QTL mapping in autotetraploid populations

C. A. Hackett[1] and J. E. Bradshaw[2]

[1]Biomathematics and Statistics Scotland, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA. Email christine@bioss.ac.uk

[2]Genome Dynamics Programme, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA

Summary

Statistical methods are well developed for locating quantitative traits (QTLs) on molecular marker maps in diploid species. However mapping in autotetraploid species, such as potato, has received less attention because of the complexities of tetrasomic inheritance. Here we propose a maximum likelihood approach for QTL mapping in an autotetraploid population, and use this to study the inheritance of blight resistance and maturity in potato.

For data consisting of molecular marker phenotypes and trait values (eg level of disease) for two parents and their offspring, the steps of the analysis are :

- partition molecular markers into independently inherited sets, corresponding to the sets of chromosomes
- order the markers within each chromosome to obtain a linkage map of each parent
- reconstruct the inheritance of chromosome segments from parent to offspring
- examine positions along each chromosome for evidence of a relationship between the presence of chromosome segments and the value of the trait.

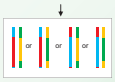
1. Inheritance in autotetraploid species by random chromosomal segregation

Chromosomes occur in sets of 4 – potato has 12 sets of 4

The four pair at random

Recombination occurs between the two chromosomes of each pair

One (recombined) chromosome from each pair is passed to the gamete



A potato gamete, containing 12 pairs of recombined chromosomes, unites with one from the other parent to give a tetraploid offspring.

2. Molecular marker information

Most molecular marker data is recorded as the presence/absence of the marker in the two parents and in the offspring of a cross between them. Autotetraploid individuals can have up to four copies of a marker. The number of copies is not directly observable but can be inferred from the segregation ratio in that individual's offspring. Markers that are absent in one parent and present as a single copy in the other are highly informative for linkage analysis.

Number of copies Parent 1	Parent 2	Offspring ratio Presence:Absence	Name
1	0	1:1	Simplex x Nulliplex
2	0	5:1	Duplex x Nulliplex
1	1	3:1	Simplex x Simplex

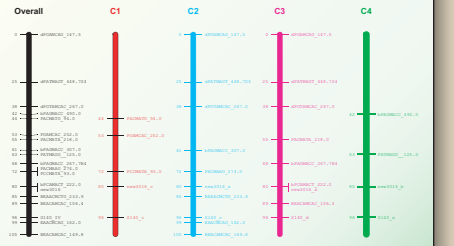
Codominant markers such as microsatellites, where the presence of more than one alleles at a single locus can be detected, are also highly informative.

3. Linkage analysis

Markers on different sets of chromosomes are passed from parent to offspring independently, but markers on the same set are passed on together unless a recombination occurs between them. The recombination frequency i.e. the probability of a recombination between the two markers forms the basis for ordering markers within a chromosome set. The recombination frequency can be estimated from the joint presences/absences of the two markers, but the form of the estimator depends on the number of copies of the marker, and whether the markers are on the same chromosome of the set (coupling phase), or different chromosomes (repulsion phase).

Cluster analysis can be used to separate markers into chromosome sets. The markers within each set can then be ordered, based on the map distances between all pairs of markers. The map distance is a transformation of the recombination frequency to achieve additivity, and is measured in centiMorgans (cM). The ordering problem is similar to a travelling salesman problem, and we use simulated annealing to find the best order.

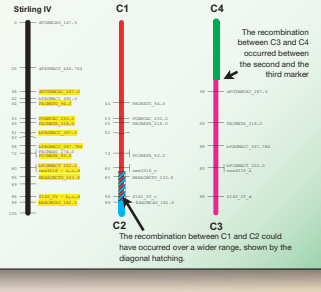
4. Map of potato linkage group IV



Map of potato linkage group IV for the cultivar Stirling, showing the overall map, and the allocation of the markers to the four chromosomes of the group (C1-C4).

