How Good Are Quantitative Complementation Tests?

Philip M. Service
The author is in the Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA. E-mail: Philip.Service@nau.edu

In this Perspective, I present a quantitative geneticist's view of life span and summarize the strengths and weaknesses of strategies that are being used by quantitative geneticists to find life-span genes in Drosophila melanogaster. I then describe the results of a recent study by Geiger-Thornsberry and Mackay (1) that used quantitative complementation tests to screen candidate life-span genes in flies. Complementation tests may seem to be a promising tool for finding life-span genes. However, I argue that they are highly problematic and likely to be misleading.

Quantitative Genetics and Life Span

A quantitative trait is measured on a numerical scale (years, centimeters, number of offspring, and so on). Life span in flies appears to be a typical quantitative trait. Variation in life span is controlled by at least several genes and is strongly affected by environmental factors. The Drosophila life span has moderate heritability, and average life span can readily be changed by selection experiments (2–9) (see Rose and Charlesworth Classic Paper). More recently, several laboratories have applied the techniques of quantitative trait locus (QTL) mapping in efforts to identify genes that affect life-span variation in flies (10–20). Taken at face value, such experiments suggest that there may be 20 to perhaps more than 100 such "life-span" genes.

It is important to understand that quantitative genetics, including QTL mapping, can reveal only segregating genetic variation that affects the trait being studied. For example, in a study using a laboratory population, only those genes that affect life span and that have two or more alleles (that is, that are segregating in the population) can contribute to the heritability of life span. Similarly, a typical QTL mapping experiment might start with a cross between two inbred wild-type fly strains. The only life-span genes that can be detected in such an experiment are those for which the two inbred lines are fixed (homozygous) for different alleles—in other words, those life-span genes that segregate in the progeny of the cross (the mapping "population").

Thus, for a quantitative geneticist, a life-span gene is any gene that affects variation in life span in a segregating (usually wild-type) population (natural or laboratory). Quantitative genetic studies tell us if and how life span can evolve in the short term (without additional mutational input) and can tell us something about the evolutionary genetic mechanisms that are responsible for senescence and for the maintenance of genetic variation for life span in wild-type populations (21–23). Finding life-span genes will enable quantitative geneticists to not only understand their effects on life span but also to evaluate their effects on other traits, such as reproduction. Such studies will help to determine
whether, for example, antagonism between life span and other traits works to limit life span and to shape patterns of senescence (24).

Quantitative Complementation Tests

Geiger–Thornsberry and Mackay report the results of a series of quantitative complementation tests designed to screen candidate life–span genes in a wild–type population of flies (1). The basic idea of a quantitative complementation test is relatively straightforward (Fig. 1). A mutant allele of the candidate gene is tested in association with alleles derived from, in this case, a natural fly population. The mutation is usually one that results in a non–functional or low–activity gene product (a loss–of–function mutation). If the candidate gene is a life–span gene, and if there is allelic variation in the natural population, then the natural alleles may "fail to complement" the mutation. More concretely, the effect of the mutation on life span may vary depending on which naturally derived allele it is tested against, as revealed by a significant interaction in an analysis of variance (ANOVA). (The logic behind this test is explained in greater detail below.)

![Fig. 1. Illustrative results of quantitative complementation tests. +/+i flies are mated to mj/Balj flies to produce two types of progeny: mj/+i and Balj/+i. +i refers to wild–type inbred line i. Results for two wild–type lines are shown, although experiments can be done with more lines. mj is a mutant allele of candidate gene j (or a deficiency that uncovers j), and Balj is a balancer chromosome. A two–factor ANOVA (or three–factor ANOVA if sex is also included as a variable) is used to analyze the results for, in this example, four different progeny genotypes (mj/+1, mj/+2, Balj/+1, and Balj/+2). (A) Results of an experiment with main effects due to chromosome type (mj versus Balj) and wild–type line (+1 versus +2), but no interaction. (B and C) Results of experiments that also include an interaction between wild–type line and chromosome type. Interactions in which the variance among the mj/+i flies is greater than the variance among the Balj/+i flies (B) are treated as failure to complement by the wild–type lines and are taken as evidence that j is a life–span gene. Interactions in which the variance among the Balj/+i flies is greater than the variance among the mj/+i flies (C) are thought to be due to epistasis between genes on Balj and genes in the wild–type genomes. They are not taken to be evidence that j is a life–span gene.

