Chapter 8

# QTL detection in designed experiments and in outbred general pedigree populations

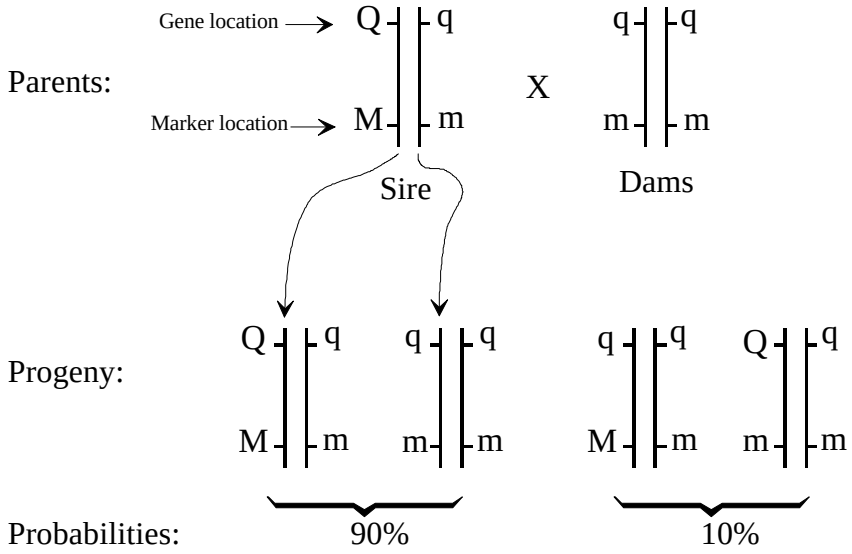
Brian Kinghorn

University of New England

**Detecting QTL in designed experiments**

**Inbred parental lines**

We have already seen in chapter 6 that there is a simple basis to infer QTL segregation in a cross that involves an ideal pattern of marker and QTL genotypes – "a good deal of the cards". Here is the diagram we used:




A backcross between inbred lines. Description is as given in Chapter 6.

### *The Backcross Design*

One way to maximise the probability of getting such a good deal is by making a backcross of inbred lines. Here the sire is a first cross between the lines and the Dams are purebred for one inbred line. The dams are all nicely homozygous and genetically identical to each other. The only things left to chance is that the two inbred lines are fixed for different alleles at both the QTL locus and the Marker locus. The Marker locus is no problem – we can tell pretty quickly from DNA test results whether the lines differ. However, for QTL loci, we can maximise the probability that the lines differ by choosing the lines appropriately – with large genetic distances and large differences for the key traits of interest. [Of course large genetic distances will also increase the chances of differences at market loci too.]

Chapter 6 showed that the difference in merit between progeny receiving M from the sire and those receiving m from the sire is $(1-2r)\alpha$, where $\alpha = a + (p-q)d$, and genotype effects are:

| | |
|---|---|
| $\mu_{QQ}$ | +a |
| $\mu_{qQ}$ | d |
| $\mu_{qq}$ | -a |

However, with inbred lines we have extra information – information that the QTL allele frequencies are 1 and 0 (or 0 and 1) in the inbred lines ***if*** the QTL is segregating.

Thus $\alpha = a + (p-q)d$ is:

$a + (1-0)d = a + d$ if the dam population is qq – going from qq to qQ adds a + d

$a + (0-1)d = a - d$ if the dam population is QQ – going from qQ to QQ adds a - d

Thus:

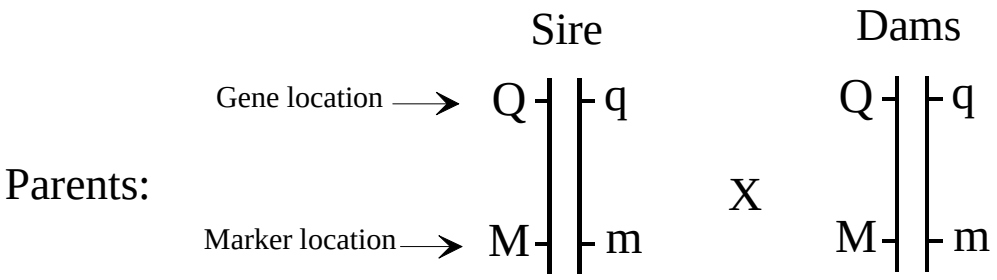
1. The effects of a gene substitution is either (a + d) if the dams hold the less favourable allele, or (a – d) if the dams hold the more favourable allele.

