CHAPTER 8

Canalization, Cryptic Variation, and Developmental Buffering: A Critical Examination and Analytical Perspective

IAN DWORKIN
Department of Genetics, North Carolina State University, Raleigh, North Carolina, USA

Introduction 132
I. A Review of the Reviews 133
II. Empirical Concerns for the Study of Canalization 134
A. The Amount of Genetic Variation must be Controlled between Lines/Populations 134
B. The Need for Multiple, Independent Samples (Across Genotypes, Not Individuals) 135
C. Genetic Background must be Controlled for Comparisons between Treatments 135
III. Definitions of Canalization 136
IV. Reaction Norm of the Mean (RxNM) Definition of Canalization 137
V. The Variation Approach to Canalization 138
VI. Partitioning Sources of Variation 139
A. Variation within Individual (V_{WI}) 140
B. Variation between Individuals, within Genotype (V_{BI}) 140
C. Between-Line (Genetic) Variation (V_G) 140
VII. Inferring Canalization: When is a Trait Canalized? 140
VIII. What are the Appropriate Tests for Making Statistical Inferences about Canalization? 142
IX. In the Interim … 144

Variation

X. Analysis for the RxNM Approach 144
XI. The Analysis of Cryptic Genetic Variation 146
XII. Mapping Cryptic Genetic Variants 147
XIII. Is the Genetic Architecture of Cryptic Genetic Variation Different from that of Other Genetic Variation Involved with Trait Expression? 149
XIV. Now that I Have All of this Cryptic Genetic Variation, What Do I Do with It? 153
XV. The Future for Studies of Canalization 154
Acknowledgments 155
References 155

INTRODUCTION

In the folklore of evolutionary biology, one of the great wedges that occurred between the advocates of the Darwinian modern synthesis and other evolutionists concerned the fundamental issue of (heritable) variation. According to Darwin (1859), the variation that natural selection acted upon was, in general, quantitative. Bateson (1894) distinguished between continuous, meristic, and discontinuous variation and on the whole thought that the latter categories of variation were the targets of evolutionary forces. These opposing views diverged further during the development of population and quantitative genetics, where a fundamental assumption for most theoretical work (and statistical models) was that a very large number of loci, each with small (additive) effects, was responsible for trait expression. From this work, several models for the maintenance of genetic variation developed, such as mutation–selection balance, balancing selection, and overdominance, among others (see Hartl and Clark, 1997; Roff, 1997; for reviews). However, Waddington (1952, 1953) suggested an alternative mechanism to explain the maintenance of some genetic variation and with it an alternative model for the evolutionary process known as "genetic assimilation." The model of genetic assimilation predicts that in the face of unusual environment conditions, phenotypes can be genetically "captured" by the process of natural selection, if strong selection occurs. Implicit to this evolutionary model was a trove of hidden (cryptic) genetic variation for the trait, which was not generally observed (without the appropriate environmental stimulus), and a buffering mechanism, referred to as canalization, which helped to "store" the genetic variation. When the buffering mechanism failed (de-canalization), the cryptic genetic variation was released for selection to act upon. In the initial formulation of the model, if selection on this novel phenotype was strong (and consistent) enough, the new trait could itself then become canalized and be produced without the environmental stimulus. However, in later derivations of the model, it has been suggested that the assimilation process may not in fact occur by the mechanism as suggested by Waddington and that

selection may act to change the threshold of trait expression or rare alleles that affect phenotypic penetrance are coselected and are in fact responsible for the genetic assimilation (Stern, 1958; Bateman, 1959). This last category is known as the "Baldwin" effect. This chapter will not deal any further with the mechanisms behind genetic assimilation and instead will focus on assessing canalization and cryptic genetic variation (see Scharloo, 1991, for review of the genetic assimilation controversy).

Regardless of the concerns with the mechanistic explanation of genetic assimilation, the plausibility of the phenomena of genetic assimilation as well as the existence of cryptic genetic variation were established via some empirical experiments. Waddington (1952, 1953) demonstrated that traits that were invariant under most (normal) environmental circumstances could be sensitized so as to express phenotypic variation for these traits. The classic example of Waddington's was the use of a high-temperature "heat-shock" in *Drosophila*, which resulted in some flies having lost their wing cross-veins. Waddington demonstrated that the cross-veinless phenotype could be selected upon, suggesting considerable hidden (cryptic) genetic variation for this trait (Waddington, 1952, 1953). Later work demonstrated that these observations could be extended to other environmental perturbations and traits (Waddington, 1956; Bateman, 1959) as well as to genetic perturbations (Rendel, 1959). However, all of these studies (and later ones) sufficiently demonstrated that the buffering mechanism (canalization) and the cryptic genetic variation being suppressed are intertwined (although this does not imply that the cryptic genetic variants are themselves responsible for the buffering).

