Multiple Correlated Traits

- Pleiotropy vs. close linkage
- Analysis of covariance
  - Regress one trait on another before QTL search
- Classic GxE analysis
- Formal joint mapping (MTM)
- Seemingly unrelated regression (SUR)
- Reducing many traits to one
  - Principle components for *similar* traits

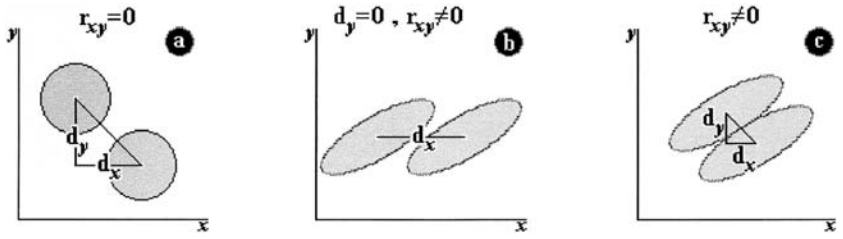
co-mapping multiple traits

- avoid reductionist approach to biology
  - address physiological/biochemical mechanisms
  - Schmalhausen (1942); Falconer (1952)
- separate close linkage from pleiotropy
  - 1 locus or 2 linked loci?
- identify epistatic interaction or canalization
  - influence of genetic background
- establish QTL x environment interactions
- decompose genetic correlation among traits
- increase power to detect QTL

Two types of data

- Design I: multiple traits on same individual
  - Related measurements, say of shape or size
  - Same measurement taken over time
  - Correlation within an individual
- Design II: multiple traits on different individuals
  - Same measurement in two crosses
  - Male vs. female differences
  - Different individuals in different locations
  - No correlation between individuals

interplay of pleiotropy & correlation



pleiotropy only *correlation only* *both*

Korol et al. (2001)

Brassica napus: 2 correlated traits

- 4-week & 8-week vernalization effect
  - log(days to flower)
- genetic cross of
  - Stellar (annual canola)
  - Major (biennial rapeseed)
- 105 F1-derived double haploid (DH) lines
  - homozygous at every locus (*QQ* or *qq*)
- 10 molecular markers (RFLPs) on LG9
  - two QTLs inferred on LG9 (now chromosome N2)
  - corroborated by Butruille (1998)
  - exploiting synteny with *Arabidopsis thaliana*

QTL with GxE or Covariates

- adjust phenotype by covariate
  - covariate(s) = environment(s) or other trait(s)
- additive covariate
  - covariate adjustment same across genotypes
  - "usual" analysis of covariance (ANCOVA)
- interacting covariate
  - address GxE
  - capture genotype-specific relationship among traits
- another way to think of multiple trait analysis
  - examine single phenotype adjusted for others