2. The differences between marker genotypes is either (1-2r)(a+d) or (1-2r)(a-d) accordingly.

3. If we make both backcrosses, we can get independent estimates of a and d.

### *The $F_2$ cross Design*

Now both sire and dam lines are heterozygous, given that there is segregation at both loci:




Given recombination fraction is r, we can work out gamete frequencies and progeny genotypes at both loci. The next table shows the genetic value (a, d or –a) and marker genotype (MM, Mm or mm) of the 16 possible 2-locus progeny genotypes:

Table 1.

| | Eggs → | QM | qm | Qm | qM |
|---|---|---|---|---|---|
| Sperm ↓ | Frequency | ½(1-r) | ½(1-r) | ½r | ½r |
| QM | ½(1-r) | a MM | d Mm | a Mm | d MM |
| qm | ½(1-r) | d mM | -a mm | d mm | -a mM |
| Qm | ½r | a mM | d mm | a mm | d mM |
| qM | ½r | d MM | -a Mm | d Mm | -a MM |

Now we have three progeny groups, organised by marker genotype. By looking at the table above, we can derive the predicted frequency and merit for these:

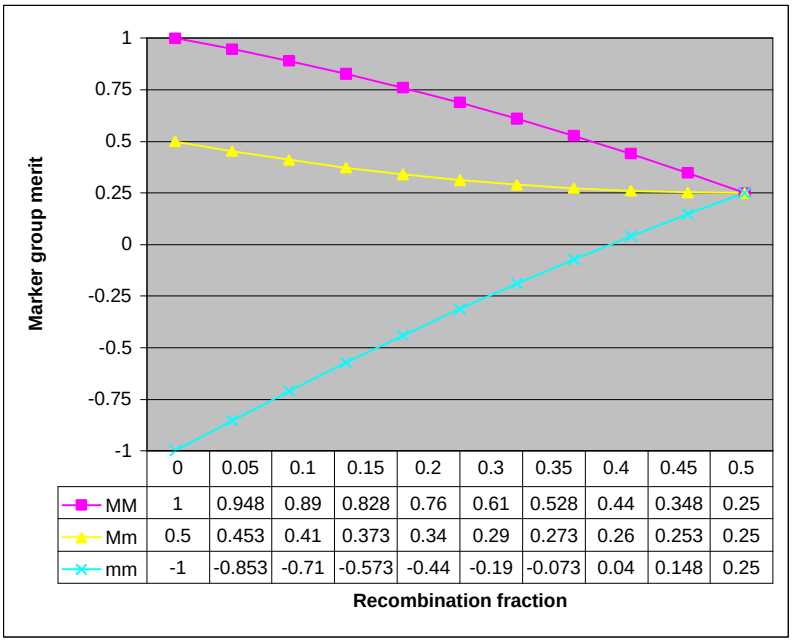
| Marker genotype | Predicted frequency | Equals |
|---|---|---|
| MM | $(½(1-r))^2 + 2½r½(1-r) + (½r)^2$ | ¼ |
| Mm | $2[(½(1-r))^2 + 2½r½(1-r) + (½r)^2]$ | ½ |
| Mm | $(½(1-r))^2 + 2½r½(1-r) + (½r)^2$ | ¼ |

| Marker genotype | Predicted merit | Equals |
|---|---|---|
| MM | $\frac{(½(1-r))^2a + 2½r½(1-r)d + (½r)^2(-a)}{¼}$ | $(1-r)^2a + 2r(1-r)d + r^2(-a)$ |
| Mm | $\frac{2[(½(1-r))^2d + 2½r½(1-r)(a-a) + (½r)^2d]}{½}$ | $[(1-r)^2 + r^2]d$ |
| mm | $\frac{(½(1-r))^2(-a) + 2½r½(1-r)d + (½r)^2a}{¼}$ | $(1-r)^2(-a) + 2r(1-r)d + r^2a$ |

This gives us some sensible predicted merits:

| Marker genotype | r = 0 | r = ½ |
|---|---|---|
| MM | a | ¼a + ½d + ¼a |
| Mm | d | ¼a + ½d + ¼a |
| mm | -a | ¼a + ½d + ¼a |

This is shown graphically below, with a = 1 and d = ½ at the QTL. With no recombination, the marker groups reflect the true QTL genotypic merits. With full recombination (r = ½) all marker groups are predicted to equal the population mean, which is (p-q)a + 2pqd = ½d – as p = q = ½.




| | 0 | 0.05 | 0.1 | 0.15 | 0.2 | 0.3 | 0.35 | 0.4 | 0.45 | 0.5 |
|---|---|---|---|---|---|---|---|---|---|---|
| MM | 1 | 0.948 | 0.89 | 0.828 | 0.76 | 0.61 | 0.528 | 0.44 | 0.348 | 0.25 |
| Mm | 0.5 | 0.453 | 0.41 | 0.373 | 0.34 | 0.29 | 0.273 | 0.26 | 0.253 | 0.25 |
| mm | -1 | -0.853 | -0.71 | -0.573 | -0.44 | -0.19 | -0.073 | 0.04 | 0.148 | 0.25 |

**Non-inbred parental lines**

### *The Backcross Design*

If our parental lines are not inbred, there can be segregation at both QTL and marker loci in the parental lines. For the backcross design, the outcome is just as we found in Chapter 6 – with a need to treat each family separately, if using simple analysis.

### *The $F_2$ cross Design*

The big problem here is that progeny that are heterozygous for the marker locus are not informative (unless we have linked markers, more extensive pedigree information, and proper method, as will be described later in the course).

For an Mm progeny, we cannot tell if M came from the sire or the dam. However, for MM progeny, we can tell that allele M was inherited from each (and similarly for mm progeny), and if the parents are heterozygous then we have useful information. We are then left to contrast MM progeny and mm progeny.

