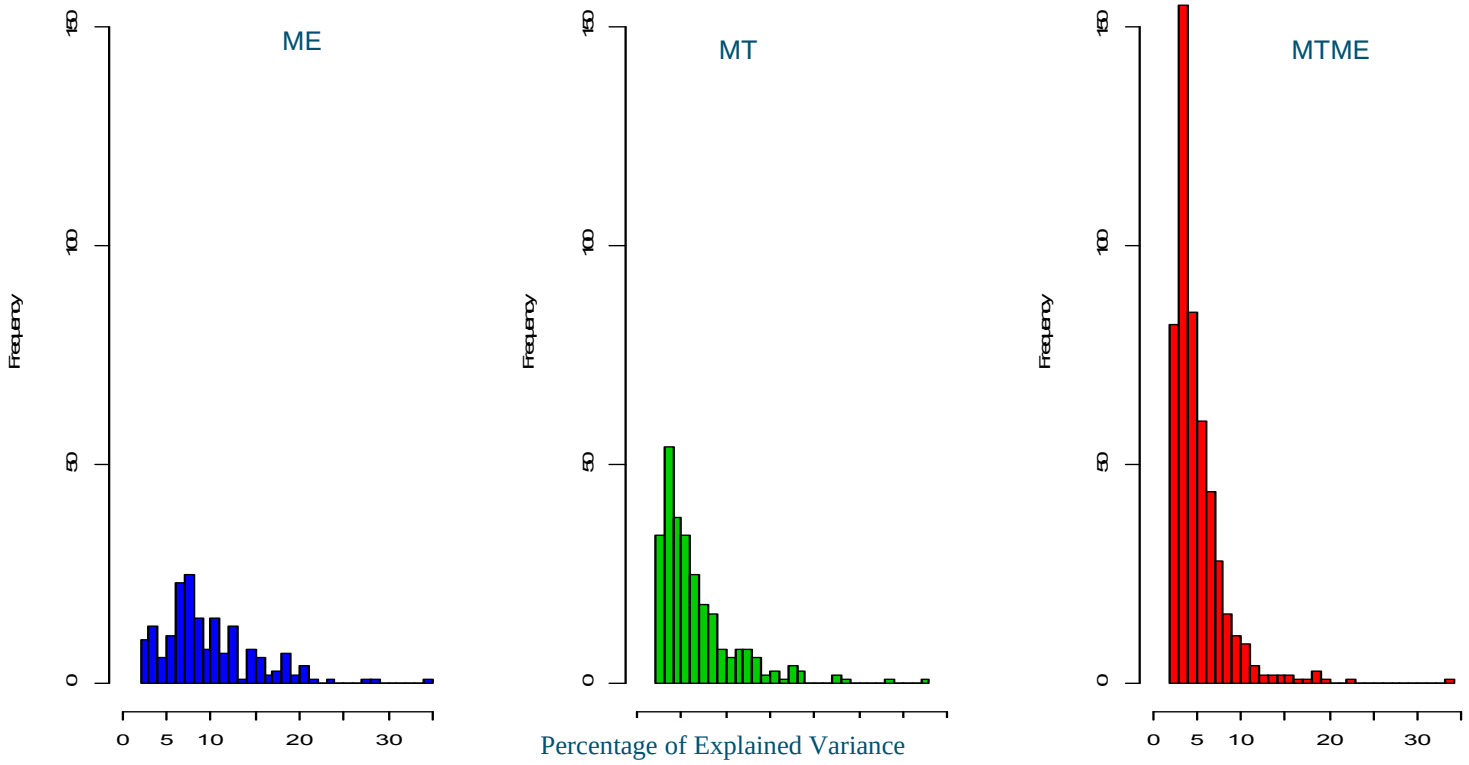
Multi-trait multi-environment QTL analysis



Nurudeen Alimi

Number of QTLs and explained variance when increasing model complexity, without going to genomic prediction

Figure 7: Histogram of Explained Variance by individual QTLs as detected by ME, MT and MTME analyses. MTME produced far more QTLs than ME and MT but many of the extra QTLs from MTME are of small effects.



WAGENINGEN UR
For quality of life

Nurudeen Alimi

QTL mapping in multi-parent populations

- Phenotype =
  - Population intercept +
  - Population specific QTLs +
  - Pop. spec. res. genetic effect +
  - Pop. spec. error

$$\underline{P}_{ic} = \mu_c + \sum x_i a_c + \underline{G}_{i \subset c} + \underline{\varepsilon}_{i \subset c}$$

i = genotype
c = cross / population

- Calculate **identity by descent** between **offspring and parents/ ancestors/ founders** for use as genetic predictors/ regressors, x_i .
- Statistical analysis by mixed models or Bayesian techniques

Literature and software

Table 1

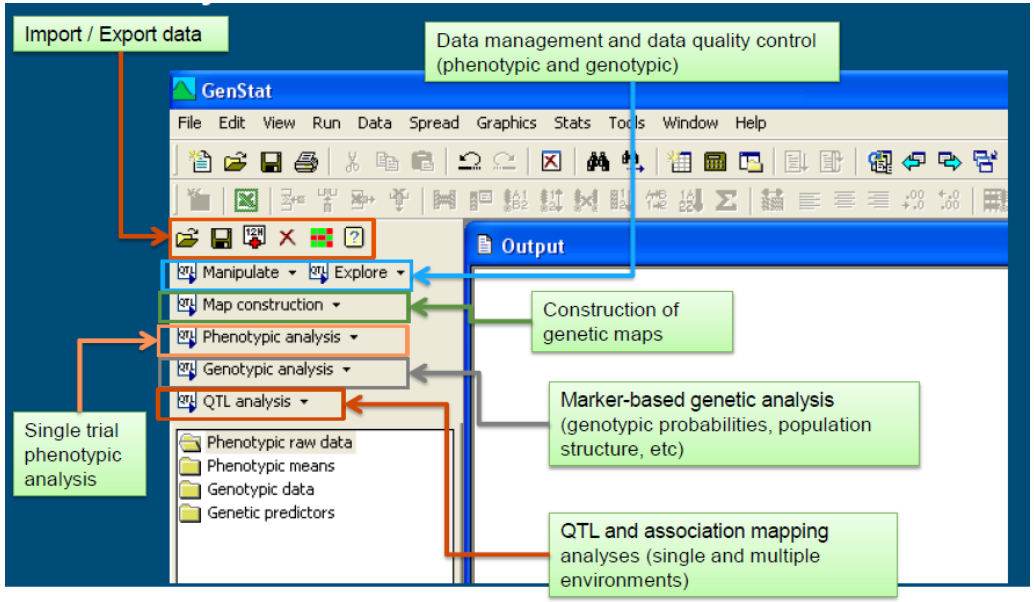
Overview of linear mixed models for QTL mapping. Abbreviations are given in the table (alphabetic order). Random terms are *P*, *e* and all terms in which *G* appears, or terms that arise from partitioning *G* or *G*-related terms (*A*, *D*, *I*). Fixed are *C*, *E*, *T* and [*TE*]. *Q* is usually fixed, but for highly polyallelic QTL in multiple connected populations and pedigreed populations *Q* is better taken random. *QE*, *QE**, *QT*, and *Q*[*TE*] are typically random. *G*(*Z*) is, in contrast, typically fixed. The VCOVs of the *G*-related terms in association analyses are structured by the relationships between genotypes on the basis of pedigree, markers, or a combination of those two [61]