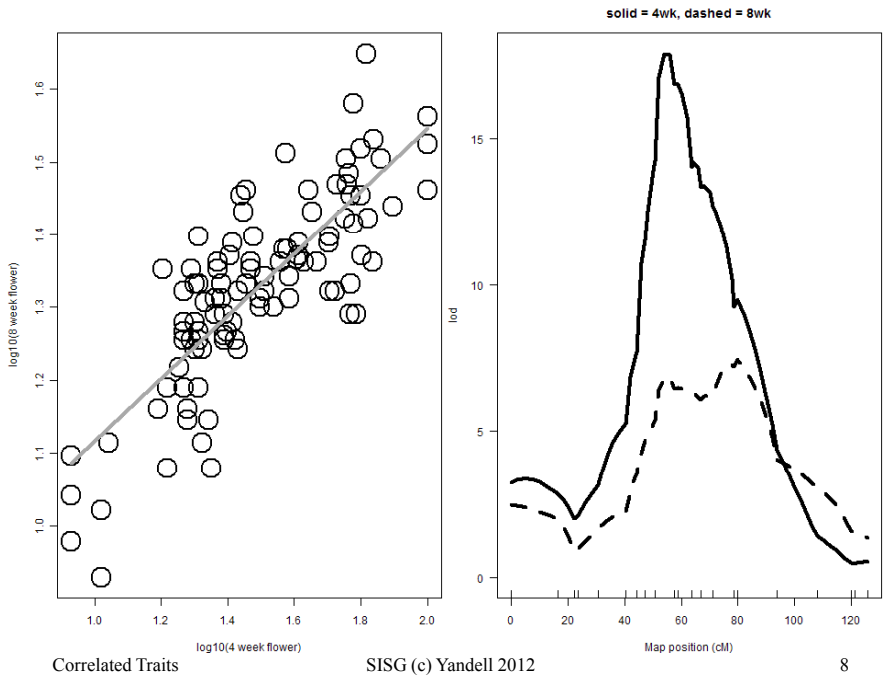
R/qtl & covariates

- additive and/or interacting covariates
- test for QTL after adjusting for covariates

```
## Get Brassica data.
library(qtlbim)
data(Bnapus)
Bnapus <- calc.genoprob(Bnapus, step = 2, error = 0.01)

## Scatterplot of two phenotypes: 4wk & 8wk flower time.
plot(Bnapus$pheno$log10flower4,Bnapus$pheno$log10flower8)

## Unadjusted IM scans of each phenotype.
fl8 <- scanone(Bnapus,, find.pheno(Bnapus, "log10flower8"))
fl4 <- scanone(Bnapus,, find.pheno(Bnapus, "log10flower4"))
plot(fl4, fl8, chr = "N2", col = rep(1,2), lty = 1:2,
  main = "solid = 4wk, dashed = 8wk", lwd = 4)
```

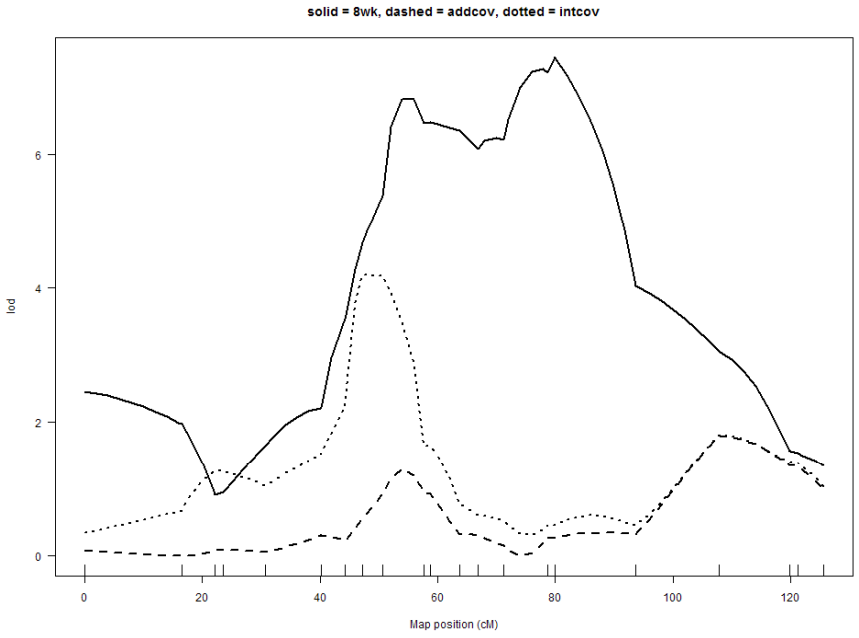
R/qtl & covariates

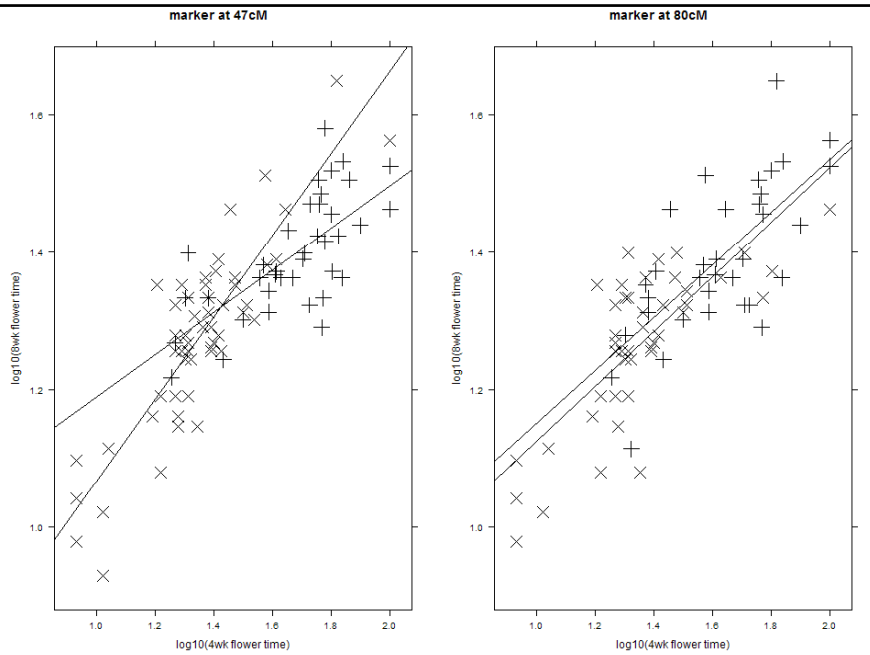
- additive and/or interacting covariates
- test for QTL after adjusting for covariates

```
## IM scan of 8wk adjusted for 4wk.
## Adjustment independent of genotype
fl8.4 <- scanone(Bnapus,, find.pheno(Bnapus, "log10flower8"),
  addcov = Bnapus$pheno$log10flower4)

## IM scan of 8wk adjusted for 4wk.
## Adjustment changes with genotype.
fl8.4 <- scanone(Bnapus,, find.pheno(Bnapus, "log10flower8"),
  intcov = Bnapus$pheno$log10flower4)

plot(fl8, fl8.4a, fl8.4, chr = "N2",
  main = "solid = 8wk, dashed = addcov, dotted = intcov")
```