Consider a sire of genotype QqMm (as shown in the diagram above). The distribution of progeny genotypes depends on the frequencies and phases of QTL and marker alleles in the population of dams. For example, assuming linkage equilibrium in the dam population, we can look at the distribution of progeny of marker genotype MM and mm. This is similar to Table 1, but with Mm and mM progeny excluded. the frequencies of Q and q are p and (1-p):

Table 2

| | Eggs → | QM | qm | Qm | qM |
|---|---|---|---|---|---|
| Sperm ↓ | Frequency within marker group→ | p | (1-p) | p | (1-p) |
| QM | ½(1-r) | a MM | | | d MM |
| qm | ½(1-r) | | -a mm | d mm | |
| Qm | ½r | | d mm | a mm | |
| qM | ½r | d MM | | | -a MM |

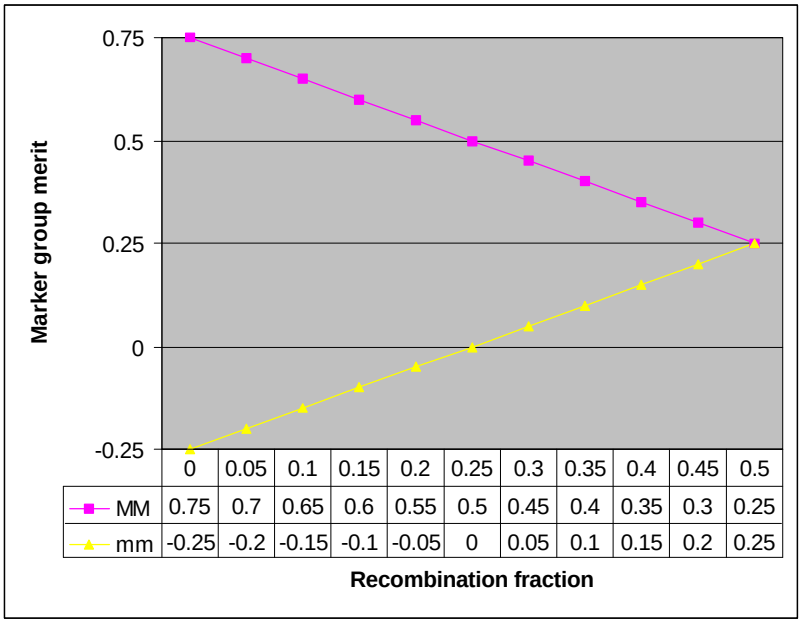
Under this assumption, the expectations of the marker group means are now:

| Marker group | Expectation | Equals |
|---|---|---|
| MM | $\frac{½(1-r)pa + ½r(1-p)(-a) + ½rpd + ½(1-r)(1-p)d}{½(1-r)p + ½r(1-p) + ½rp + ½(1-r)(1-p)}$ | $\frac{½(p-r).a + (r.p+½(1-p-r)).d}{½}$ |
| mm | $\frac{½rpa + ½(1-r)(1-p)(-a) + ½r(1-p)d + ½(1-r)pd}{½rp + ½(1-r)(1-p) + ½r(1-p) + ½(1-r)p}$ | $\frac{½(p+r-1).a + [½(r+p)-rp]d}{½}$ |

This is shown graphically below, with a = 1 and d = ½ at the QTL. With no recombination, the two marker groups no longer reflect the true QTL genotypic merits (as they did for inbred parental lines). This is because, even with no recombination, we do not know which maker allele is associated with which QTL allele in each dam. However, we can find this information for the sire, given sufficient progeny – not that it matters if the sire is heterozygous at both loci.

With full recombination (r = ½) both marker groups are predicted to equal the population mean, which is (p-q)a + 2pqd = ½d – as p = q = ½.

Of course, results will differ when there is some linkage disequilibrium in the dam population.




| | 0 | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.35 | 0.4 | 0.45 | 0.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MM | 0.75 | 0.7 | 0.65 | 0.6 | 0.55 | 0.5 | 0.45 | 0.4 | 0.35 | 0.3 | 0.25 |
| mm | -0.25 | -0.2 | -0.15 | -0.1 | -0.05 | 0 | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 |

With no recombination, and linkage equilibrium in the dams, MM progeny have a probability p of being QQ (merit +a) and (1-p) of being Qq (merit d). This can be seen by inspection of Table 2. At p=½ in the graph above, this comes out at a mean merit of 0.75.

Unfortunately, we have not been able to get independent estimates of a and d. We can do better than this – if we have large full sib families then we have a basis to infer linkage phases in each dam (as we do for the sire in this example). This can lead us to independent estimates of a and d.

**Outbred populations**

We can also do better if we have more than one marker locus, a richer pedigree, and good analysis methods.

With more loci we can often get information about which allele is inherited from which parent – even when the parents and progeny are all heterozygous for the same alleles. We should cover that later.

With richer pedigree and good analysis methods, we can infer the probabilities of being QQ Qq and qq for each animal in the pedigree. We will also cover that later, in chapter 18.