Unfortunately, appropriate mutations were not available for seven of the candidate genes that Geiger–Thornsberry and Mackay wanted to screen. To circumvent this problem, they used deficiencies that "uncovered" those genes. A deficiency is a chromosome with a deleted segment. A gene that is located within the segment that has been deleted is said to be uncovered when the deficiency is heterozygous with a chromosome that does not have the deletion. In the context of these complementation tests, a deficiency is analogous to a loss–of–function mutation for the candidate gene and for all other genes uncovered by the deficiency. Of the 17 mutant alleles and deficiencies that they tested, Geiger–Thornsberry and Mackay found five that were not complemented, suggesting five naturally occurring life–span quantitative trait loci. These were Alcohol dehydrogenase (Adh), which is involved in intermediary metabolism, and four genes or gene groups that were included within deficiencies: Accessory gland–specific peptide 26A (Acp26A) and Acp70A, both of which are involved in reproduction, a group of small Heat–shock proteins (Hsp22–Hsp28), which help the organism respond to stress, and the
In what appears to be a notable success of this approach, De Luca et al. (10) chose Dopa decarboxylase

**Insulin-like receptor** gene (InR), which modulates insulin signaling. Candidate genes that were complemented and thus apparently do not contribute to natural variation in life span were Glucose-6-phosphate dehydrogenase (G6pd), Phosphogluconate dehydrogenase (Pgd), α-Glycerol phosphate dehydrogenase (Gpdh), Superoxide dismutase (Sod), Catalase (Cat), rosy (ry), Punch (Pu), Epidermal growth factor receptor (Egfr), mutagen-sensitive 306 (mus306), Phosphogluconate mutase (Pgm), period (per), and Presenilin (Psn) (the latter three being tested by using deficiencies).

Although these complementation tests make use of mutations, they share an important feature with other quantitative genetic methods: Because the mutations are being evaluated against wild-type inbred lines, the tests are designed to detect naturally occurring variation in genes that affect life span. Geiger-Thornsberry and Mackay could have detected a life-span gene only if the 10 lines that they used were not all fixed for the same allele of that gene. This is an important consideration when assessing the value of their negative results. For example, suppose that in the source population there was a life-span gene with two alleles with frequencies 0.8 and 0.2. There would be about an 89% chance that this allelic variation would be "captured" in a sample of 10 inbred lines. That seems reasonably good. However, if the allele frequencies were 0.9 and 0.1, the corresponding probability of capture would be only about 65%.

**Methods for Mapping Life-Span Quantitative Trait Loci**

Most life-span QTL mapping studies in flies have used either linkage mapping or mutation/deficiency complementation mapping. Linkage mapping is typically accomplished by crossing pairs of inbred lines that differ at a large number of marker loci, and then evaluating recombinant progeny (such as F2 individuals or a series of recombinant inbred lines) to find statistical associations between marker loci and life span (10–13, 16, 17). There is a well-developed theory for mapping quantitative traits this way and for obtaining point estimates of QTL positions (25–28). Furthermore, familiar methods, such as permutation tests (randomly associating genotypes with life span) and bootstrapping (resampling one's original data with replacements to generate virtual data sets for reanalysis, so as to obtain some idea of the precision of a population parameter that one has estimated from one's data) can be used to control the probability of incorrectly identifying a QTL and to obtain valid 95% confidence intervals for QTL location (29–33). However, there are some serious problems with this approach: Most important, it has not, so far and by itself, identified any gene as a life-span QTL. Even with dense marker maps and large numbers of recombinant genotypes, the regions that have been identified as containing life-span quantitative trait loci are too large to be more than suggestive. To appreciate this problem, consider that the *D. melanogaster* genome consists mostly of three major linkage groups with a total map length of about 270 centimorgans (cM). There are estimated to be about 13,600 genes (34), or roughly 50 genes/cm. In studies in my laboratory, for example, most of the 95% confidence intervals for life-span QTL location have been at least 10 cm wide—a distance that includes perhaps 500 genes (17). There are ways to reduce the size of QTL confidence intervals [for example, see (35)], but this manipulation still will not allow the identification of specific genes.