Code	Model	Comments
1	*Single trait, single environment, single population model*	
1.1	$P = G + e$	Basic quantitative genetic model. Genotype is random and VCOV of genotypic effects needs to be structured by genotypic relationship matrix for association panels.
1.2	$P = A + D + I + e$	Partitioning of genotypic effect in additive, dominance and gene interaction (epistatic) effects.
1.3	$P = R + Ib + G + e$	Model 1.1 with terms accounting for experimental design, for example, replicate and incomplete block within replicate. Design terms are mostly taken random.
1.4	$P = Q + G^* + e$	QTL model in marker regression and SIM for biparental offspring populations. Genotype in 1.1 is partitioned in QTL and a residual genotype. For association analysis, VCOV of G^* needs structuring as in 1.1.
1.5	$P = \mathrm{Sum}(Q\_bg) + G^* + G^* + e$	Model for CIM, with test for QTL at particular position being corrected for background QTL elsewhere.
1.6	$P = \mathrm{Sum}(Q\_add) + \mathrm{Sum}(Q\_dom) + \mathrm{Sum}(Q\_epi) + G^* + e$	*G* in 1.1 can be partitioned in multiple QTL showing additive, additive × dominance, and additive × dominance × epistatic effects, and a residual genotypic effect.
2	*Multi-population*	
2.1	$P = C + Q + CQ + G^* + e$	For multiple biparental populations with common parents or association panels with population substructure. QTL are corrected for population means or population substructure, respectively. Epistasis can be tested in the form of QTL by background interaction.
3	*Multi-environment*	
3.1	$P = E + G + GE + e$	Multi-environment generalization of 1.1; fixed environmental main effect and random GE are added.
3.2	$P = E + GGE + e$	For defining mixed model with heterogeneous genetic variances and correlations between environments, 3.2 is more convenient than 3.1. For association panels, relations between genotypes need to be accounted for explicitly.
3.3	$P = E + Q + G^* + QE + GE^* + e$	Detecting genetic basis of GEI can be done by partitioning GEI in QTL part and residual GEI.
3.4	$P = E + QQE + GGE^* + e$	Model for genome scan for multi-environment trial data (biparental offspring populations). Main-effect QTL and QE are jointly modeled. Equally so, the residual genotypic main effect and GEI are modeled together.
3.5	$P = E + \mathrm{Sum}(QQE) + GGE^* + e$	Multiple QTL form of model 3.4.
3.6	$P = E + Q + Q(Z) + QE + GGE^* + e$	QE in 3.4 is modeled by regression on environmental covariable (*Z*).
3.7	$P = E + CCE + \mathrm{Sum}(QQE) + GGE^* + e$	Multi-environment generalization of 2.1, ignoring QTL by background interactions. Populations (multiple connected biparental) and population substructure groups (association panels) interact with environment. VCOV for the GGE^* needs structuring to account for genotypic relationships, as in 3.2.
4	*Multi-trait*	
4.1	$P = T + GGT + e$	4.1 is equivalent to 3.2, traits can take the role of environments and vice versa. However, multi-trait analysis is more sensitive to scaling issues, so main effect *T* term should always be in model and VCOV for GGT term should adequately account for heterogeneity of genotypic variances and correlations between traits.
4.2	$P = T + \mathrm{Sum}(QQT) + GGT^* + e$	Extension of 4.1 and equivalent to 3.4, with same remark on VCOV of GGT as for 4.1.
5	*Multi-trait multi-environment*	
5.1	$P = [TE] + GG[TE] + e$	As 3.2 and 4.1, but treating each trait by environment combination either as an environment in 3.2 or a trait in 4.1 that is fitting a fixed intercept term for each trait by environment combination. Again, VCOV for $GG[TE]$ term should be flexible.
5.2	$P = [TE] + \mathrm{Sum}(QQ[TE]) + GG[TE]^* + e$	QTL extension of 5.1.

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Plant Biology
ELSEVIER

Detection and use of QTL for complex traits in multiple environments

Fred A van Eeuwijk[1,2], Marco CAM Bink[1], Karine Chenu[3] and Scott C Chapman[4]



Multi-population QTL analysis: Case 1
Arabidopsis Multi-Parent RIL population

- Statistics
  - Fred van Eeuwijk
  - João Paulo
  - Martin Boer
  - Paul Keizer
  - Marco Bink
- Biology
  - Xueqing Huang
  - Sigi Effgen
  - Maarten Koornneef

Analysis of natural allelic variation in *Arabidopsis* using a multiparent recombinant inbred line population

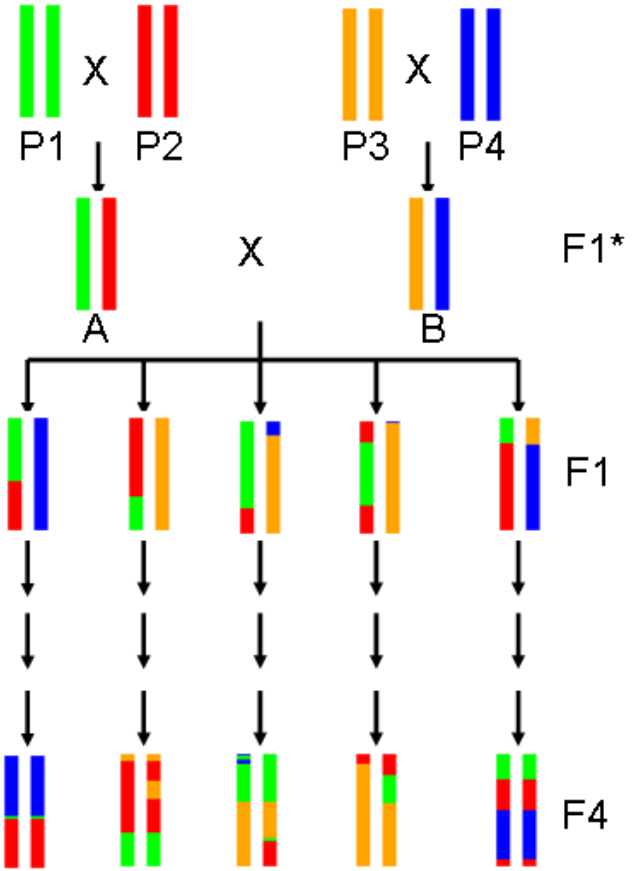
Xueqing Huang[a,1], Maria-João Paulo[b,c], Martin Boer[b], Sigi Effgen[a], Paul Keizer[b], Maarten Koornneef[a,d,1], and Fred A. van Eeuwijk[b,c]

4488–4493 | PNAS | March 15, 2011 | vol. 108 | no. 11

Arabidopsis Multi Parent RIL population (AMPRIL)

```
A = P1xP2 = Col    x Kyo-1
B = P3xP4 = Cvi    x Sha
C = P5xP6 = Eri-1 x An-1
D = P7xP8 = Ler    x C24
```

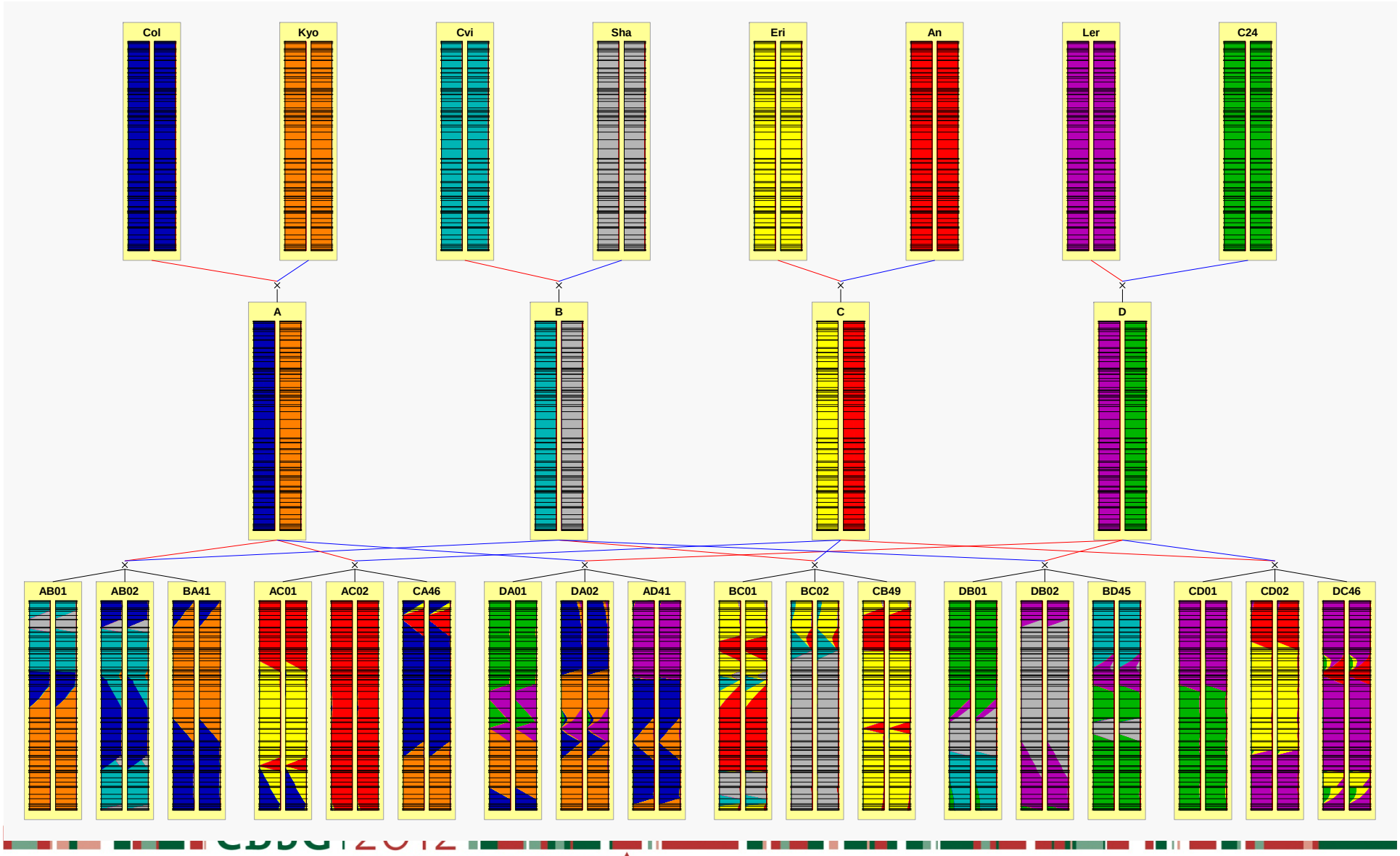
	A	B	C	D
A		x	x	x
B	x		x	x
C	x			x
D	x	x	x	



