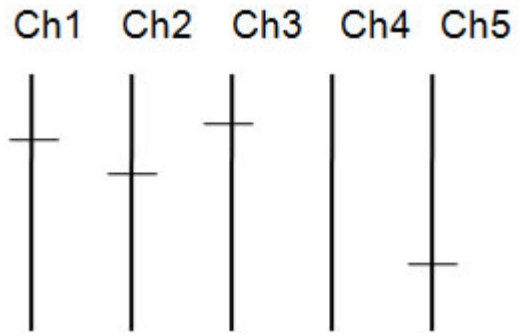
Quantitative traits

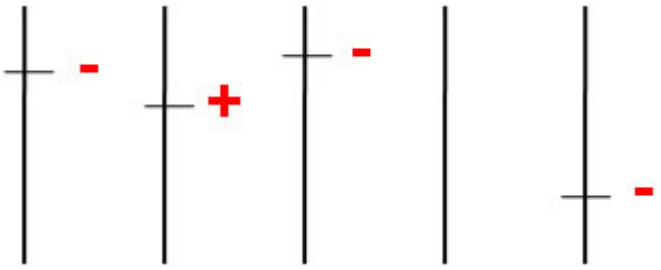
Most traits that plant breeders wish to improve are quantitative, rather than qualitative, in nature. Quantitative traits vary continuously (e.g., yield, quality, stress tolerance etc.), whereas qualitative ones are typically (not always) binary (yes vs no - e.g., resistance to a fungus, colour of flower etc.). Quantitative traits are usually governed by a number of genes, while qualitative ones are often simply inherited (one or two genes). The loci involved in the inheritance of quantitative traits are commonly called QTL (quantitative trait loci).



There may be many regions in a genome affecting a particular trait. For some traits, as many as 50 QTL may be identified, but usually most of these are of too small effect to be useful in MAB; typically the number of useful QTL is less than 10.

Trait effects

A variety may have some QTL that increase a trait (for example, increase yield) and others that decrease the trait. These work together to create the phenotype of the plant.



In this example genome with 5 chromosomes, there are 3 loci that are associated with decreased yield, and one associated with higher yield. The phenotype, depending on the size of the effect of each QTL and how they work together, may be low yield.

Use of wild relatives

Crop wild relatives generally have a poor phenotype from a human perspective (e.g. small fruit, low yield), but this does not mean that they carry negative alleles at all the important QTL. The presence of a "good" (i.e. positive from a breeding perspective) QTL allele can only be detected by backcrossing the wild relative with the crop, which allows the QTL to be detected (by separating the effects) in a crop genetic background. So even though the phenotype of a wild species may not look very promising, there may be hidden allelic variation for the traits you are interested in.

QTL increasing fruit size have been identified from small-fruited wild relatives of tomato (top left) and introgressed into cultivated varieties (top right), creating new lines with larger fruit (bottom row).

Finding “good” QTL

So the key is identifying the “good” QTL – those that affect the trait in the direction you want, and then separating those from the negative ones. This is where QTL identification techniques are important.

Note that these techniques are simply statistical correlations, just like genetic mapping and any marker-trait correlations; however, because we are looking for many markers that correlate with a single trait, it is somewhat more complex statistically.

Prerequisites for QTL mapping

•Availability of a good linkage map (this can be done at the same time the QTL mapping)
•A segregating population derived from parents that differ for the trait(s) of interest, and which allow for replication of each segregant, so that phenotype can be measured with precision (such as RILs or DHs)
•A good assay for the trait(s) of interest
•Software available for analyses

Population types for QTL mapping (1)

There are many possible types of populations that can be used for QTL mapping. Here we compare a few of the most widely used:

Backcross (BC):

Advantages: it is easier to identify QTL as there are less epistatic and linkage drag effects; especially useful for crosses with wild species

Disadvantages: difficult or impossible in species that are highly heterozygous and outcrossing; best when inbred lines are available

Example: Huang XQ, Cöster H, Ganal MW, Röder MS (2003) Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (Triticum aestivum L.). Theoretical Applied Genetics 106: 1379-1389.

Population types for QTL mapping (2)

Recombinant inbred lines (RILs) and Double Haploids (DHs):

Advantages: fixed lines so can be replicated across many locations and/or years; can eliminate problem of background heterozygosity

Disadvantages: Can take a long time to produce or, in the case of DHs, be impossible to produce (not all species are amenable).