scatterplot adjusted for covariate

```
## Set up data frame with peak markers, traits.
markers <- c("E38M50.133","ec2e5a","wg7f3a")
tmpdata <- data.frame(pull.geno(Bnapus)[,markers])
tmpdata$fl4 <- Bnapus$pheno$log10flower4
tmpdata$fl8 <- Bnapus$pheno$log10flower8

## Scatterplots grouped by marker.
library(lattice)
xyplot(fl8 ~ fl4, tmpdata, group = wg7f3a,
  col = "black", pch = 3:4, cex = 2, type = c("p","r"),
  xlab = "log10(4wk flower time)",
  ylab = "log10(8wk flower time)",
  main = "marker at 47cM")
xyplot(fl8 ~ fl4, tmpdata, group = E38M50.133,
  col = "black", pch = 3:4, cex = 2, type = c("p","r"),
  xlab = "log10(4wk flower time)",
  ylab = "log10(8wk flower time)",
  main = "marker at 80cM")
```

Multiple trait mapping

- Joint mapping of QTL
  - testing and estimating QTL affecting multiple traits
- Testing pleiotropy vs. close linkage
  - One QTL or two closely linked QTLs
- Testing QTL x environment interaction
- Comprehensive model of multiple traits
  - Separate genetic & environmental correlation

Formal Tests: 2 traits

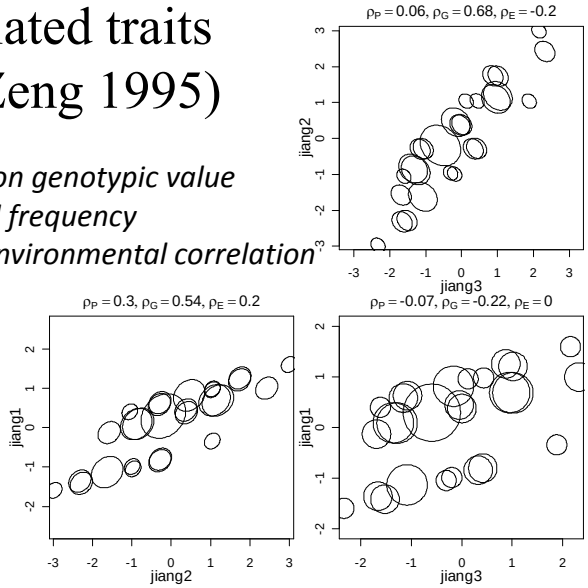
$y_1 \sim N(\mu_{q1}, \sigma^2)$ for group 1 with QTL at location λ_1
$y_2 \sim N(\mu_{q2}, \sigma^2)$ for group 2 with QTL at location λ_2

- Pleiotropy vs. close linkage
  - test QTL at same location: $\lambda_1 = \lambda_2$
  - likelihood ratio test (LOD): null forces same location
- if pleiotropic ($\lambda_1 = \lambda_2$)
  - test for same mean: $\mu_{q1} = \mu_{q2}$
  - Likelihood ratio test (LOD)
    - null forces same mean, location
    - alternative forces same location
  - only make sense if traits are on same scale
  - test sex or location effect

3 correlated traits (Jiang Zeng 1995)

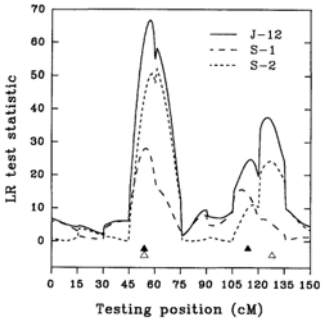
ellipses centered on genotypic value
width for nominal frequency
main axis angle environmental correlation
3 QTL, F2
27 genotypes

note signs of genetic and environmental correlation



pleiotropy or close linkage?

