

Question & Answer

Q&A: Genetic analysis of quantitative traits

Trudy FC Mackay

What are quantitative traits?

Quantitative, or complex, traits are traits for which phenotypic variation is continuously distributed in natural populations, with population variation often approximating a statistical normal distribution on an appropriate scale. Quantitative traits include aspects of morphology (height, weight); physiology (blood pressure); behavior (aggression); as well as molecular phenotypes (gene expression levels, high- and low-density cholesterol levels).

What causes the continuous distribution of phenotypes for quantitative traits?

The continuous variation for complex traits is due to genetic complexity and environmental sensitivity. Genetic complexity arises from segregating alleles at multiple loci. The effect of each of these alleles on the trait phenotype is often relatively small, and their expression is sensitive to the environment. Allelic effects can also depend on genetic background and sex. Because of this complexity, many genotypes can give rise to the same phenotype, and the same genotype can have different phenotypic effects in different environments. Thus, there is no clear relationship between genotype and phenotype.

Does this mean you can't see Mendelian ratios for quantitative traits?

Yes, because of the small magnitude of the allelic effects on the phenotype.

Mendelian variants have large effects on the phenotype so there is a clear correspondence between genotype at a locus and trait phenotype. For any trait there is a continuum of allelic effects from small to large: the large effects segregate as Mendelian variants, while the small effects segregate as quantitative genetic variation. For example, human height is a classic quantitative trait, but achondroplasia (dwarfism) is caused by a Mendelian autosomal dominant mutation in the *fibroblast growth factor receptor 3* gene.

Why are quantitative traits important?

Quantitative genetic variation is the substrate for phenotypic evolution in natural populations and for selective breeding of domestic crop and animal species. Quantitative genetic variation also underlies susceptibility to common complex diseases and behavioral disorders in humans, as well as responses to pharmacological therapies. Knowledge of the genetic basis of variation for quantitative traits is thus critical for addressing unresolved evolutionary questions about the maintenance of genetic variation for quantitative traits within populations and the mechanisms of divergence of quantitative traits between populations and species; for

increasing the rate of selective improvement of agriculturally important species; and for developing novel and more personalized therapeutic interventions to improve human health.

How can you identify genes affecting quantitative traits?

This is usually done in stages. In the first stage, we map quantitative trait loci (QTLs) affecting the trait. QTLs are genomic regions in which one or more alleles affecting the trait segregate. In the second stage, we focus in on each QTL region to further narrow the genomic intervals containing the gene or genes affecting variation in the trait. The final and third stage is most challenging: pinpointing the causal genes.

How do you map QTLs?

There are two basic approaches: linkage mapping and association mapping. Both approaches are based on the principle that QTLs can be tracked via their genetic linkage to visible marker loci with genotypes that we can readily classify. The most common markers used today are molecular markers, such as single nucleotide polymorphisms (SNPs), polymorphic insertions or deletions (indels), or simple sequence repeats (also known as microsatellites). If a QTL is linked to a marker locus, then on average individuals with different marker locus genotypes will have a different mean value of the quantitative trait (Figure 1). Linkage

Trudy F C Mackay, Department of Genetics,
North Carolina State University, Raleigh
NC 27695-7614, USA.
Email: trudy_mackay@ncsu.edu