Example: He P, Li JZ, Zheng XW, Shen LS, Lu CF, Chen Y, Zhu LH (2001) Comparison of molecular linkage maps and agronomic trait loci between DH and RIL populations derived from the same rice cross. Crop Science 41: 1240-1246.

See also: Tanksley and Nelson 1996

Population types for QTL mapping (3)

F2:3 (and related) populations:

Advantage: Fast and easy to construct

Disadvantage: F3 families are still very heterozygous, so the precision of the estimates can be low (because of the high standard error); can't be replicated

Example: Jampatong C, McMullen MD, Barry BD, Darrah LL, Byrne PF, Kross H (2002) Quantitative trait loci for first- and second-generation European corn borer resistance derived from the maize inbred Mo47. Crop Science 42: 584-593.

Population types for QTL mapping (4)

Near Isogenic Lines (NILs):

Advantage: Very precise and statistically strong, as background is constant; especially useful for validation experiments

Disadvantage: Can take time to construct; only useful for specific target QTL

Example: Szalma SJ, Hostert BM, LeDeaux JR, Stuber CW, Holland JB (2007) QTL mapping with near-isogenic lines in maize. Theoretical and Applied Genetics 114: 1211-1228.

See also later slide on using NILs for validation.

These are just a few of the most used types – there are many other possible populations types. Also note that outcrossing species require special considerations. See the references in the Resources at the end of this section for more details.

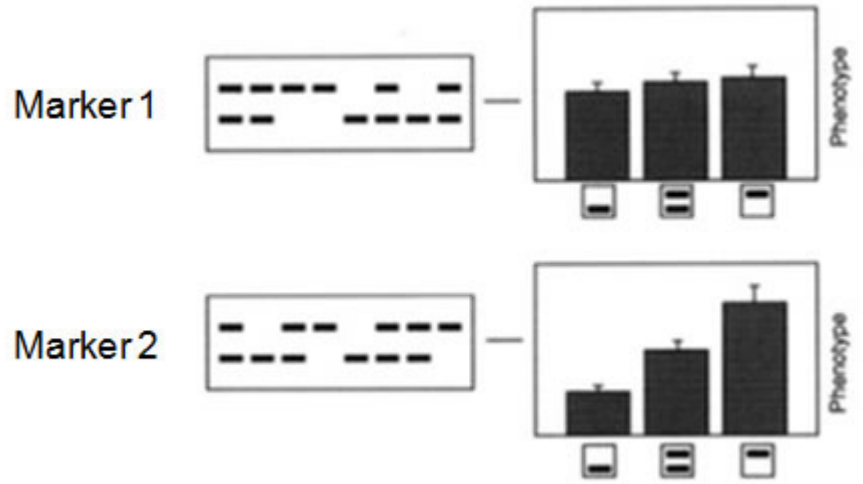
Principles of QTL mapping

The purpose of the phenotyping experiment is to assign a trait value to each mapping population member. This value is then combined with the allele score at the set of marker loci distributed throughout the genome. A datafile is then created which includes all the trait data and all the marker data for the entire population.

Various software applications can be applied to this data file to identify statistical associations between the presence of alternative alleles and the trait value.

Principles of QTL mapping, contd.

To calculate the strength of the association between genotype and phenotype, the mapping population is split into two groups, according to the allele they carry for that trait at each marker in turn. Then the mean trait value of these two classes is compared. If the difference is significant, then this provides initial evidence for the location of a QTL in the neighbourhood of the marker (Young 1996).



In this example from Young (1996), there is no significant difference in the mean trait value whether the progeny are homozygous for either of the Marker 1 parental alleles or are heterozygous. But at Marker 2, it is clear that the mean trait value of the three allelic groups is marker allele-dependent, indicating a probable QTL for this trait near Marker 2.

Methods to detect QTLs: single point regressions

There are a number of statistical methods used for QTL identification. The simplest, and probably most common, is the single point (ie. single marker) linear regression.

As in the previous example from Young, this approach looks at every marker-trait combination, one marker at a time. Using a linear regression statistical test (as opposed to a simple t-test, for example) gives the additional information of the phenotypic variance (R2) of that QTL – that is, how much of the phenotypic variation is associated with that specific QTL. An example is shown on the next slide.

R^2 values

Recall that by definition there are many QTL contributing to the phenotype of a plant. Note that in this example, the R^2 values, or the percent of phenotypic variation, range from 10% to 2%.