2 traits, 2 qtl/trait
pleiotropy @ 54cM
linkage @ 114,128cM
Jiang Zeng (1995)



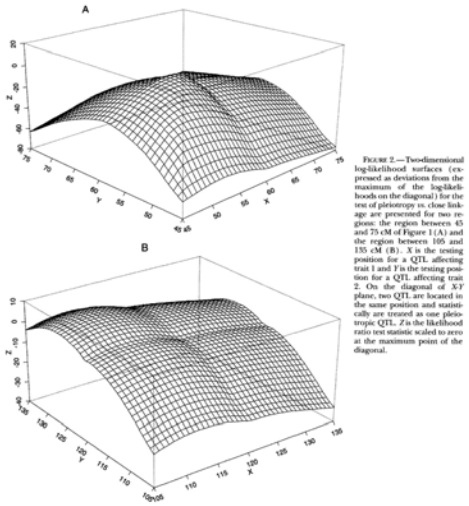


FIGURE 2.—Two-dimensional log-likelihood surfaces (expressed as deviations from the maximum of the log-likelihoods on the diagonal) for the test of pleiotropy vs. close linkage are presented for two regions: the region between 45 and 75 cM of Figure 1(A) and the region between 105 and 135 cM (B). X is the testing position for a QTL affecting trait 1 and Y is the testing position for a QTL affecting trait 2. On the diagonal of X/Y plane, two QTL are located in the same position and statistically are treated as one pleiotropic QTL. Z is the likelihood ratio test statistic scaled to zero at the maximum point of the diagonal.

More detail for 2 traits

$y_1 \sim N(\mu_{q1}, \sigma^2)$ for group 1

$y_2 \sim N(\mu_{q2}, \sigma^2)$ for group 2

- two possible QTLs at locations λ_1 and λ_2
- effect β_{kj} in group k for QTL at location λ_j

$\mu_{q1} = \mu_1 + \beta_{11}(q_1) + \beta_{12}(q_2)$

$\mu_{q2} = \mu_2 + \beta_{21}(q_1) + \beta_{22}(q_2)$

- classical: test $\beta_{kj} = 0$ for various combinations

seemingly unrelated regression (SUR)

$\mu_{q1} = \mu_1 + \gamma_{11}\beta_{q11} + \gamma_{12}\beta_{q12}$

$\mu_{q2} = \mu_2 + \gamma_{21}\beta_{q21} + \gamma_{22}\beta_{q22}$

indicators γ_{kj} are 0 (no QTL) or 1 (QTL)

- include γs in formal model selection

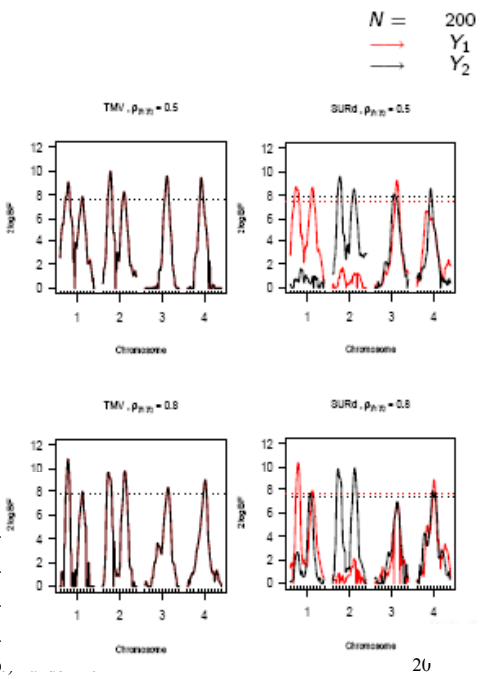
SUR for multiple loci across genome

- consider only QTL at pseudomarkers (lecture 2)
- use loci indicators γ_j (=0 or 1) for each pseudomarker
- use SUR indicators γ_{kj} (=0 or 1) for each trait
- Gibbs sampler on both indicators
  - Banerjee, Yandell, Yi (2008 *Genetics*)