AMPRIL
Multi-parent, multi-population

- Inclusion of a certain amount of natural genetic variation
- Possibility to study new combinations of alleles
- Higher power due to higher probability of segregation (diallel of F1*'s)
  - Segregation in at least one, but hopefully more sub populations
- Higher precision
  - 4-way cross followed by 4 generations of inbreeding
- Residual heterozygosity in F4 allows fine mapping by producing NILs (HIFs)
- Epistasis: QTL x population & QTL x QTL

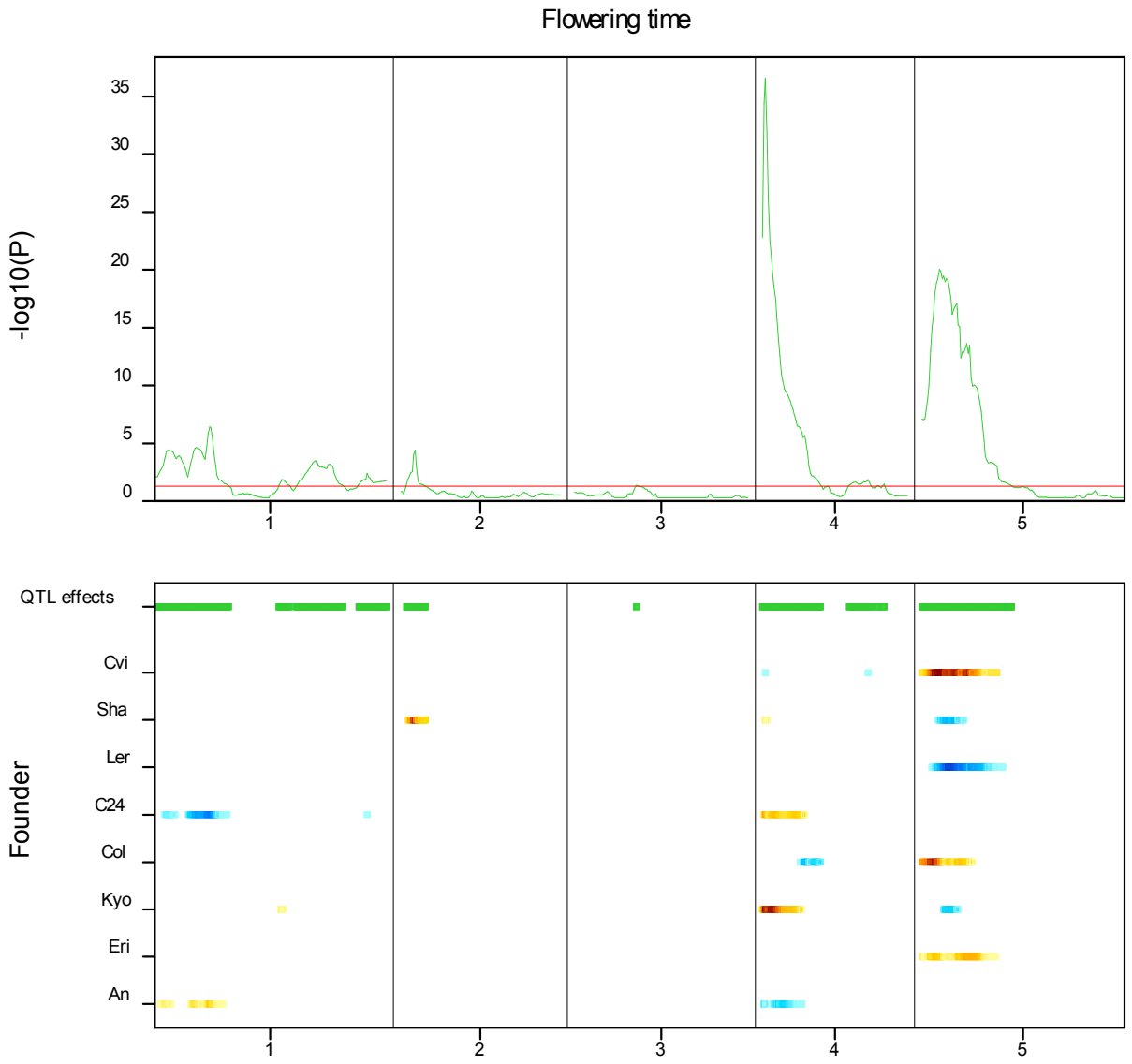
Founder Allele IBD probability : LG 1



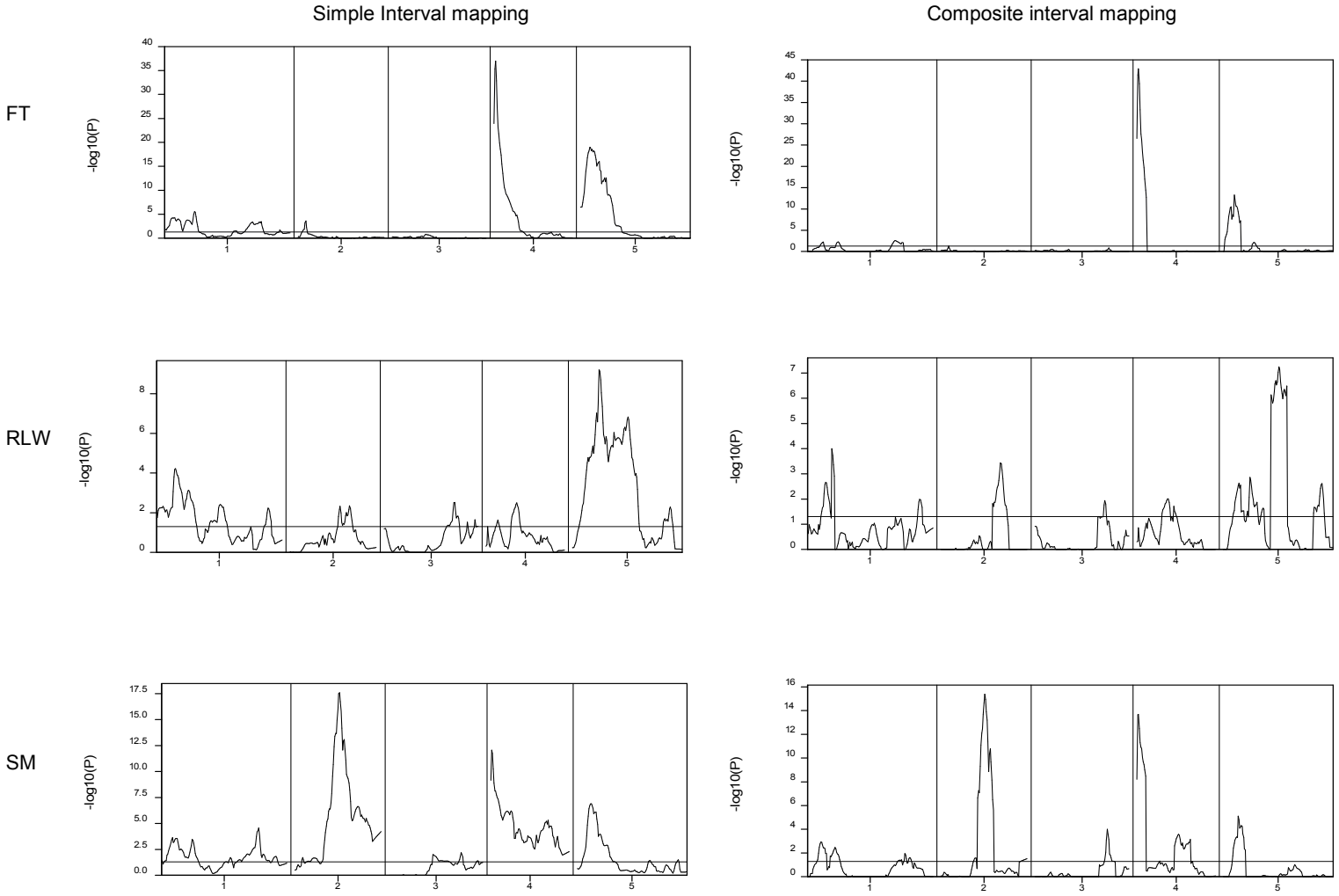
QTL mapping procedure

- Simple Interval mapping: testing single QTL
  - $\underline{y}$ = Pop + {$\underline{X}$ + $\underline{X.Pop}$} + $\underline{residual}$
  - $X_{532 \times 8}$: IBD information; row = F4 line; column = parent/founder
  - For convenience the QTL effects are not written explicitly alongside the genetic design matrix
  - Residual is cross dependent
- Deviance test for QTLs and QTL by cross interactions
- Composite Interval Mapping