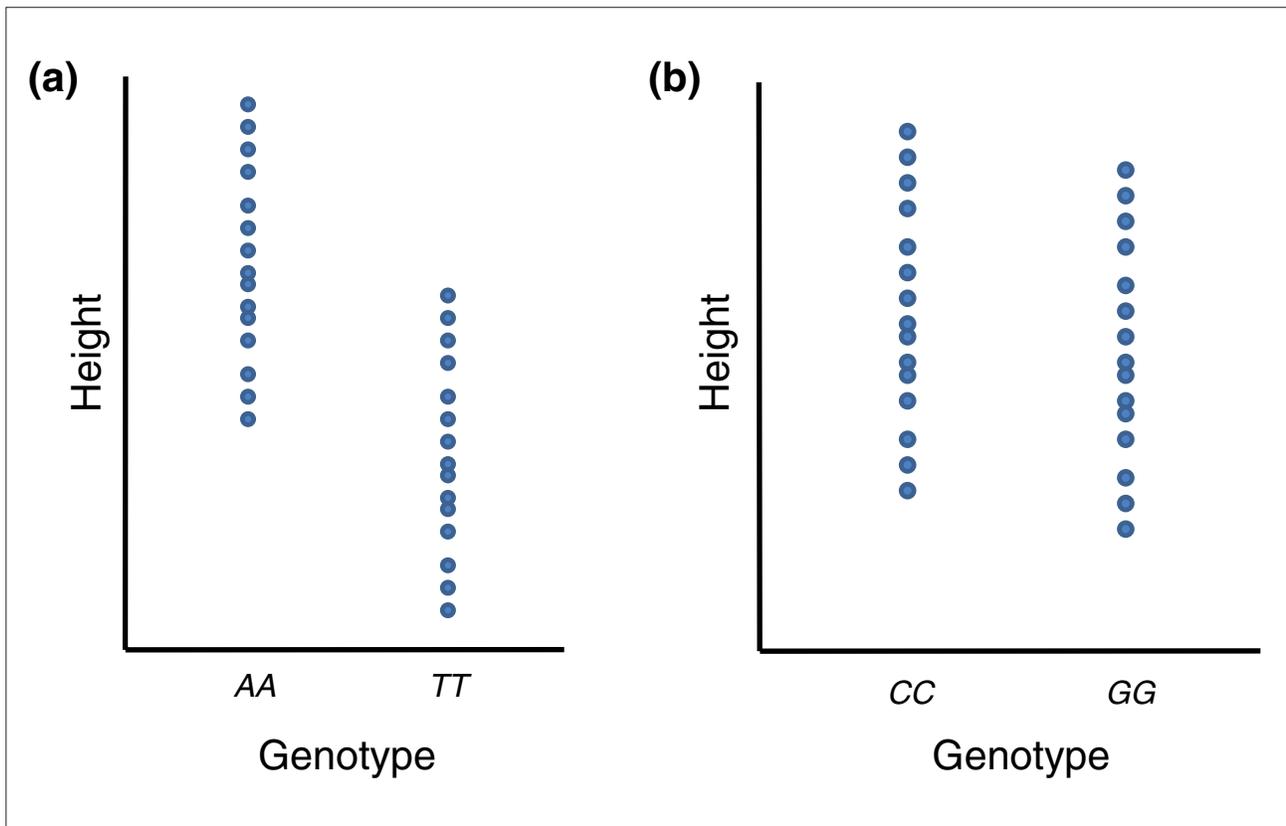
**Figure 1**

Illustration of hypothetical data on height for 15 individuals at each of two marker loci, one with alleles *A* and *T*, the other with alleles *C* and *G*. **(a)** Individuals with the *AA* genotype are taller than those with the *TT* genotype. Therefore, a QTL affecting height is linked to this marker locus. **(b)** There is no significant difference in height between individuals with the *CC* and *GG* genotypes. Therefore, no QTLs affecting height are linked to this marker locus.

mapping involves tracing the linkage of a trait with a marker either through families in outbred populations (such as human populations), or by breeding experiments in which animal or plant strains that vary for the trait are crossed through several generations. By contrast, association mapping looks for associations between a marker and different values of a trait in unrelated individuals sampled directly from a population. In both cases, we need to obtain measurements of the phenotype and determine the marker locus genotypes for all individuals in the mapping population, at all marker loci. Then we use a statistical method to determine whether there are differences in the value of the

quantitative trait between individuals with different marker locus genotypes; if so, the QTL is linked to the marker. We repeat this for every marker (or pair of adjacent markers) to perform a genome scan for QTLs. The results of a genome scan are depicted graphically, as shown in Figure 2.

So mapping QTLs depends crucially on statistical expertise?

It is important to understand the principles of the experimental design to measure the quantitative trait phenotypes in the mapping population, and consultation with a statistician is recommended if you have any questions about these principles. The actual mapping methods do not

require strong statistical expertise. There are many freely available statistical programs for implementing QTL mapping methods and using permutation to determine appropriate significance thresholds. Two popular software suites are QTL Cartographer (<http://statgen.ncsu.edu/qtlcart>) and R-QTL (<http://www.rqtl.org>).

If statistical tests are needed for mapping, you must need a lot of individuals to map quantitative traits?