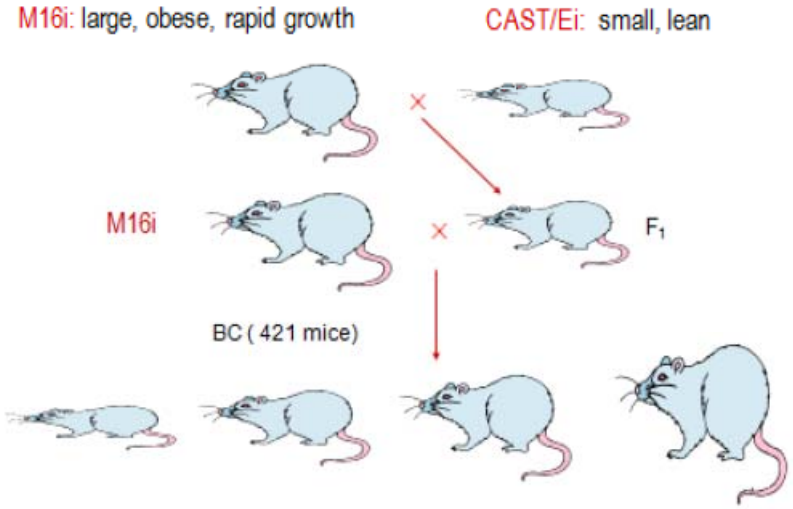
$$\mu_{q1} = \mu_1 + \gamma_1\gamma_{11}\beta_{11}(q_1) + \gamma_2\gamma_{12}\beta_{12}(q_2) + \ldots$$

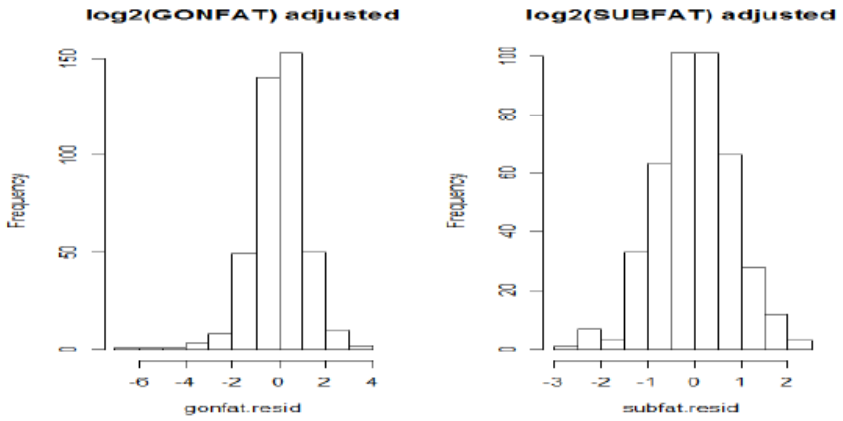
$$\mu_{q2} = \mu_2 + \gamma_1\gamma_{21}\beta_{21}(q_1) + \gamma_2\gamma_{22}\beta_{22}(q_2) + \ldots$$

Simulation
5 QTL
2 traits
n=200
TMV vs. SUR

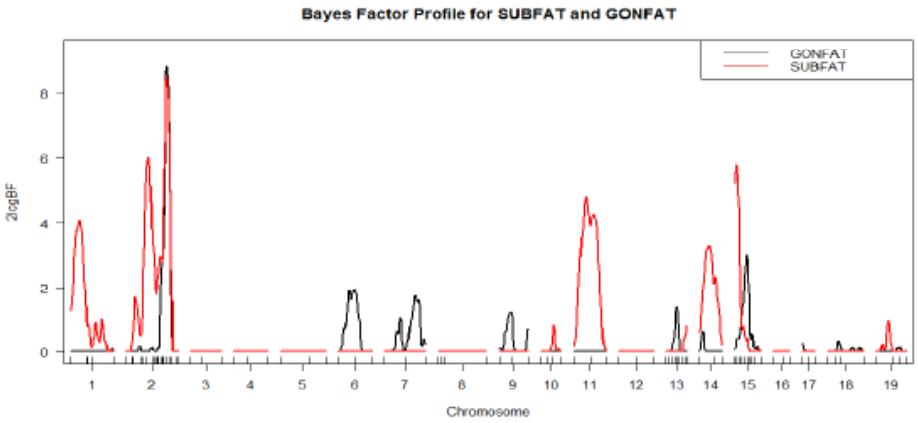


	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6
Chr	1	1	2	2	3	4
Pos(cM)	22	55	22	65	65	45
y_1	0.8	0.6	0	0	0.8	0.6
y_2	0	0	-0.8	-0.6	0.8	0.6
y_1	8.8%	4.9%	0	0	8.8%	4.9%
y_2	0	0	9.3%	5.2%	9.3%	5.2%