  - Model: $\underline{y}$ = Pop + {$\sum$ $\underline{QTL}$ + $\sum$ $\underline{QTL.Pop}$}+ {$\underline{X}$ + $\underline{X.Pop}$} + $\underline{residual}$
- Final model by back selection

Flowering time, SIM profile



QTL profiles obtained from Simple Interval Mapping (SIM) and from Composite Interval Mapping (CIM), for three traits.

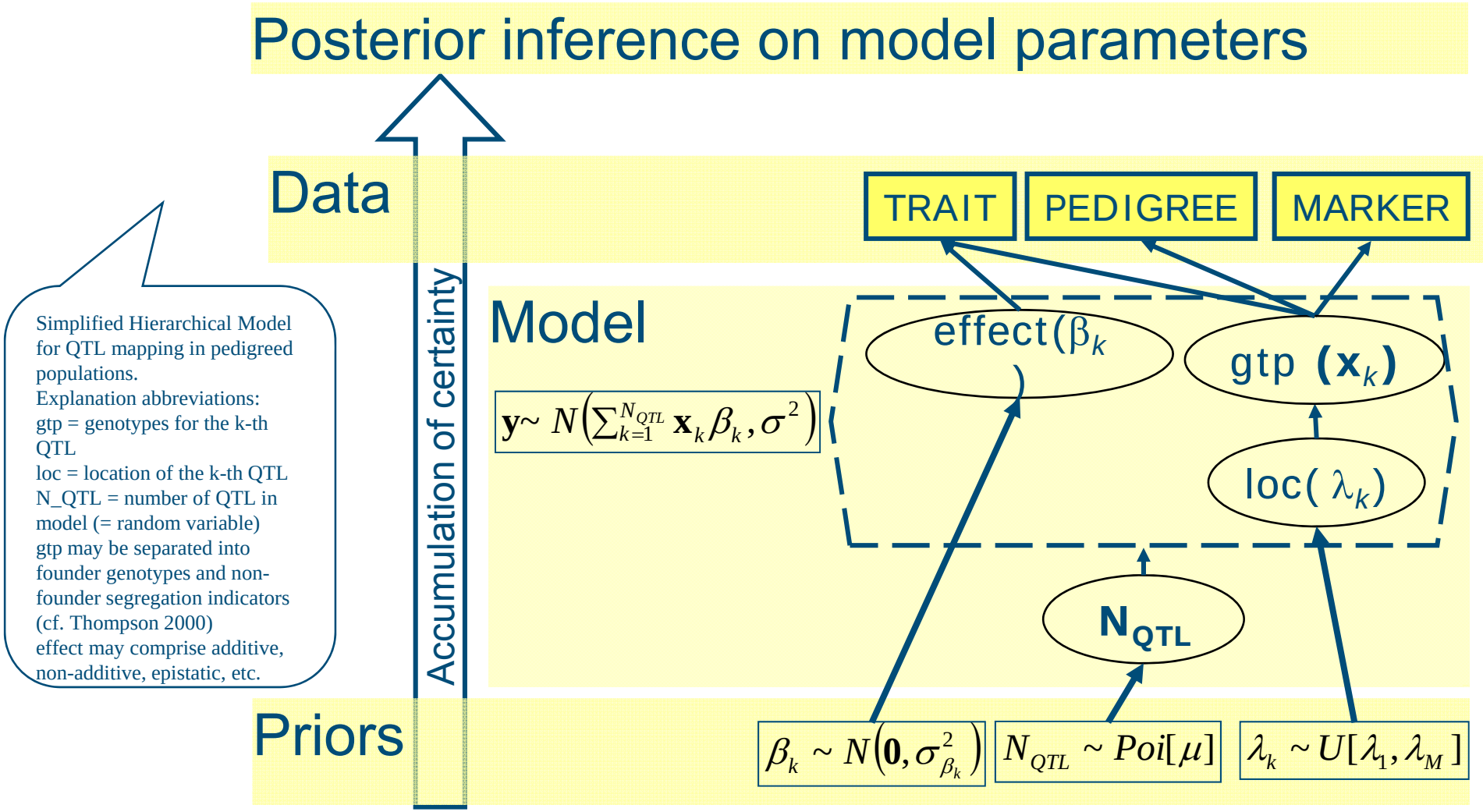


Effects and standard errors of four QTL retained in the final model for **flowering time** expressed in days, % genotypic variation explained

Chr	Pos	h^2_{QTL}	*Cross*	Cvi	Sha	Ler	C24	Col	Kyo	Eri	An	SE (min,max)
1	*39*	1.6%		0.43	-0.24	0.30	-1.73	-0.65	0.34	-0.50	1.99	(0.56,0.65)
1	*112*	3.5%		-1.31	1.78	0.11	-1.38	-0.59	2.73	0.09	-0.88	(0.62,0.79)
4	*2*	28.4%		-3.35	3.40	-2.58	4.30	-2.63	7.26	-1.60	-3.25	(1.38,1.49)
4[(a)]	2	2.8%	AB	-0.98	-0.55	-	-	-0.63	-1.93	-	-	(1.45,2.01)
			AC	-	-	-	-	-0.69	3.47	0.25	-0.94	
			AD	-	-	-1.49	0.84	0.70	0.18	-	-	
			BC	0.27	1.53	-	-	-	-	0.86	1.75	
			BD	-0.09	-0.17	1.80	-1.54	-	-	-	-	
			CD	-	-	-0.92	1.71	-	-	-1.49	-1.58	
5	*12*	17.5%		6.79	-3.98	-3.66	-1.30	5.47	-0.42	2.75	-1.23	(0.79,1.05)