Marker SSR218 is statistically significant ($p = 0.05$) but has an R^2 of approximately 2%. Is it worth your time to use marker-assisted breeding if it will only increase your trait phenotype by 2%? This must be your decision.

Pros and cons of single point QTL mapping

Pros:

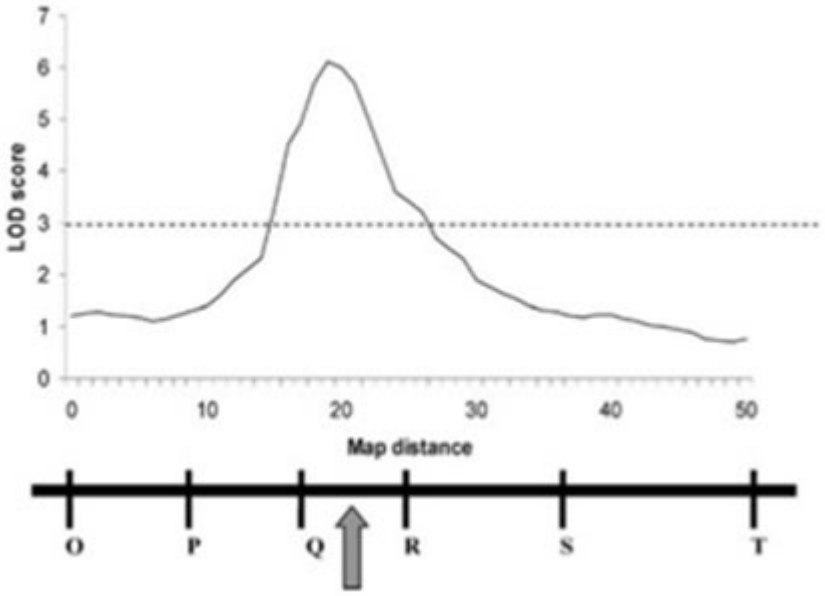
•Is the most simple method, can be performed by most statistical software programs
•Does not require a complete genetic linkage map

Cons:

•The further away from a marker the QTL is, the less likely it is to be detected (due to the chance of recombination between the marker and QTL); thus, QTL effects may be underestimated (this can be somewhat mitigated by having testing more markers)
•Is less precise about the location of the QTL

Methods of QTL detection: interval mapping

Interval mapping (IM) (Lander and Botstein 1989) analyses pairs of linked markers rather than one marker at a time. It is statistically more powerful. The output is more typically in a graphical form, as in the figure below.



A hypothetical example of interval mapping output from Collard et al. (2005). IM gives a slightly more specific location of the QTL.

Composite Interval Mapping

Composite interval mapping (CIM) (Zeng 1994) combines interval mapping and linear regression, and looks at more markers at a time. This can give greater power in identifying key QTL, but is more statistically complicated and requires more computational power.

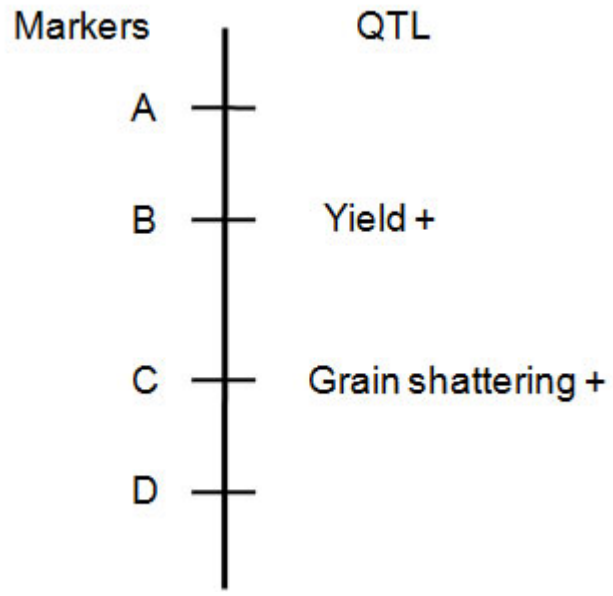
It is recommended that single marker QTL identification is completed first, than CIM may be used for further information.

Which method to use?

Either of these methods are perfectly legitimate to use for Marker-Assisted Breeding. The results should be quite similar. For MAB, in the end what is needed is a marker to use to track the QTL or trait; it doesn't really matter which method or software program you use to arrive at that point. However, for using a marker identified by someone else in your own work, you would of course want to know the methods and statistical levels they used so you can feel comfortable using the marker(s) in your work.