Quantitative complementation mapping (similar in concept to the candidate-gene complementation tests of Geiger-Thornsberry and Mackay) has been proposed as a solution to the difficulties of linkage mapping. Pasyukova et al. (19) used overlapping deficiencies to dissect two broad life-span QTL regions that had been identified in an earlier recombination mapping study (10). They suggested that a minimum of 11 life-span genes lie within this pair of regions. De Luca et al. (20) used additional deficiencies to map one of the 11 QTL regions identified by Pasyukova et al. and concluded that there were at least three life-span genes within that cytogenetic region (36E; 38B). Complementation tests using deficiencies do not, by themselves, identify specific genes. However, the regions that are obtained can be much smaller than those identified by recombination mapping. They may, in fact, be small enough to start making educated guesses about specific loci that are known to be located within them, and to generate manageable lists of candidate life-span genes for further testing.

In what appears to be a notable success of this approach, De Luca et al. (20) chose Dopa decarboxylase...
(Ddc) as a candidate gene within a subsection of the 36E; 38B region. They then found that three mutants of Ddc were not complemented in tests similar to that shown in Fig. 1. Finally, they examined a large number of Ddc alleles from a natural population and found significant associations between single-nucleotide polymorphism haplotypes and life span. In short, Ddc appears to be a paradigm for mapping life-span genes: (i) a fairly broad chromosome region is identified by recombination mapping as containing at least one life-span QTL; (ii) the broad region is dissected by one or more rounds of deficiency complementation mapping in order to narrow estimates of QTL location and to reduce the number of possible candidate genes; (iii) candidate genes are then screened using loss-of-function mutations in quantitative complementation tests (the same test as used by Geiger-Thornsberry and Mackay); and (iv) candidate genes that pass the complementation tests are then examined for correlations between sequence variation and life-span variation in wild flies.

Why Quantitative Complementation Tests Are Problematic

Despite the apparent success of complementation tests (first with deficiencies and then with gene-specific mutations) in identifying Ddc as a life-span gene, there are reasons to believe that the Ddc story may be difficult to repeat. I suggest that failure to complement may be frequently observed because the statistical interaction that is used as the criterion can arise from epistasis rather than from "allelism." By allelism we mean that the mutation that is being tested actually is an allele of a life-span gene. Epistasis, on the other hand, refers to interactions among two or more genes. In this case, for example, the mutation being tested might interact with other genes to modify life-span. Epistasis of this sort would result if the effect of the mutation on life span depended on the genetic background at other loci. If life span is a complex trait that is affected by interactions among a large number of gene products, as most quantitative geneticists probably think, then we might expect epistasis for life span to be common, particularly when we are testing loss-of-function mutations. From the perspective of a quantitative geneticist, the gene being tested would not qualify as a life-span gene unless the epistatic interaction were characteristic of a wild-type population. In short, I believe that complementation tests are likely to yield many false positive results because there are several sources of epistasis. For the most part, my discussion here focuses on single-gene mutation complementation tests, although my arguments apply to deficiency complementation tests with even greater force.

The tests conducted by Geiger-Thornsberry and Mackay involved introgressing mutant alleles into a partially standardized genetic background. For example, a mutation for a chromosome 2 gene such as Adh was placed in a standard background for chromosomes 1 and 3. However, the chromosome 2 that contained the mutant Adh allele was not standardized. It was the chromosome that was supplied with the mutation. Therefore, all chromosome 2 mutations had the same genetic background for chromosomes 1 and 3, but each had a different chromosome 2 background. Prior to the complementation tests, mutants were maintained against balancer chromosomes. Balancers carry dominant mutations, so that they can be identified by their phenotypic effects and inversions that "suppress" crossing over with the chromosome that carries the candidate-gene mutation that is being tested.