(a) interaction effects of cross by QTL

Bayesian Analysis



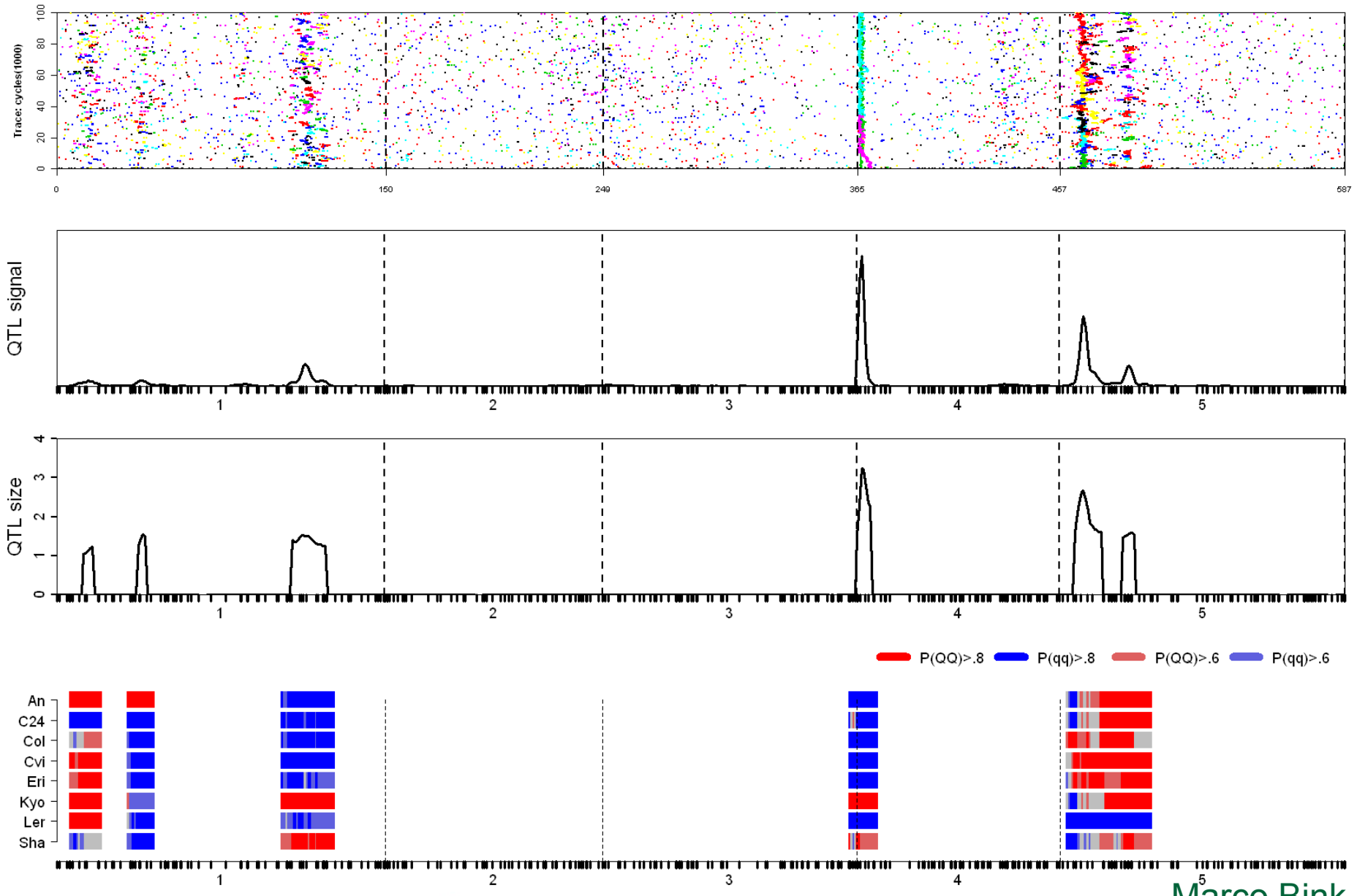
Marco Bink

M2

Simplified Hierarchical Model for QTL mapping in pedigreed populations.
Explanation abbreviations:
gtp = genotypes for the k-th QTL
loc = location of the k-th QTL
N_QTL = number of QTL in model (= random variable)
gtp may be separated into founder genotypes and non-founder segregation indicators (cf. Thompson 2000)
effect may comprise additive, non-additive, epistatic, etc.

Bink, 8-11-2009

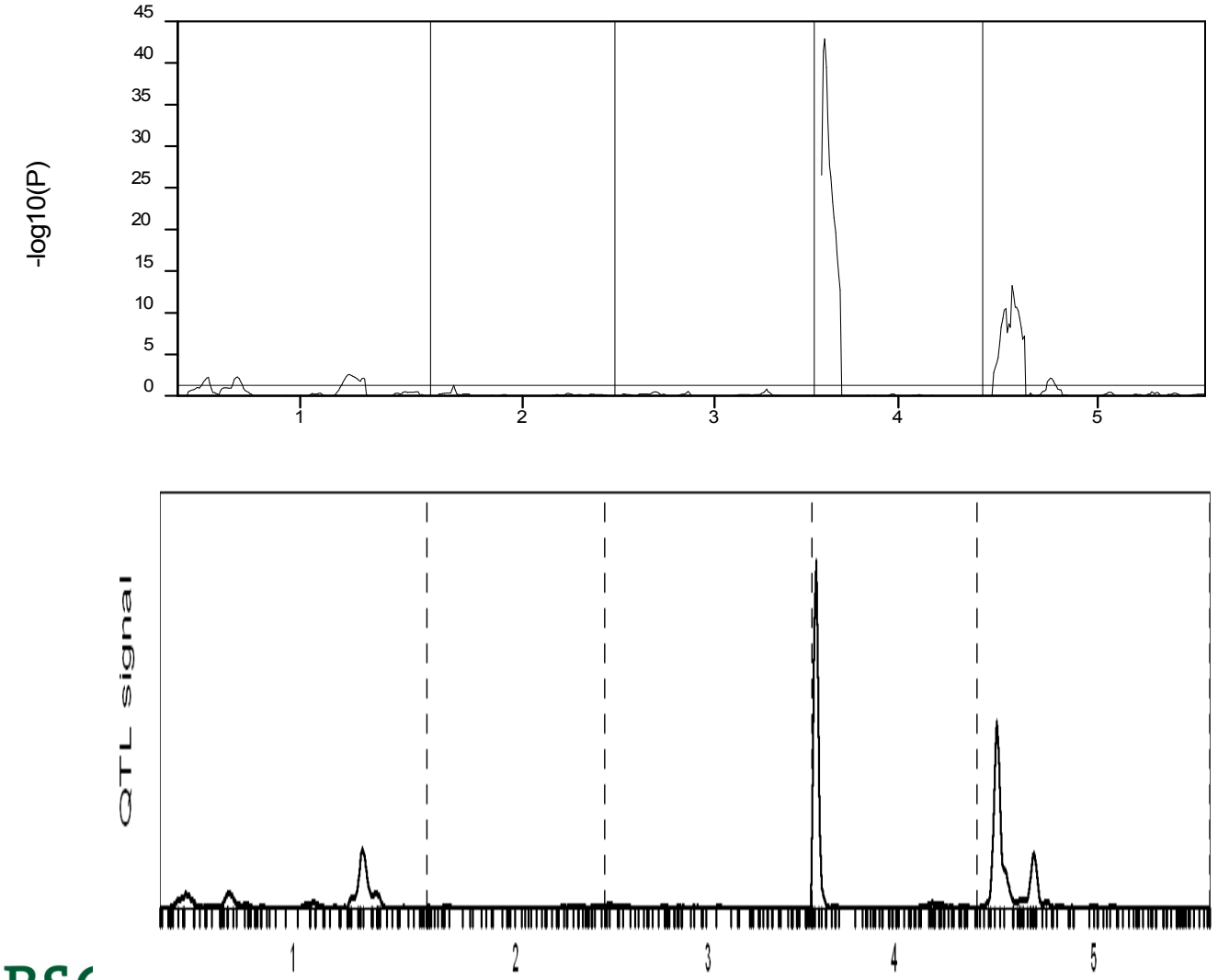
Bayesian QTL mapping: trait FT



Marco Bink

Comparing mixed model and Bayesian analysis

Composite interval mapping



Epistasis

- Start off with the final QTL subset, i.e.