This is a key question. The answer has two components: the number of individuals needed to detect a QTL and the number required to localize the gene or genes at the QTL. The

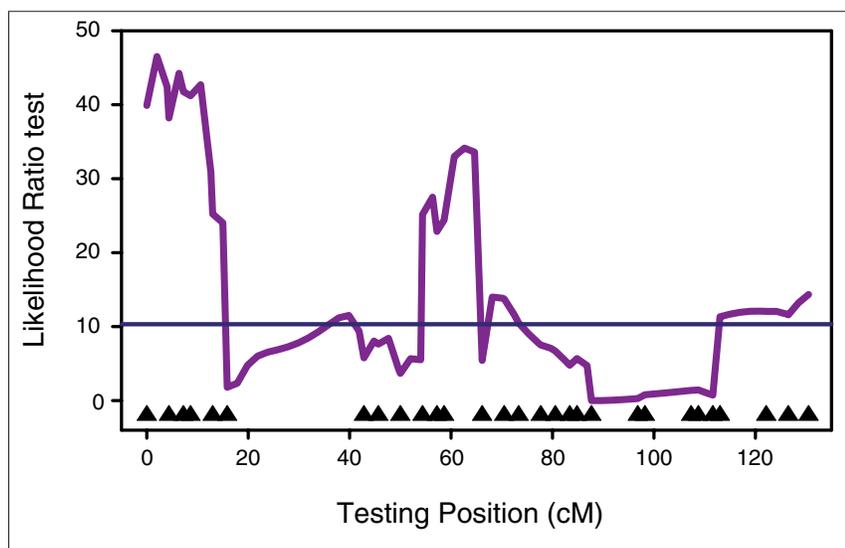


Figure 2
 The results of a genome scan are depicted graphically, where the locations of the markers are given on the x-axis (black triangles), and the result of the statistical test is indicated on the y-axis (here a likelihood ratio test). The significance threshold is given by the horizontal line parallel to the x-axis and intersecting the y-axis at the appropriate value. The significance threshold has been adjusted to account for the number of independent tests performed, and was determined by a permutation test. Evidence for linkage of a QTL with markers occurs when the test for linkage generates a significance level that exceeds the permutation threshold. The best estimate of the QTL location is the position on the x-axis corresponding to the greatest significance level.

answer also depends on whether you are doing a linkage study or an association study. To detect a QTL in a linkage study, you need to identify a reliable difference in the average value of the trait between marker genotypes. How many individuals you need for this depends broadly on the frequency of the QTL alleles in the population you are looking at, and how large their effects are. (More precisely - the power to detect a difference in the mean value of the trait between two marker genotypes depends on δ/σ_w , where δ is the difference in mean between the marker classes, and σ_w is the standard deviation of the trait within each marker genotype class.) In a linkage-mapping study, the different alleles are generally at intermediate frequency, and in this case, the marker genotype and quantitative trait phenotype must be recorded for more than 500-1,000 individuals if the QTL has a moderate effect ($\delta/\sigma_w = 0.25$).

For QTLs with small effects ($\delta/\sigma_w = 0.0625$), much larger sample sizes (more than 10,000 individuals) are needed. Allele frequencies can be more extreme with association mapping designs, and this translates to greater sample sizes required to detect QTLs. For example, more than 30,000 individuals would be necessary to detect a moderate effect QTL ($\delta/\sigma_w = 0.25$) for which the frequency of the rare allele was 0.1.

So what about the numbers required to localize a QTL?

To localize a QTL you need individuals in which recombination has occurred in the vicinity of the QTL so that only markers very close to the QTL on the chromosome remain linked to it. The bottom line is that the more precisely we want to localize a QTL by linkage (in terms of the recombination fraction, c), the larger

the number of individuals necessary. For example, we would only need 29 individuals to detect at least one recombinant in a 10 cM interval ($c = 0.10$), but 2,994 individuals to detect at least one recombinant in a 0.1 cM interval ($c = 0.001$).

Wouldn't you also need a lot of markers, to be sure that some were very close to the QTL?