Geiger-Thornsberry and Mackay took the trouble to use the same balancer chromosome for all mutations on a given chromosome (for example, CyO was used as the balancer for chromosome 2 mutants). In short, considerable pains were taken to reduce the likelihood of artifactual results. Even so, the experiments are problematic. The criterion for failure to complement is a statistical interaction in an ANOVA that compares four (or more) heterozygous "genotypes" (Fig. 1): (i) the mutant allele $m_j$ of candidate gene $j$ (together with its partially standardized background) placed against the entire haploid genome from wild-derived inbred line $+1$; (ii) a similar pairing of $m_j$ and wild-derived inbred line $+2$; (iii) the balancer chromosome $Bal_j$ used for $m_j$ (together with the same partially standardized background as used for $m_j$) placed against the haploid genome from inbred line $+1$; and (iv) the corresponding pairing of $Bal_j$ and $+2$. (For the sake of simplicity, I assume only two wild-type lines, as in Fig. 1. The essential points of my argument are not changed by the fact that Geiger-Thornsberry
and Mackay used 10 lines, although using 10 lines rather than two gave them more chances to include naturally occurring allelic variation in their experiment.)

There are at least four potential sources of genetic interaction in these tests, each of which can give rise to the statistical interaction that is detected by ANOVA. First, there can be a statistical interaction, because the two different wild-type alleles of gene \( j \) in lines \(+1\) and \(+2\) produce different life spans when each is made heterozygous with the loss-of-function mutation \( m_j \) for the candidate gene. This is the interaction that is attributed to allelism and that provides evidence that gene \( j \) is a life-span gene. I refer to it as a type 1 interaction. In order to detect the interaction, we also need to compare the effects of the two wild-type alleles of gene \( j \) in lines \(+1\) and \(+2\) when they are tested in association with the functional allele of \( j \) that is provided by the balancer chromosome \( \text{Bal}_j \). Fig. 1A illustrates a case where there is no statistical interaction and therefore no evidence that \( j \) is a candidate gene. Fig. 1B shows a statistical interaction that Geiger–Thornsberry and Mackay argue is evidence for allelism. A type 2 interaction involves epistasis between \( m_j \) and one or more other loci anywhere else in the wild-derived genomes of lines \(+1\) and \(+2\). Gene \( j \) would not, in that case, be considered a life-span gene unless its naturally occurring alleles also contributed to epistasis for life span. A type 3 interaction is due to epistasis between the genetic background of the chromosome that carries \( m_j \) and the wild-type genomes. A type 4 interaction involves the epistatic effects of the balancer chromosomes in conjunction with different wild-type genomes. Only a type 1 interaction identifies \( j \) as a life-span gene. However, all four types of interaction (one allelic and three epistatic) can produce statistical interactions that are indistinguishable from one another. There appears to be no conclusive way to distinguish between allelism and epistasis in quantitative complementation tests, a point that is conceded by Geiger–Thornsberry and Mackay. These authors do discount interactions when the variance in mean life span among the \( \text{Bal}_j/+/j \) flies is greater than the variance among the \( m_j/+/j \) flies (Fig. 1C). They attribute such interactions to epistasis between genes on the balancer chromosome and genes in the wild-type genomes (that is, type 4 interactions). The logic behind this argument is not clear: There seems to be no reason why all four types of genetic interaction cannot give rise to either pattern of statistical interaction (Fig. 1, B and C). Even if the logic behind this variance comparison test is correct, it may be a weak method for screening out epistatic interactions because (i) the power to detect unequal variances may be low; (ii) type 4 epistatic interactions might also give rise to the pattern shown in Fig. 1B; and (iii) the variance ratio test would not, in any event, appear to screen out type 2 and type 3 interactions. Deficiency complementation tests, as distinct from mutation complementation tests, are subject to similar sources of confounding epistasis, only perhaps more so. In particular, we might expect that type 2 interactions would be even more prevalent, because a block of genes is uncovered by a deficiency (typically 50 to >100 in the tests by Geiger–Thornsberry and Mackay). It is as if there were loss–of–function mutations simultaneously for all of those genes. That could represent a powerful source of artifactual epistatic interactions.