  $\underline{y} = Pop + \underline{X}_{40} + \underline{X}_{113} + \underline{X}_{396} + \{\underline{X}_{396}.\underline{Pop}\} + \underline{X}_{507} + \underline{residual}$

- Test epistasis between identified QTLs i and j

  $\underline{y} = cross + \underline{X}_{40} + \underline{X}_{113} + (\underline{X}_{396} + \underline{X}_{396}.\underline{Pop}) + \underline{X}_{507}$
  $+ \underline{X}_i + \underline{X}_j + \underline{X}_i \cdot \underline{X}_j + \underline{residual}$

- Test the epistatic effect corresponding to the product matrix $\underline{X}_i \cdot \underline{X}_j$ via deviance test for variance component

Significant epistatic interactions in the AMPRIL population

Trait	QTL_i (chr,pos)	QTL_j (chr,pos)	-log10(P)	$h^2_{QTLxQTL}$
FT	(1,112)	(4,2)	1.98	1.7%
FT	(4,2)	(5,12)	3.12	2.6%
TLN	(1,121)	(4,2)	1.45	1.4%
TLN	(4,2)	(5,13)	1.97	2.5%
RLN	(1,19)	(1,121)	1.44	1.2%
RLN	(1,121)	(4,2)	1.60	1.4%
RLN	(4,2)	(5,13)	1.52	2.2%
CLN	(4,2)	(5,14)	2.23	2.6%
LA	(1,17)	(3,56)	1.60	2.4%
LW	(1,123)	(2,83)	1.37	0.7%
PW	(1,18)	(1,144)	2.03	2.3%
PW	(3,56)	(5,115)	1.37	1.7%
RPA	(1,37)	(2,55)	2.25	3.2%
SM	(2,56)	(3,92)	1.51	2.0%
SM	(2,56)	(4,2)	1.46	1.6%
TBN	(1,121)	(4,2)	2.39	2.3%
RBN	(1,123)	(4,2)	1.60	2.2%
RBN	(4,2)	(5,13)	2.70	4.8%
SBN	(4,2)	(5,13)	2.10	2.5%

Conclusions use AMPRIL

- Power: QTLs with explained variance below 2% detected
- Epistasis: effects with explained variance between 1 and 2% were detected
- Precision: expected nr of cross-overs 3.625
  - (larger than for 2-way RIL, 2.5; 4-way RIL, 3.0)
- Confirmation of detected: by HIFs

Multi-population QTL analysis: Case 2
NAM population in maize (Buckler)

- Marcos Malosetti
- Amy Jacobson
- Joao Paulo
- Martin Boer
- Fred van Eeuwijk

NAM populations

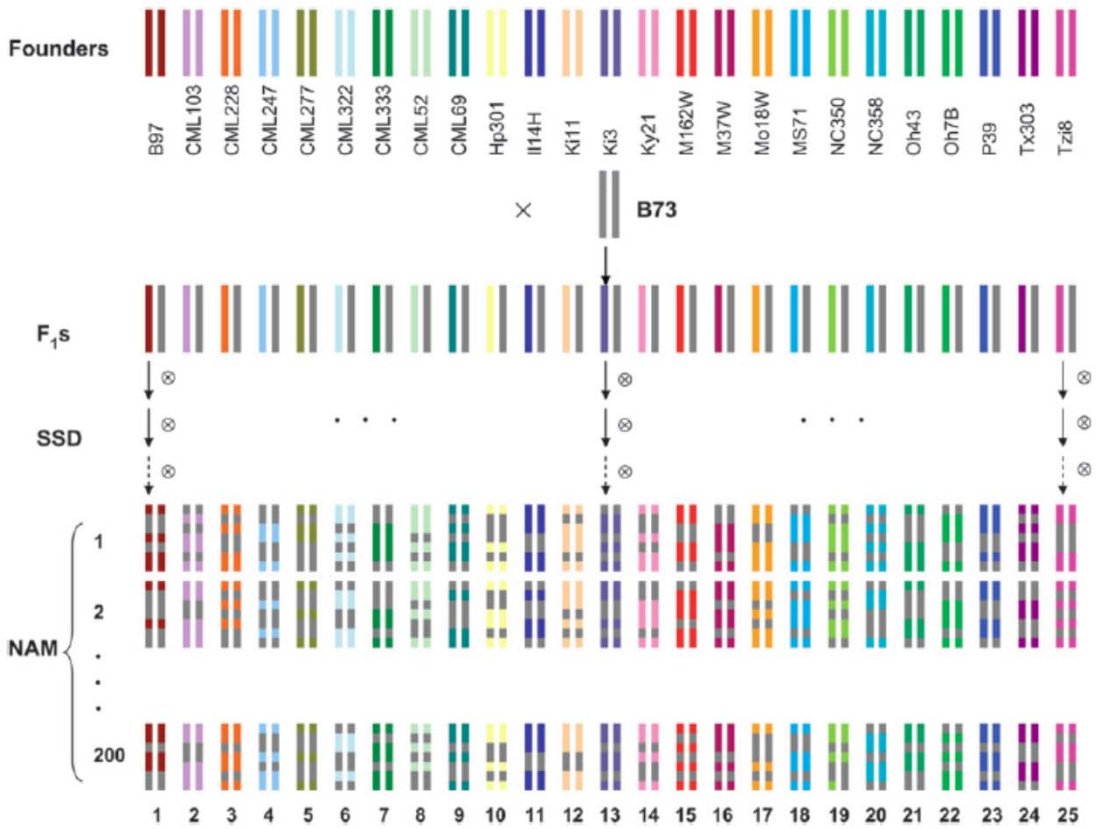


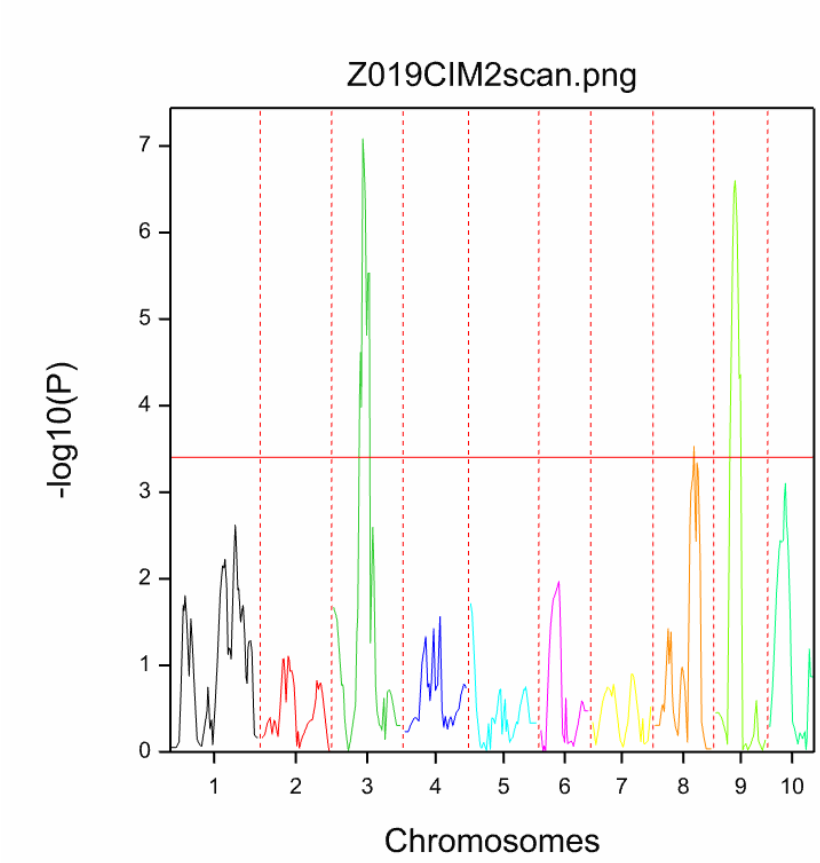
FIGURE 1.—Diagram of genome reshuffling between 25 diverse founders and the common parent and the resulting 5000 immortal genotypes. Due to diminishing chances of recombination over short genetic distance and a given number of generations, the genomes of these recombinant inbred lines (RILs) are essentially mosaics of the founder genomes. ×, crossing; ⊗, selfing; SSD, single-seed descent.