- GONFAT → Right Gonadal fat pad
- SUBFAT → Subcutaneous fat pad



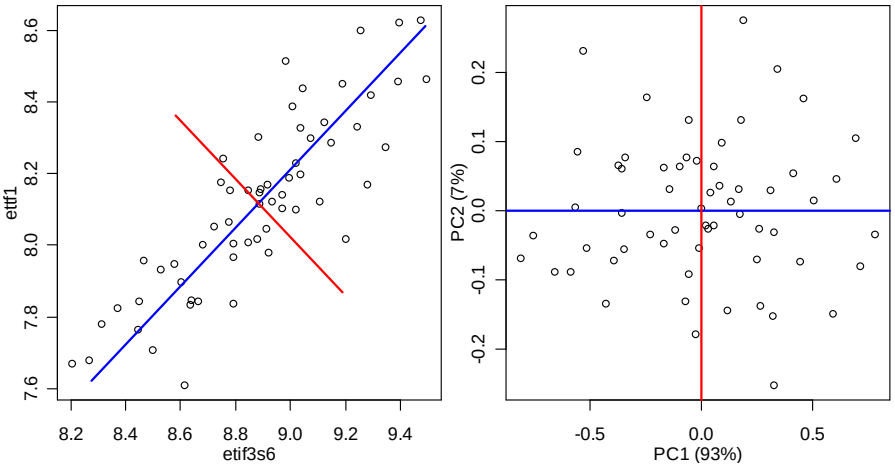
R/qtlbim and GxE

- similar idea to R/qtl
  - fixed and random additive covariates
  - GxE with fixed covariate
- multiple trait analysis tools coming soon
  - theory & code mostly in place
  - properties under study
  - expect in R/qtlbim later this year
  - Samprit Banerjee (N Yi, advisor)

reducing many phenotypes to 1

- *Drosophila mauritiana* x *D. simulans*
  - reciprocal backcrosses, ~500 per bc
- response is "shape" of reproductive piece
  - trace edge, convert to Fourier series
  - reduce dimension: first principal component
- many linked loci
  - brief comparison of CIM, MIM, BIM

PC for two correlated phenotypes



shape phenotype via PC

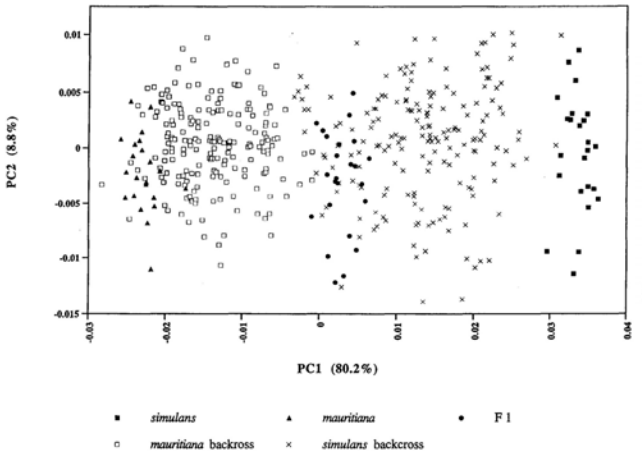


FIGURE 5.—A plot of the first two principal components of the Fourier coefficients from posterior lobe outlines. Many individuals from each of five genotypic classes are represented. Each point represents an average of scores from the left and right sides of an individual (with a few exceptions for which the score is from one side only). The percentage of variation in the Fourier coefficients accounted for by each principal component is given in parentheses. *Liu et al. (1996) Genetics*

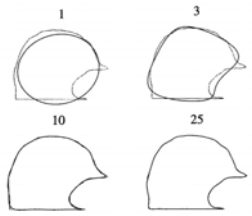


FIGURE 2.—The effect of harmonic number on the accuracy of reconstruction of a posterior lobe outline by elliptical Fourier analysis.