5. Reconstruction of offspring chromosomes

The configuration of markers in each offspring is examined to see how the chromosomes could have been derived from the parents. A branch and bound search is used to identify configurations with the smallest number of recombinations leading to the offspring. There may be more than one such configuration.

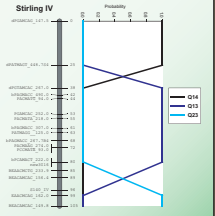


The figure shows the reconstruction of the chromosomes inherited from Stirling linkage group IV for one offspring. This offspring carried the markers shown in yellow in the overall chromosome. Some marker names are omitted as the offspring was not scored for these. The branch and bound algorithm shows that the chromosomes must have paired as C1 with C2, and C3 with C4, and that two recombinations occurred, one for each pair.

6. QTL configuration probability

The reconstruction (left) shows that this offspring carries material from Stirling chromosomes C1 and C4 on the top section, chromosomes C1 and C3 on the middle section, and chromosomes C2 and C3 on the lower section. In the regions of the inferred crossovers more than one configuration is possible.

This figure (right) shows the possible configurations, designated Q14, Q13 and Q23, and their probabilities at different positions along the chromosome.



7. Modelling the quantitative trait data

If we knew the location of a gene affecting a quantitative trait, and the parental origin of the chromosomes at that location, q_i, for each offspring i, we could model the trait values y_i as

$$y_i = F(q_i)$$

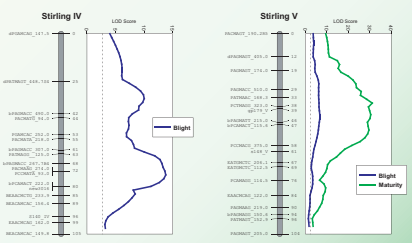
for some function F. We only know the marker information o_i for each offspring. However we can write the likelihood of the data as

$$f(y_i, o_i) = \sum_{g} \sum_{q \in Q_i} f(y_i \mid q_i)\, P(q_i \mid g_i)\, P(g_i \mid o_i) P(o_i)$$

where g_i is the reconstructed chromosome for offspring i. This is a mixture model for the trait data. It is fitted at a series of positions along the chromosome, using the EM algorithm at each position. This separates the mixture model into a weighted regression of the trait values on the inferred QTL genotypes at that position, followed by an updating of the QTL genotype probabilities. This gives a likelihood profile along the chromosome. This is compared with the likelihood of no QTL at that point to give the log of the likelihood ratio, or LOD score.

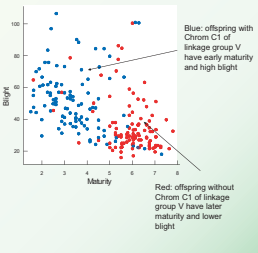
8. Genetic control of foliage blight in potato

These figures show the likelihood profiles for foliage blight on Stirling linkage groups IV and V, with peaks at about 72cM and 44cM respectively. The dotted line shows the 95% point from a permutation test. Linkage group V also shows the profile for maturity, with its peak close to that for blight, but there is no evidence for a QTL affecting maturity on group IV. If the residuals from a regression of blight on maturity are mapped, then no QTL is found on linkage group V, showing that the QTL here is a maturity effect. The fitted QTL model shows that for linkage group V chromosome C1 was associated with significantly earlier maturity and increased blight than the other three, while for linkage group IV offspring inheriting chromosomes C1 and C4 had significantly more blight than other offspring.



9. Relationship between maturity, blight and genotype on linkage group V

There is a significant negative correlation between the maturity and the blight. The factor for maturity on group V explains more than 50% of the variation in maturity.



Conclusions and further work

It is now possible to assemble maps of molecular markers and locate QTLs affecting quantitative traits in autotetraploid species, as well as diploids.

Here we have assessed significance of the QTLs by means of a permutation test to control the chromosome type I error rate, but further work is needed to determine the significance thresholds to control the overall false discovery rate.

Software to carry out these analyses, TetraploidMap, is available from the authors.