QTL mixed model for multiple environments (QTLxE)

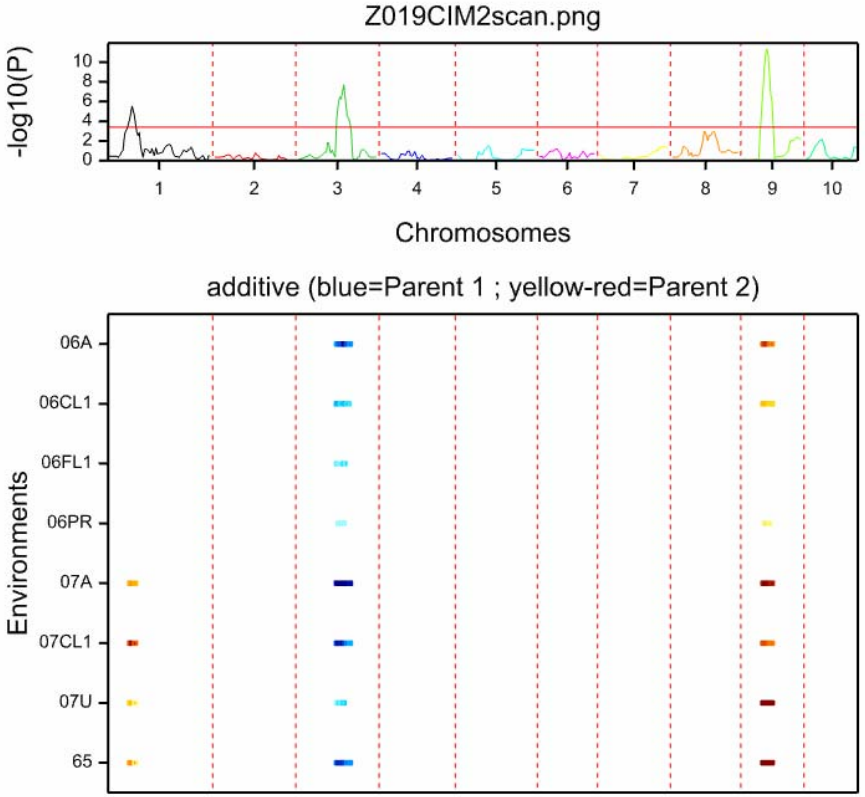
- Phenotype =
  - Environment mean +
  - Environment specific QTLs +
  - Environment specific residual genetic effect +
  - Environment specific error

$$\underline{P}_{ij} = \mu_j + \sum x_i a_j + \underline{G}_{ij} + \underline{\varepsilon}_{ij}$$

Ear height: population 19 (B73xMS71)



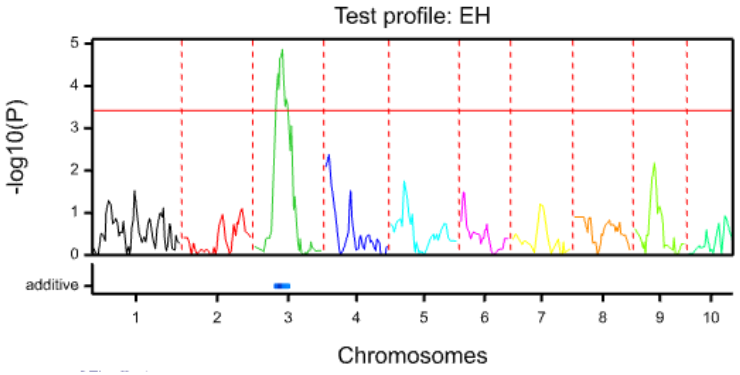
Main Effects analysis for EH



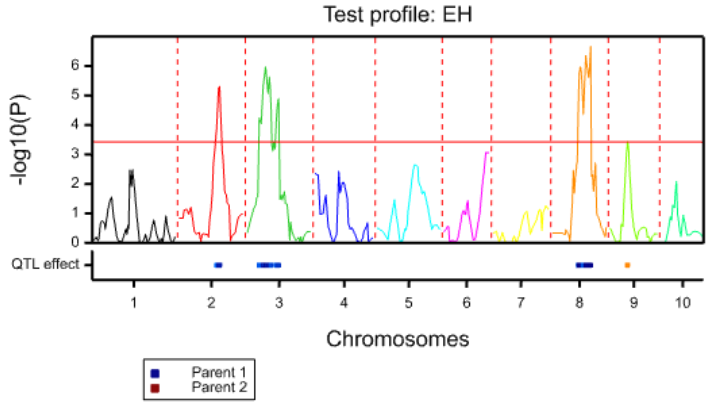
QTL+QTLxE analysis for EH
(CIM; VCOV = FA model)

Population 24 (B73xP39)

Single Trait Single Environment QTL analysis

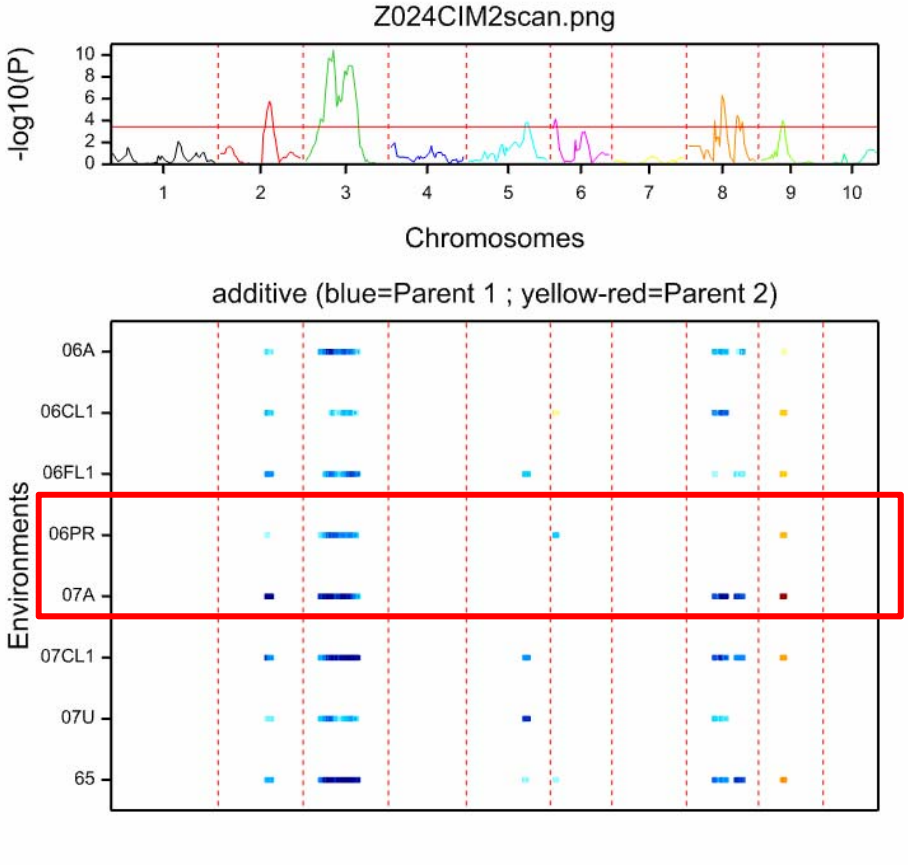


06PR



07A

QTL+QTLxE analysis for EH (CIM; VCOV = FA model)