shape phenotype in BC study indexed by PC1

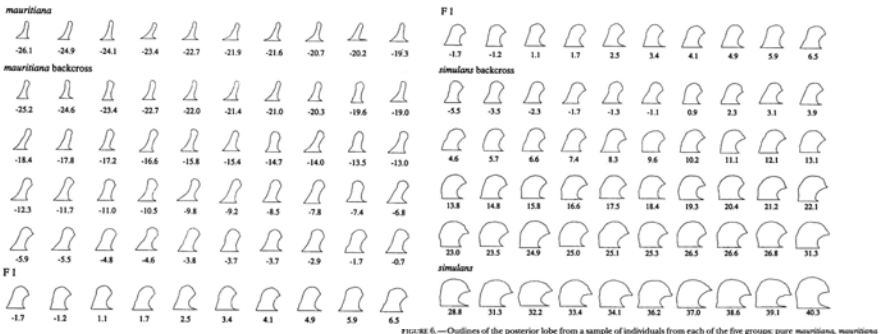


FIGURE 6.—Outlines of the posterior lobe from a sample of individuals from each of the five groups: pure *mauritiana*, *mauritiana* backcross, F_1, *simulans* backcross, and pure *simulans*. Within each group, the outlines are presented in order of their PC1 score (sampled at even intervals from the range of variation). The number below each specimen is its PC1 score. The outlines are drawn to scale with the origin at the centroid of each outline and with all baselines parallel.

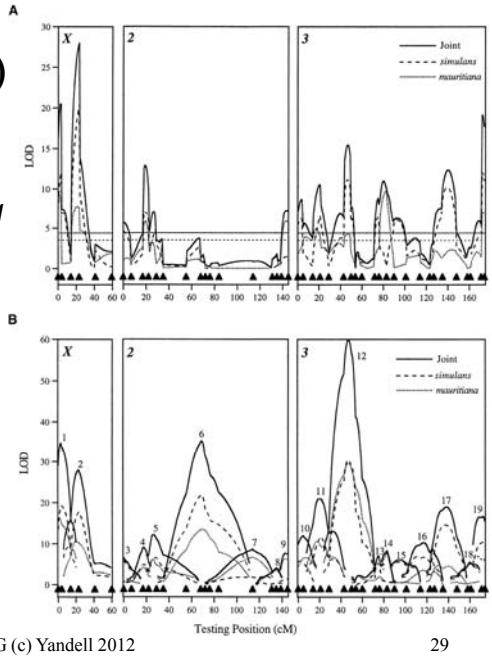
Liu et al. (1996) Genetics

Zeng et al. (2000) CIM vs. MIM

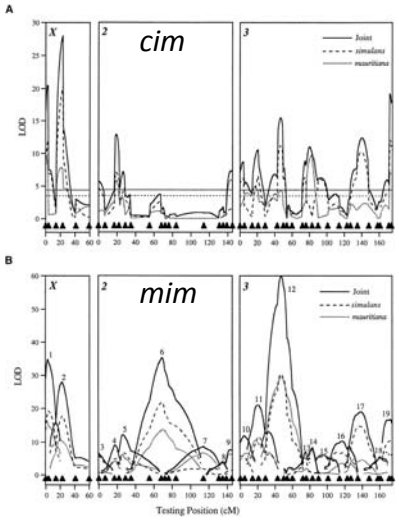
composite interval mapping
(Liu et al. 1996)
narrow peaks
miss some QTL

multiple interval mapping
(Zeng et al. 2000)
triangular peaks

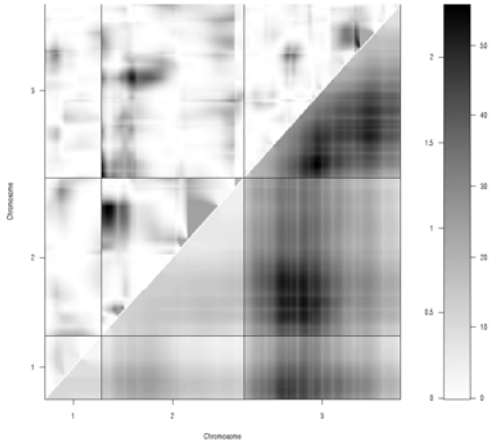
both conditional 1-D scans
fixing all other "QTL"



CIM, MIM and IM pairscan



2-D im



multiple QTL: CIM, MIM and BIM