Yes. The smaller the physical distance on the chromosome, the smaller the number of recombinants will be, and the larger the marker density we need to identify them. The relationship between recombination fraction and physical distance varies between species and across the genome within species. We can infer the scale of mapping using the *Drosophila* genome as an example, where a QTL localized to a 5 cM interval would span 2,100 kb and include on average 245 genes, whereas a QTL localized to a 1 cM interval would span 420 kb and include 49 genes. Clearly, extremely large linkage-mapping populations would be needed if we attempted to simultaneously detect QTLs and localize them to small chromosomal regions. That is why linkage mapping of QTLs is typically an iterative procedure where we first determine the general location (in 10-20 cM intervals) of QTLs in a mapping population of several hundred to approximately a thousand individuals. We then narrow down the regions that we know contain the QTLs, and determine their location more precisely by focusing on individuals in which recombination has occurred between the markers flanking the QTL - and then essentially repeat the whole procedure on the smaller genomic regions. This phase requires breeding many more individuals to obtain the necessary recombination, and identifying molecular markers within the region of interest. These experiments are very laborious and

rarely result in positional cloning of QTLs.

What is the difference between linkage and association mapping from the point of view of numbers of individuals and markers needed?

Association mapping is done on random-mating, and thus much more heterogeneous, populations, so there will be more recombinant individuals, and thus fewer individuals are necessary to localize QTLs. The number of markers required in an association mapping study depends on the scale and pattern of linkage disequilibrium (LD) - that is, the correlation of allele frequencies at two or more polymorphic loci, or the tendency of a particular pair or group of alleles to be found together in different individuals. If a group of markers is in high LD, we only need to genotype one of them as a proxy for all the others in the LD block. Thus, in species with large LD blocks, such as pure breeds of dogs, only a few markers may be required for QTL detection, but it will not be possible to localize QTLs very precisely by within-breed association mapping. In contrast, knowledge of all sequence variants is necessary for association mapping in species like *Drosophila*, where LD can decline very rapidly over short physical distances. Under this scenario, however, QTL localization can be quite precise. In humans, commercial genotyping arrays with many hundreds of thousands of markers spanning the whole genome have been developed, based on tagging SNPs in LD blocks, facilitating a new era of genome-wide association studies in people. The requirement for genotyping large numbers of markers in large numbers of individuals has meant that, until recently, most association-mapping studies have been for a candidate gene or candidate gene region, and used

only a subset of all possible molecular polymorphisms.

Which is better, linkage mapping or association mapping?

Both methods have advantages and disadvantages. Linkage mapping, particularly in controlled crosses (as opposed to, say, human families), has the advantage of increased power to detect QTLs because all segregating alleles are at intermediate frequency, whereas allele frequencies in a population used for association mapping can vary throughout the entire range. On the other hand, association mapping can give increased power to localize QTLs because of the higher recombination between markers and QTL alleles in random-mating populations. Recombination can be increased in linkage-mapping designs by random mating of F_2 or backcross populations for several generations (so-called advanced intercross lines). Linkage mapping also has the disadvantage of reduced genetic diversity, especially when crosses between a pair of lines are used to create the mapping population. Association mapping samples the whole gamut of genetic diversity in the population. The reduced genetic diversity in linkage-mapping populations can be somewhat alleviated by starting from crosses of four or eight initial parental strains. Finally, association mapping relies on LD between marker alleles and QTL alleles, and any mixing of different populations can cause LD that is not due to close linkage, thus leading to incorrect conclusions.

How do you identify the genes corresponding to QTLs?

QTL mapping will identify a genomic region containing one or more candidate genes affecting the trait. Determining which one(s) are causal is the next step. The most

straightforward method is high-resolution recombination mapping. However, this method is limited to QTL alleles with large effects and to organisms amenable to the experimental generation of tens of thousands of recombinants. Otherwise, we need to seek corroborating evidence, such as DNA polymorphisms between alternative alleles of one of the candidate genes that could change the protein, a difference in mRNA expression levels between genotypes, or expression of RNA or protein in tissues thought to be relevant to the trait. Associations of markers in candidate genes with the trait that are replicated in independent studies also constitute strong evidence that the gene affects variation in the trait. In model organisms, it is possible to test whether a mutation in one of the candidate genes affects the trait, or whether the mutant gene fails to complement QTL alleles. Formal proof that a specific allelic substitution affects the trait comes from replacing the allele of a candidate gene in one strain with that of the other, without introducing any other changes in the genetic background, but this is not possible in very many organisms.

What have we learned from QTL mapping?