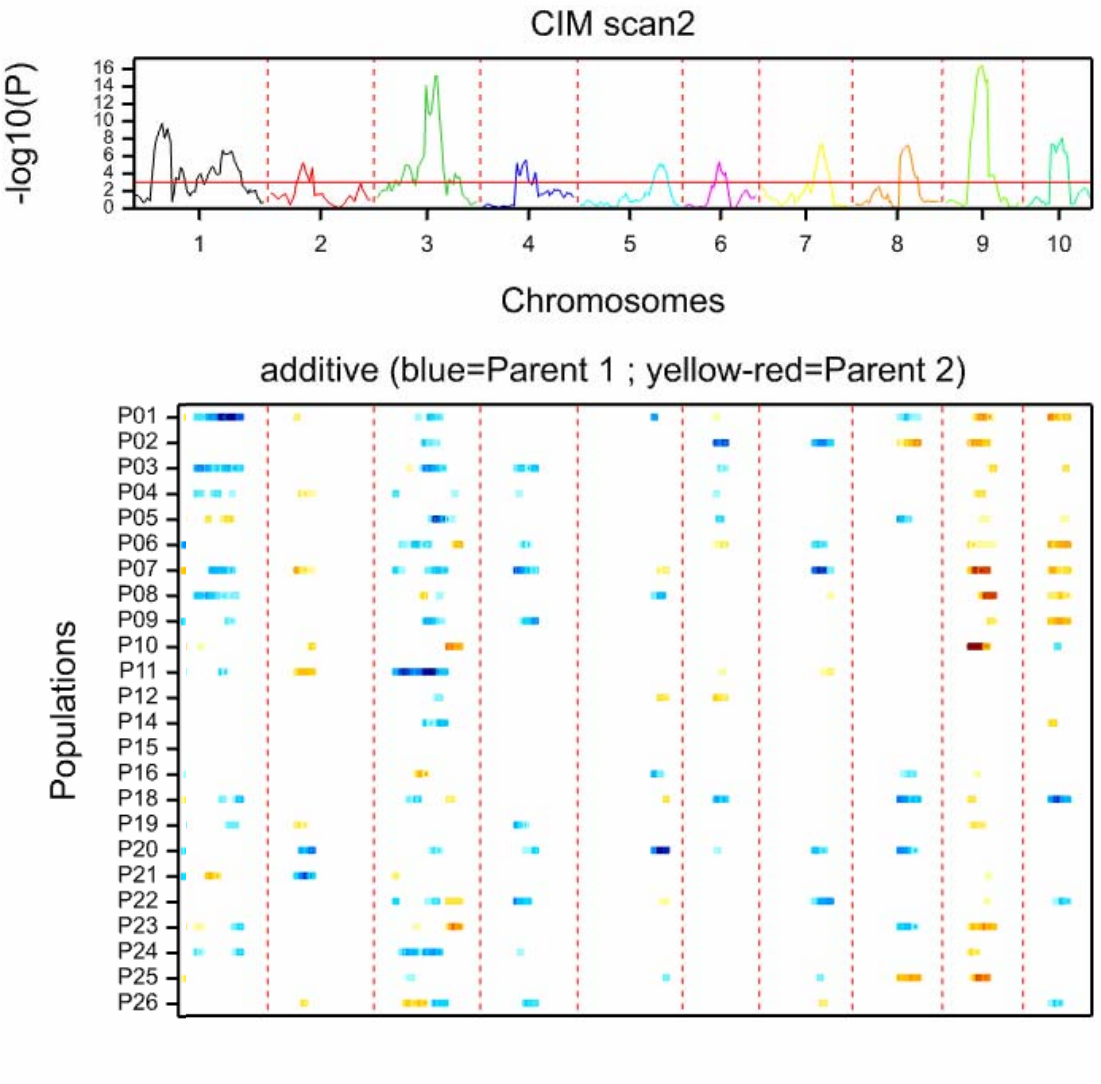


Multiple populations in single environment

- Linkage Analysis
- Phenotype =
  - Population Mean +
  - Population specific QTLs +
  - Population specific residual genetic effect +
  - Population specific error

$$\underline{P}_{ik} = \mu_k + \sum x_i a_k + (\underline{G}_{i \subset k} + \underline{\varepsilon}_{i \subset k})$$

Linkage Analysis for Environment 06PR



Multiple populations in multiple environments

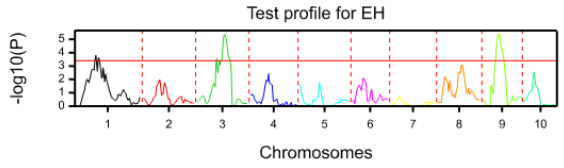
- Linkage Analysis
- Phenotype =

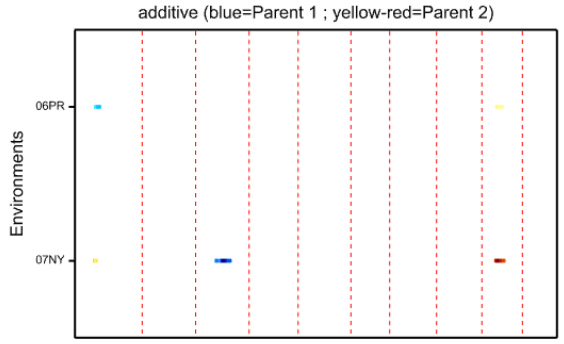
$$\underline{P}_{ijk} = \mu_{jk} + \sum x_i a_{jk} + \{\underline{G}_{i(jk)} + \underline{\varepsilon}_{i(jk)}\}$$

  - Pop.Env Mean +
  - Pop.Env specific QTLs +
  - Pop.Env specific residual genetic effect +
  - Pop.Env specific error

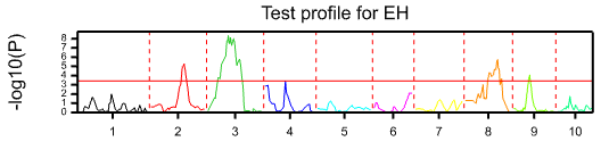
  - Residual genetic effect and error: unstructured VCOV

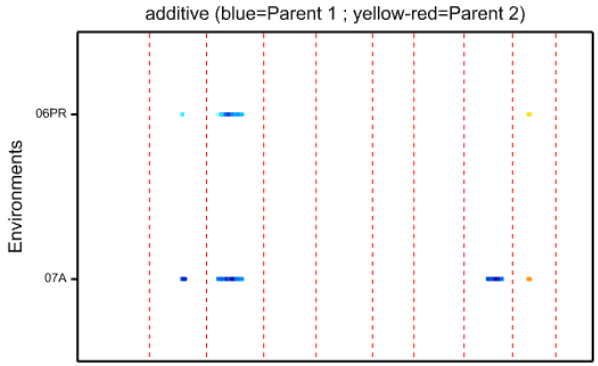
Multi-Population Multi-Environment



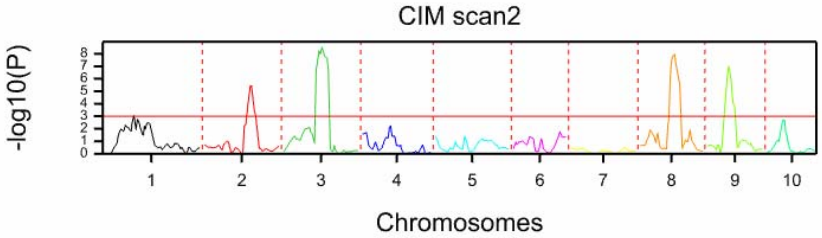


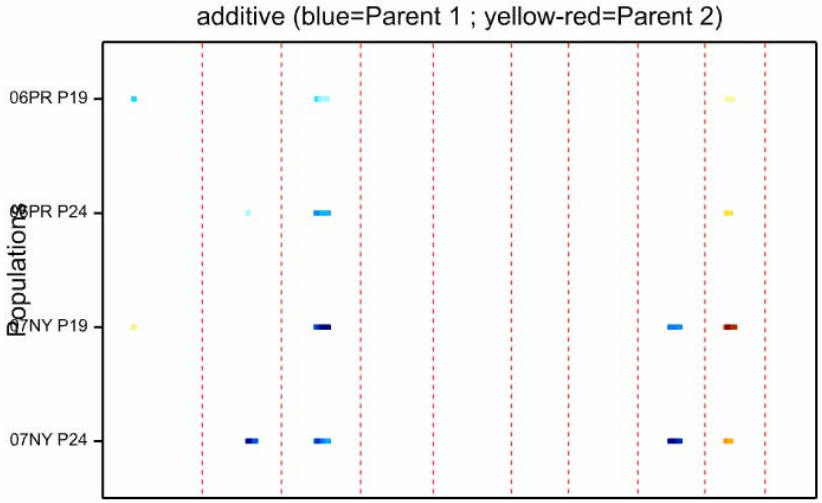
Population 19





Population 24





Conclusions

- Mixed model QTL mapping is a flexible tool to analyse multi-population data
  - Set-up design vectors/matrices to represent genetic data (IBD/IBS)
  - Test for environment and/or population specific QTLs (fixed/ random)
  - Structure residual genetic effects (environment, population) in appropriate ways for valid testing
  - Take care to model within environment/ trial error structure (randomization based + spatial)
  - Use of multiple populations offers straightforward way of testing QTL x background interactions