An Examination of Quantitative Traits in *Brassica*

Brian S. Yandell*
Departments of Horticulture and Statistics
University of Wisconsin–Madison

*Research supported by USDA-CSRS grant 511-100.

Outline

- *Brassica* study
- linkage maps
- classical *vs.* mapmaker QTL
- testing with likelihood ratios
- confidence interval for major QT gene
- major and minor QT genes
- stepwise location of QT loci

Brassica study at UW–Madison
Tom Osborn and Keming Song, Agronomy

95 families
297 RFLP markers
24 (20) quantitative traits

- Song, Suzuki, Slocum, Williams and Osborn (1990)

- Song, Slocum and Osborn (1990)

Testing for join in Linkage Maps

two strands *vs.* one?

$LOD(one) - LOD(two) \sim \chi^2_1/2log(10)$

level	LOD
.05	0.834
.01	1.441
.001	2.351
.0001	3.287
.00001	4.237

multiple comparisons problem: doing many tests

with 100 markers there are about 4950 tests possible

exptwise error rate $= 1 - (1 - \alpha)^{4950}$

$=$ .39 for $\alpha =$ .0001

$=$ .05 for $\alpha =$ .00001

one actually looks for MLE of whole map

Testing for QTLs

$$trait = \left\{ \begin{array}{lll} \text{mean} & & \text{alleles} \\ \mu + a & & AA \\ \mu + d & if & ABorBA \\ \mu - a & & BB \end{array} \right\} + error$$

$$y_i = \mu + ax_i + d(1 - |x_i|) + \epsilon_i$$

μ = reference mean
a = additive effect
d = dominance effect
x_i = +1 (AA), 0 (AB), –1 (BB)
$\epsilon_i \sim N(0, \sigma^2)$

null hypothesis: no QTL ($a = d = 0$)

nuisance parameter:
position along chromosome t
recombinant frequency btw. flanking markers

$\widehat{y}_i = \widehat{\mu} + \widehat{a}x_i + \widehat{d}(1 - |x_i|)$
$SSModel = \sum_i (\bar{y} - \widehat{y}_i)^2$
$SSError = \sum_i (y_i - \widehat{y}_i)^2 = (n - 3)\widehat{\sigma}^2$

$F = SSModel/2\widehat{\sigma}^2 \sim F_{2,n-3}]$

Distribution of LODs

$L(\hat{a}, \hat{d}) = \log_{10} Pr\{trait|\hat{model}\}$
$= \frac{1}{2\hat{\sigma}^2} SSError / \log(10) +$ terms in n, σ^2

$LOD = L(0,0) - L(\hat{a}, \hat{d})$
$= \frac{1}{2} SSModel / (\hat{\sigma}^2 \log(10))$
$\approx \chi^2_2 / (2 \log(10))$

luck! LOD=2 for 1% level test

level	LOD
.05	1.301
.01	2.000
.001	3.000
.0001	4.000
.00001	5.000

counts and generalized linear models:
use $LOD = deviance / \log(10)$

Confidence Interval for Major QT locus

pointwise estimate at ML peak
must have $LOD > 2$ at peak
to be significant at 1% level
99% confidence interval

$\{t \ ; \ \max_t(LOD(\hat{a}, \hat{d})) - LOD(\hat{a}, \hat{d}|t) > 2\}$

equivalent to normal interval in additive case
based on profile likelihood
need locally quadratic LOD

profile actually quite bumpy
less (?) information at markers

CI may include regions on other chromosomes

Confidence Intervasl for Major and Minor QT loci

Major Gene

confidence interval for major gene

twin genes close together:
difficult to distinguish
could use χ^2 test with df=4

Minor Gene on another chromosome

add LODs for both chromosomes
2-D contour map
1-D look conditional on major gene
minor gene must add signif. amount
test with χ^2 at df=2

problem:
chromosomes not independent
due to nature of sample

Stepwise Location of QTLs

Finding loci

remove effect of major gene
locate max conditional LOD across genome
repeat in stepwise fashion

when to stop?
same issues as stepwise regression

Effect on Estimates

additive vs. dominance near MLE
very nonlinear
note sensitivity to loci
on other chromosomes