While literally thousands of studies have been published reporting QTLs for all imaginable traits (including biochemical traits, such as transcript abundance) and in a wide range of organisms, few actual genes corresponding to QTLs have been identified, and these represent alleles with large effects and thus only a very small proportion of QTLs. We now know that most alleles affecting quantitative traits have very small effect, and it is clear that most experimental efforts to map QTLs have not been large enough to detect them. Furthermore, QTLs that have

been detected often break down into multiple linked QTLs with smaller effects when subjected to high-resolution mapping. It is also clear that mapping studies so far are likely to have missed much of the genetic variation responsible for quantitative traits. This follows from the fact that the number of QTLs detected is usually positively correlated with the sample size of the mapping population, so if the smaller studies were enlarged more QTL would presumably emerge. Thus, it appears that large numbers of loci are responsible for quantitative genetic variation. Some surprises have come from QTL mapping: many genes corresponding to QTLs are previously unknown genes predicted computationally from genome sequences, genes affecting development associated with adult quantitative traits, or even genes occurring in otherwise 'gene deserts'. QTLs often have allelic effects that vary depending on background genotype, environment and sex. All kinds of molecular polymorphisms (SNPs, indels, microsatellites and transposable genetic elements) have been associated with variation for quantitative traits. While some variants have potentially functional effects on the translated protein, others are synonymous substitutions in protein-coding regions, or variants in non-coding regions with presumed regulatory effects.

What hope is there for dissecting the genetic basis of variation of quantitative traits?

In the past 20 years, there has been a shift from optimism to pessimism. At first, it seemed possible that QTL mapping could identify something like several to tens of loci with alleles of moderate to large effect that could explain quantitative traits and

complex diseases. Latterly, it has become clear that the task will be to identify unambiguously hundreds of genes with alleles with small effects affecting any one trait, and success seems more remote. The challenge becomes particularly arduous given context-dependent effects and the prospect of drilling down from QTL region to candidate gene one QTL at a time.

Several recent technical developments offer the hope of overcoming the difficulties, however. Two major obstacles have been the need for a dense panel of molecular markers for high-resolution mapping in the organism of interest, and for a way of genotyping these markers economically and in parallel in tens of thousands of individuals. Next-generation sequencing methods make possible the rapid identification of large numbers of polymorphisms in parental strains used in linkage-mapping studies, or a sample of individuals from a population targeted for association mapping, and several companies offer custom genotyping designs for massively parallel genotyping. As the cost of sequencing plummets, we can conceive of eventually determining the whole-genome sequence of every individual in a large population, pushing the challenge of genetic dissection of quantitative traits towards accurate and high-throughput phenotyping. In addition, molecular polymorphisms do not directly affect quantitative traits, but do so by altering levels of transcript abundance, amount and activity of proteins, metabolites and other 'intermediate' phenotypes. Incorporating measures of variation in intermediate phenotypes with genetic variation in molecular markers and quantitative phenotypic variation will provide a biological context in which to

interpret the phenotype. Finally, quantitative traits do not exist in a vacuum, but are connected to other traits via the pleiotropic effects of functional variants. Projects to develop sequenced genetic reference panels for model organisms as community resources for QTL mapping (for example, the mouse Collaborative Cross consortium, the *Drosophila* Genetic Reference Panel, and the *Arabidopsis* 1001 Genomes Project) will make possible large-scale measurement of multiple phenotypes, including intermediate phenotypes, in multiple environments. These resources offer the prospect of elucidating the genetics of the interdependence of multiple phenotypes, and addressing the longstanding question of the genetic basis of genotype-environment interaction.

Where can I go for more information?

Reviews

Mackay TFC: **The genetic architecture of quantitative traits.** *Annu Rev Genet* 2001, **35**:303–339.

Weiss KM: **Tilting at quixotic trait loci (QTL): an evolutionary perspective on genetic causation.** *Genetics* 2008, **179**:1741–1756.

Textbooks

Falconer DS, Mackay TFC: *Introduction to Quantitative Genetics*, 4th edition. Harlow, Essex: Longman; 1996.

Lynch M, Walsh B: *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer; 1998.

Published: 17 April 2009

Journal of Biology 2009, **8**:23

(doi:10.1186/jbiol133)

The electronic version of this article is the complete one and can be found online at <http://jbiol.com/content/8/3/23>

© 2009 BioMed Central Ltd