Conclusions

A major implication of my critique of quantitative complementation tests is that they are likely to yield false positives, especially when deficiencies are used. In this light, it may not be all that surprising that Pasyukova et al. (19) found evidence for 11 putative life–span genes within two relatively small genomic regions, or that De Luca et al. (20) found evidence for three more genes within one of the 11 regions identified by Pasyukova et al. The likelihood that at least some of those results are false positives seems even greater when it is realized that both studies used only two wild-type laboratory strains in their tests and that those strains did not have significantly different life spans. The implication is that those two "off–the–shelf" laboratory strains (essentially a sample of only two wild-type copies for each gene) were fortuitously fixed for different alleles for at least 13 different life–span genes in regions that accounted for only about 15% of the total linkage map for chromosomes 1 to 3. To understand this issue more fully, imagine that we sample two copies of a life–span gene that is segregating for two different alleles in frequencies 0.5 and 0.5. Then the probability that our sample of two gene copies will contain two different alleles (thus making it possible to detect the life–span
gene) is only 50%, and that is the best-case scenario when there are two different alleles. That suggests that the 13 life-span quantitative trait loci proposed in those two studies is actually a substantial underestimate of the number of life-span genes that would be revealed within those chromosome regions with a larger sample of gene copies. Alternatively, as I suggest, some, perhaps many, of those 13 putative life-span quantitative trait loci really are not.

Geiger-Thornsberry and Mackay tested seven deficiencies and found that four failed to complement. Even though the deficiencies were chosen specifically because they uncovered particular candidate genes, the "success" rate for deficiencies (4 out of 7) was markedly higher than for single-gene mutations (1 out of 9). (Two mutations were tested together, so there were only 16 rather than 17 separate complementation tests.) Although the difference in "success" rate between deficiencies and single-gene mutations was not statistically significant, it was in the direction that I predict. Geiger-Thornsberry and Mackay found that an Adh mutation was not complemented, suggesting that Adh might be a life-span gene. However, Deckert-Cruz et al. found no difference in ADH allozyme profiles between sets of long- and short-lived laboratory fly populations that had been derived from a common ancestor. That result does not necessarily contradict the finding of Geiger-Thornsberry and Mackay—the relevant variation with respect to life span may not be the amino acid substitution that distinguishes the two common ADH allozymes—but neither does it lend support. In a similar vein, Deckert-Cruz et al. did observe differences for the SOD and phosphoglucomutase (PGM) enzymes. Neither of those genes was implicated as a life-span gene by Geiger-Thornsberry and Mackay. Again, there are good reasons for such disagreement. For example, it is possible that the 10 inbred lines used by Geiger-Thornsberry and Mackay were all fixed for the same alleles of Sod and Pgm or that the effects of variation in SOD and PGM allozymes on life span were too small to be detected in their experiments.

An important test of quantitative complementation that appears not to have been done is an appropriate control experiment: What would the results be if deficiencies and mutations were chosen arbitrarily, rather than because they coincided with regions identified in recombination QTL mapping experiments or were thought, for some other reasons, to have a possible effect on life span? That experiment would begin to address the unanswered question of the prevalence of confounding epistasis for life span. If the work done to date is any guide, the search for a relatively complete set of genes that account for the natural variation of life span in flies is likely to be long and hard. Neither recombination mapping nor complementation tests are sufficient by themselves. I have focused on the problems of complementation tests, because the limitations of recombination mapping are more widely appreciated.

March 24, 2004
-------------

Comment on Article

References


*Copyright © 2004 by the American Association for the Advancement of Science.*