QTL Mapping Notes

Nathan Hughes (nah31@aber.ac.uk)

July 5, 2017

Contents

1 What is QTL Analysis?

Quantitative Trait Locus Analysis is a statistical method that links two types of information

- Phenotypic data (i.e. trait measurements)
- Genotypic data (usually molecular markers)

1.1 What is a single QTL

An individual QT is a secton of DNA, it is arbitrarily described by correlation and not a physical element that can be seen.

1.2 What is it used for?

QTL attempts to explain the genetic basis of variation in complex traits. They allow researches in many fields, from agriculture to evolutionary biology, to link complex phenotypes to specific regions of chromosomes.

1.3 What is the goal of it?

The end goal of the QTL process is to identify the action, interaction number, and precise location of regions on a chromosome which "code" for specific phenotypic expression

More specifically the principle goal is to answer the question of whether phenotypic differences are primarily due to few loci with farily large effects, or to many loci, each with minute effects. It appears that a substantial proportion of the phenotypic variation in many quantitative traits can be explained with few loci of large effect, with the remainder due to numerous loci of small effect. For example, in domesticated rice (Oryza sativa), studies of flowering time have identified six QTL; the sum of the effects of the top five QTL explains 84% of the variation in this trait.

2 What can you do with QTL Data

Once QTL have been identified, molecular techniques can be employed to narrow the QTL down to candidate genes. One important emerging trend in these analyses is the prominent role of regulatory genes, or genes which code for transcription factors and other signalling proteins. For instance, in rice, three flowering time QTL have been identified at the molecular level, and all of these loci encode regulatory proteins know from studies of *Arabidopsis thaliana*.

3 How does QTL work?

In order for QTL to be conducted, two things are required:

1. Two or more strains of organisms which differ genetically with regard to the trait of interest (Fig1. a shows variation in a population over time).
    - For example selecting lines which have alleles that influence flowering colour.
2. Genetic markers for that specific species which distinguish between parental lines.
    - Molecular markers are preferred for genotyping, because these markers are unlikely to affect the trait of interest.

4 Generating populations for study

This involves using the identified parents/individuals/organisms and running a breeding program to build up a population suitable for use in genetic difference identification.

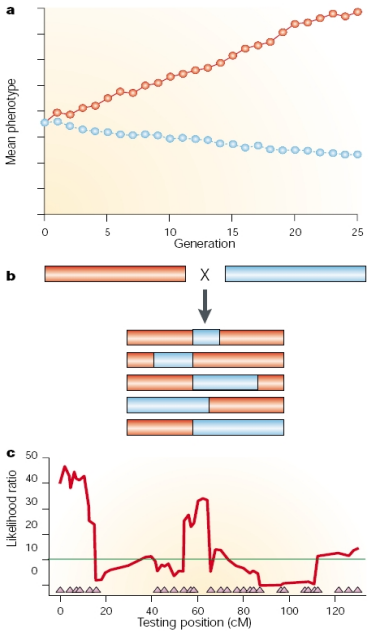

Figure 1:

4.1 Initial selection and crossing

Carrying out the QTL analysis is done by crossing the, genetically different, identified strains and a f^1 population of **heterozygous** individuals is produced (Fig 1. B). These are then selectively chosen and used to produce further generations to a f^n generation. Next the genotypes' and phenotypes' of the f^n generation individuals are scored. Markers that are genetically linked to a QTL influencing the trait of interest will segregate more frequently with the trait values (flowering colour, in the previously mentioned example). Whereas unlinked markers will not show significant association with phenotype (Fig 1. C).

4.2 Multi-gene traits

For traits controlled by tens or hundreds of genes, the paternal lines need not actually be different for the phenotype in question; rather they must simply contain different alleles, which are then resorted by recombination in the derived population to produce a range of phenotypic values. Consider for example, a trait which is controlled by four genes, wherein the upper-case alleles increase the value of the trait being studied. The lower-case alleles decrease the vale of that trait. Here, if the effects of the alleles of the four genes are similar, individuals with the AABBccdd and aabbCCDD genotypes might have roughly the same phenotype measurement. The members of the f^1 population (AaBbCcDd) would be invariant and would have an intermediate phenotype. However the f^2 generation, or the progeny from a backcross would have anywhere from zero to eight upper-case alleles. The backcross progeny would have anywhere from four to eight upper-case alleles.

4.3 Problems with population sizes

A small population size may lead to problems and as such is a critical factor to the validity of experiments. A small sample size may lead to overestimation of the effect of a QTL.

5 Types of markers

As mentioned there are several different kinds of markers, these are just a few commonly used examples

5.1 SNPs

Single nucleotide polymorphisms, are the most common type of genetic variation among people. Each SNP represents a difference in a single DNA building block (nucleotide). For example a SNP may replace the nucleotide cytosine (C) with the nucleotide thymine (T) in a certain region of DNA.

- The wikipedia article on this is quite good:

- https://www.wikiwand.com/en/SNP_genotyping

5.2 SSR

Simple sequence repeats, these have the ability to be done at a medium throughput.

5.3 RFLPs

Restriction fragment length polymorphisms. These were used as the primary markers up to the late 80's but couldn't be done through automated means, was expensive and thus became obsolete.

6 Performing QTL Analysis with a set of markers

7 Biology Dictonary