I. A REVIEW OF THE REVIEWS

Given that a number of excellent reviews on canalization have appeared recently, I will provide a short overview of these papers, before proceeding with this chapter. Scharloo (1991) reviews the classic canalization literature spanning the 1950s through the 1980s. As I will discuss at length, there is little or no consensus as to the definitional issues regarding canalization and the apparently related biological processes of phenotypic plasticity and developmental stability. Two recent reviews (Debat and David, 2002; Nijhout and Davidowitz, 2003) discuss many of the definitional concerns, and historical constraints with the literature regarding canalization. Within the larger context of buffering, the mechanisms governing canalization and its evolution, De Visser *et al.* (2003) is quite comprehensive. For an alternative introduction to the principles of buffering, see Hartman *et al.* (2001).

In terms of exploring plausible mechanisms for genotype to phenotype mapping and canalization, Rutherford (2000) and Nijhout and Davidowitz (2003) both offer useful ideas to consider. Gibson and Wagner (2000) explore some of the methods for the detection of canalization and provide a general overview of the canalization literature. One of the few papers to focus almost exclusively on

the evolutionary forces that may be responsible for canalization is the one by Meiklejohn and Hartl (2002). Several reviews focused more on the cryptic genetic variation part of the relationship (Wagner *et al*., 1999; McLaren, 1999; Gottleib *et al*., 2002).

For the remainder of this chapter, I will focus on some of the experimental and inferential issues involved with the empirical study of canalization and cryptic variation. I will not discuss at any length some of the theoretical work that has been done with respect to canalization, although this does not imply that I feel it should be in any way neglected. I simply do not have sufficient room to deal with this issue.

II. EMPIRICAL CONCERNS FOR THE STUDY OF CANALIZATION

Before delving into the statistical concerns involved with making inferences with regard to canalization, I feel it is important to address some issues about experimental design. In the next section, I will discuss the alternative definitions of canalization and the consequential effects on inference. However, regardless of the definition or metric used for canalization, there are certain experimental concerns to be addressed. As with so many subjects of inquiry, there is no single optimal design for experiments of canalization. However, there are a few specific areas of concern, which can have an impact on the interpretation of results garnered from the experiment. The three major concerns are as follows:

- The amount of genetic variation must be controlled between lines/populations.
- The need for multiple, independent samples (across genotypes, not individuals)
- Genetic background must be controlled for comparisons between treatments.

A. The Amount of Genetic Variation must be Controlled between Lines/Populations

It must always be kept at the forefront of the mind of researchers that canalization is not a property of a species or population, but of a genotype. For most species, each individual has a unique genotype, which means testing questions of canalization can be extremely difficult. If a population is studied under several environments, the results may have little to do with canalization, but the differential response of different genotypes (i.e., different genotypes in the population have differential fitness across environments). Thus it is important to employ some method to control for within-line genetic variation.

In most genetically tractable systems and in those species that reproduce clonally, it is possible to get closely related individuals via inbreeding, or controlled genetic crosses (i.e., via chromosome extraction procedures in *Drosophila*). For other

species in which some genetic manipulation is possible, basic crossing schemes can be used to control for at least some of the genetic effects. This is of fundamental importance, so that not only can the same genotype be examined under multiple environments, but reasonable sampling of a given genotype can be performed.

B. The Need for Multiple, Independent Samples (Across Genotypes, Not Individuals)

This is one of the more common oversights in studies of canalization. Given that each line represents a single genetic sample, measuring multiple individuals within a line essentially increases the sampling (providing a better estimate) of a single measure. Thus, if two lines are used, the effective sample size can be considered two (not $2 * n$, where n = number of individuals sampled within line). Debat *et al.* (2000) used this approach (only two lines) when examining the patterns of within-individual variation and canalization. For some situations, this may be sufficient; however, it must generally be regarded with caution. For a feasible experiment, a balance must be found between the number of individuals sampled within a line (to get precise estimates) and number of independent lines to use (for statistical power). These issues are too complex to consider here, but needless to say depend highly on what central question is being addressed.

