

QTL Mapping

CS741
2009
Jim Holland

What is a QTL?

- Quantitative Trait Locus
- Locus, meaning region of the genome – not necessarily a single gene, could be several linked genes.
- Allelic variation at a QTL region causes phenotypic variation in a quantitative trait.

Objectives of QTL Analysis

1. Identify regions of the genome containing QTL
2. Estimate the effects of the QTL on the quantitative trait:
 - how much of the variation for the trait is caused by a specific region?
 - what is the gene action associated with the QTL - additive effect? Dominant effect?
 - which allele is associated with the favorable effect?
3. Use QTL estimates and genotypic information to assign breeding values to lines

QTL analyses

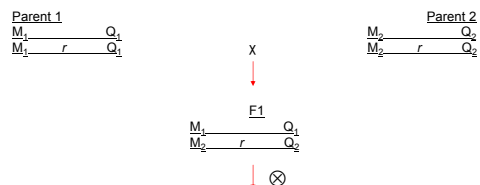
- Single marker ANOVA
- Multiple marker model building
- Interval mapping
- Composite Interval Mapping
- Multiple Interval Mapping
- Association Analysis

Simple Model of One QTL

- Genetic Model:

QTL genotype	Value
Q_1Q_1	$m+a$
Q_1Q_2	$m+d$
Q_2Q_2	$m-a$

One QTL, F2 Population



We can compare marker genotype class means (M_1M_1 vs M_1M_2 vs M_2M_2) to test if a QTL is linked to marker locus. But what does this test really mean?

To figure this out, we need a model to relate marker genotypes to QTL effects.

What are the F2 generation two-locus genotypes (for marker and QTL?)

2-Locus Gamete Frequencies

- Gamete** **Frequency**
- M_1 — Q_1 $1/2(1-r)$
- M_1 — Q_2 $1/2(r)$
- M_2 — Q_1 $1/2(r)$
- M_2 — Q_2 $1/2(1-r)$

F2 Genotype Frequencies and Values

Genotype	Genotypic Value	Frequency
$M_1M_1Q_1Q_1$	+a	$1/4 \cdot (1-r)^2$
$M_1M_1Q_1Q_2$	+d	$1/2 \cdot r(1-r)$
$M_1M_1Q_2Q_2$	-a	$1/4 \cdot r^2$
$M_1M_2Q_1Q_1$	+a	$1/2 \cdot r(1-r)$
$M_1M_2Q_1Q_2$	+d	$1/2 \cdot [(1-r)^2 + r^2]$
$M_1M_2Q_2Q_2$	-a	$1/2 \cdot r(1-r)$
$M_2M_2Q_1Q_1$	+a	$1/4 \cdot r^2$
$M_2M_2Q_1Q_2$	+d	$1/2 \cdot r(1-r)$
$M_2M_2Q_2Q_2$	-a	$1/4 \cdot (1-r)^2$

Expected Values of F₂ Marker Genotype Classes

Marker Genotype	Genotypic Value	Frequency	Estimated Phenotypic Mean
M_1M_1	$a[(1-r)^2 - r^2] + 2d[r(1-r)]$	1/4	$\bar{Y}_{M_1M_1}$
M_1M_2	$d[(1-r)^2 + r^2]$	1/2	$\bar{Y}_{M_1M_2}$
M_2M_2	$-a[(1-r)^2 - r^2] + 2d[r(1-r)]$	1/4	$\bar{Y}_{M_2M_2}$

Can perform ANOVA to test if variation among marker genotype classes is significant.

Use genotype class means to estimate additive and dominant effects.

What are we really comparing when we compare marker genotype class means?

- Additive effect estimate:

$$E[\bar{Y}_{M_1M_1} - \bar{Y}_{M_2M_2}] / 2 = a[(1-r)^2 - r^2] = a(1-2r)$$

- Dominant effect estimate:

$$E\left[\bar{Y}_{M_1M_2} - \frac{(\bar{Y}_{M_1M_1} + \bar{Y}_{M_2M_2})}{2}\right] = d[(1-r)^2 + r^2] - 2dr(1-r) = d(1-2r)^2$$

If the marker is unlinked to the QTL, $r = 0.5$ and the expected differences are zero. That's good!

But if the marker is linked to the QTL and $r > 0$, then the additive and dominant effects are underestimated!

Single Marker ANOVA Example

- Tomato F₂ population (333 individuals), 12 marker loci (2 linkage groups)

LINKAGE GROUP 1:

Locus T175-----C35-----T93-----C66-----T50B
 Map distance (cM) 4.0 15.0 11.9 12.2

LINKAGE GROUP 2:

Locus T24-----C15-----T125-----T71-----T83-----T209-----T17
 (cM) 14.8 6.4 18.9 24.0 18.1 28.6

Example Data Set

Plant	*T175	*T93	*C35	*T24	*C66	...etc...	weight
1	1	1	1	1	1		4.949
2	0	0	0	1	1		3.58
3	1	1	1	0	1		6.212
4	0	2	1	0	2		6.14
5	1	1	1	0	1		5.33
6	1	1	.	.	1		5.761
7	0	0	0	1	0		5.47
8		7.897
9	1	1	1	0	1		
...	etc...						
330	.	0	1	.	0		11.13
331	.	1	1	.	.		4.208
332	.	1	1	.	1		3.524
333

Is there a relationship between locus T24 and the phenotype?

Single Marker ANOVA in SAS

```
proc glm;
class T24;
model logwt = T24;
estimate "additive effect" T24 -0.5 0 0.5;
estimate "dominant effect" T24 -0.5 1 -0.5;
means T24;
```

Dependent Variable: LOGWT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1.37	0.69	12.97	0.0001
Error	246	13.0	0.05		
Total	248	14.37			

R-Square	C.V.	Root MSE	LOGWT Mean
0.095395	31.84404	0.22990	0.72194

Source	DF	Type III SS	Mean Square	F Value	Pr > F
T24	2	1.37	0.69	12.97	0.0001

Effect Estimates

Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
additive effect	-0.10712396	-5.08	0.0001	0.02108124
dominant effect	-0.00742012	-0.25	0.7994	0.02917147

Level of T24	N	Mean	SD
0	61	0.83165121	0.20674180
1	130	0.71710712	0.24033564
2	58	0.61740328	0.22887678

Figure 5. Results of Single-Factor ANOVAs of Marker Loci On Two Linkage Groups