Breaking linkage drag

Linkage drag refers to the (usually undesirable) effects of genes linked to the genes or QTL we are trying to introgress. If a desirable QTL for trait X lies close to an undesirable gene affecting trait Y, you will want to "break" the linkage drag – that is, separate the good QTL from the bad.



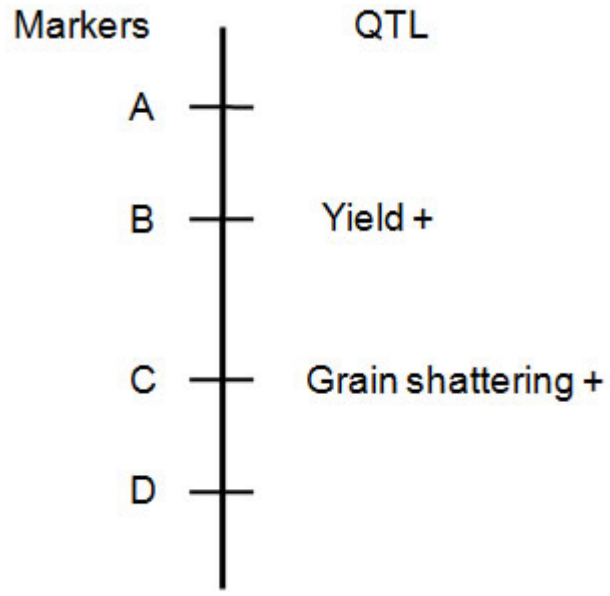
Overcoming linkage drag requires passing the material through meiosis again, and searching for recombinants (possibly rare) between the target QTL (which is tracked by the DNA markers used originally to identify it) and the undesirable gene/QTL, such that you can select a plant that has the desirable QTL but not the undesirable one.

Here a QTL that increases yield is near a QTL that increases grain shattering. In the next population, you would look for a plant that still had the positive allele at Marker B, but has lost the grain shattering allele at Marker C.

Breaking linkage drag, contd.

A major advantage of using advanced backcross populations for QTL mapping is that each time a cross is made back to the recurrent parent (usually the cultivated variety, if the other parent is a wild relative), less of the donor genome remains.

Once a locus in the segregating population becomes “fixed” (homozygous) for the recurrent parent allele (AA), you no longer need to check it with markers, because it can only stay homozygous in each subsequent backcross generation.



Therefore if your recurrent parent is AA, and in your segregating population Marker A and Marker D are AA, after a cross back to the recurrent parent, you only need to check the progeny with Markers B and C, and hopefully find a plant that is Aa only at B (linked to the yield QTL is) and now AA at Marker C (has lost the bad QTL due to recombination).

QTL validation

Before being used for MAB, a QTL needs to be validated (confirmed) to rule out the possibility of statistical anomalies or errors. The validation process tests whether the same QTL appears when the material is grown in other locations and/or years, and whether its effect can still be detected when introduced into a series of different genetic backgrounds.

A clear and thorough example of this can be seen in Landi et al. 2005.

Near-Isogenic Lines (NILs)

A good way of validating a putative QTL is by creating a Near-Isogenic Line (NIL). This is a new line that differs from its parent in only one genomic location: where the QTL is.

Using the marker identified for that QTL, backcrosses are made to the recurrent parent (usually the cultivated variety, in a wide cross) until the entire genome of the line is exactly like the recurrent parent except in the region around the marker locus. Any phenotypic differences between these 2 lines is then most probably due to the QTL linked to the marker locus, thereby validating the QTL.

Pyramiding of QTL

After identifying good QTL for use in a breeding programme, often it may be useful to combine 2 or more QTLs into the same line. This is called pyramiding (or “stacking”).

Plant with good yield + plant with good colour + plant with good flavour

= Plant with good yield + good colour + good flavour!

See an example in the next section.

Fine mapping and cloning of QTL

Markers allow a QTL to be manipulated in a breeding programme, but they seldom give any clues as to function or molecular mechanism. To obtain this information, it is necessary to clone the QTL, which requires it to be first mapped to a much higher resolution position. This is achieved by a combination of larger population size (up to a few thousand) and additional markers known - or discovered - to map in the region of the QTL.

See Tuberosa et al. 2007 for a good discussion of this, and Frary et al. 2000 for an example.