On a further note of caution, when choosing the multiple lines to use, it is important to consider the "independence" of the lines. In this context, independent means genetically unrelated. Otherwise, there can be "pseudoreplication" of the data, where the same genotype is being resampled, even though they are being treated as statistically independent. There are several studies that examine patterns of variation that may suffer from artifactual effects because of this type of problem (Woods *et al.*, 1999; Bourget, 2000). However, whether this approach leads to a "fatal" flaw is unclear and must be investigated further. It is also unclear how "unrelated" the individuals must be. Independent sampling of wild-caught individuals (even from the same population) will likely be sufficient for outcrossing species. However, this too must be empirically addressed.

C. Genetic Background must be Controlled for Comparisons between Treatments

This issue is pertinent to studies of genetic canalization where different mutants (or chromosomes) are being compared for their effect on canalization. For genetically tractable systems, it may be possible to have controlled genotypes where a given chromosome is crossed into many different genetic backgrounds. However, an approach that, although requiring some initial time and energy, is generally

acceptable is to introgress the mutation via repeated backcrossing to an inbred line into the genetic backgrounds of interest. Several studies have demonstrated the efficacy of this approach in a number of model systems (Gibson and van Helden, 1997; Alonso-Blanco *et al*., 1998; Polaczyk *et al*., 1998; Gibson *et al*., 1999; Bolivar *et al*., 2001, Sershen *et al*., 2002; Dworkin, 2005). In a species such as *D. melanogaster*, 10 generations of backcrossing the mutant (or other marker) leads to ~90% of the genome being replaced. The exact amount will depend largely on the rate of recombination in that region of the chromosome. This will allow ideal comparison between mutant and wild-type conspecifics (see Gibson and van Helden, 1997, for details on the procedure), who are essentially identical except for the allele of study. If genetic background is not controlled, then it will be unclear if the observed effect results from the mutation in question or some other loci in the genetic background of the mutant.

III. DEFINITIONS OF CANALIZATION

If one is to study a subject, it would be expected that there be a clear definition of the object of study. However, although there have been numerous studies on canalization, there is no agreement as to what defines it. Several studies have outlined the historical and logical reasons for the different definitions (Debat and David, 2001; Nijhout and Davidowitz, 2003). After providing a brief overview of these definitions, I will outline some empirical metrics for canalization given the different definitions.

Broadly speaking, there are two main definitions, which I call the reaction norm of the mean (RxNM) and the variation approach. Why should we care which definition of canalization we are using? As we will see, among other disparities, the different definitions lead to different metrics for the study of canalization. Before we continue with this discussion, several issues must be made clear about some definitions that we are using. First, when I speak of "line" effects, I am assuming that each individual within the line is genetically identical (or very close relatives), while different lines are sufficiently diverged from one another that they are "independent" in the statistical sense. Furthermore, the use of the term *environment* throughout this chapter refers not only to typical external macroenvironmental variables (such as temperature, salinity, or density), but also can refer to an internal environmental variable such as a mutation (i.e., the two environments could be wild type and mutant). The use of this common term *environment* to refer to both genetic and external causes of effects is not meant to imply a common mechanism (*sensu* Wagner *et al*., 1997; Ancel and Fontana, 2000; Meiklejohn and Hartl, 2002), but is simply used as a matter of convenience.

Although I hope the reader will clearly recognize the distinction between the following definitions, at this point it is unclear which definition is in fact the

correct one. Thus I suggest that, for future studies of canalization, both definitions (and derived metrics) be employed to address questions of canalization (and the relationship between the metrics).

IV. REACTION NORM OF THE MEAN (RxNM) DEFINITION OF CANALIZATION

Figure 8-1 (A, B) helps illustrate the first definition of canalization and its relationship to phenotypic plasticity. Figure 8-1A illustrates a "typical" reaction norm for a trait under study. Each line represents a genetic "line" (genotype) for which we have sampled multiple individuals in each of two environments (E1 and E2). Although we do not need to assume that there is no microenvironmental variance within E1 or E2, we do need to assume that it is equal. In Figure 8-1A, lines A, C, and E all have been observed with different means in each environment. Furthermore, these particular lines display what is commonly known as crossing of line means. However, lines B and D (Figure 8-1A) show virtually no difference in their trait means across environments (although the line means differ from one another). According to the RxNM definition, canalization is the opposite of phenotypic plasticity (Nijhout and Davidowitz, 2003). Lines B and D (Figure 8-1A) would be considered canalized with respect to environments E1 and E2, i.e., there is no environmental effect on trait expression.

How do we infer canalization for the RxNM approach? The majority of studies infer canalization by the decanalization of a system via an environmental (genetic or exogenous) stress that causes perturbation to normal trait expression. If we

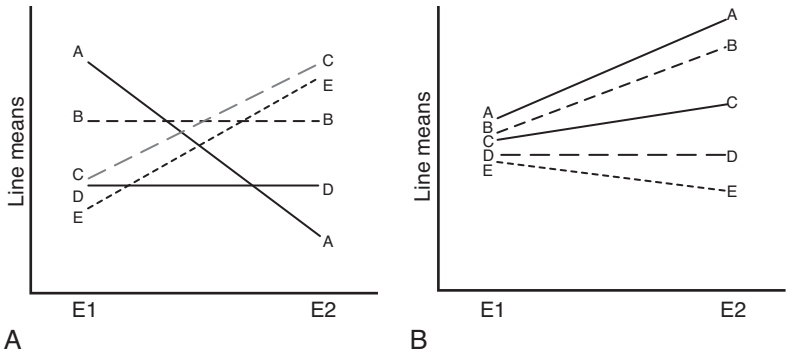

FIGURE 8-1. Reaction norm demonstrating phenotypic plasticity and canalization as opposite characteristics of the same phenomenon. (A) Lines A, C, and E all display classic phenotypic plasticity (change in trait mean across environments), while lines B and D show a form of canalization (no change in trait values). (B) The test of canalization for this metric is a change in the between-line (genetic) variation from one environment to another. E1 and E2 represent the two environments.

observe a change in phenotypic variation, we can ask whether we infer the release cryptic genetic variation for the trait. In the case of inferring canalization by decanalization, Figure 8-1B illustrates an idealized example of what we are seeking. In this example, lines A–E show relatively little between-line variation in E1. However, when individuals from these same lines are "exposed" to E2 (the stressful environment), we see a significant increase in the between-line variation (because of a release of cryptic genetic variation for trait expression). Fundamentally, when we are using the RxNM definition of canalization, this is what we are looking for (details on the process of statistical inference will be discussed in a following section). By this definition, the more canalized a line is, the less its mean should change across environments, and this in fact can become our definition of canalization for each line (denoted *C*). Specifically, we are interested in the unsigned deviation of the line means in environment E1 and E2 or

$$Cj = |\mu_{j,\,E1} - \mu_{j,\,E2}|$$

where *Cj* is the measure of canalization for line *j*, $\mu_{j,\,E1}$ is the line mean of *j* in E1.

V. THE VARIATION APPROACH TO CANALIZATION

The other major definition of canalization, which I simply define as the "variation" metric, is fundamentally concerned not with how the line mean changes across environments, but how the measure of variation *within* line changes (Stearns and Kawecki, 1994). This idea is illustrated in Figure 8-2(A, B). In Figure 8-2A, we observe the measures for two genetic lines (solid and dashed), where we have "error" bars, which represent some measure of within-line variation. The dashed line shows no difference in within-line variation in E1 or E2 (again the stressful environment). However, the solid line shows a significant increase in within-line variation in E2. We can illustrate this as a "reaction norm" (Figure 8-2B). However, it is important to note that the Y axis no longer represents line means, but within-line variation. The line means have been illustrated as invariant in Figure 8-2A for purposes of simplicity and do not necessarily reflect observed biological pattern.

Unlike the RxNM approach, for the within-line variation definition of canalization, the appropriate metric of canalization is less clear. For instance, if a trait is measured in a number of genetically distinct lines in a single environment, we may observe that some lines show more within-line variation than others (Figure 8-2A, environment E2). It could be argued that the lines that show lower levels of within-line variation are better canalized than other lines, even though all traits were measured in a single environment (even though this is far from the traditional definition of canalization and has been given other names; see Debat and David, 2001,

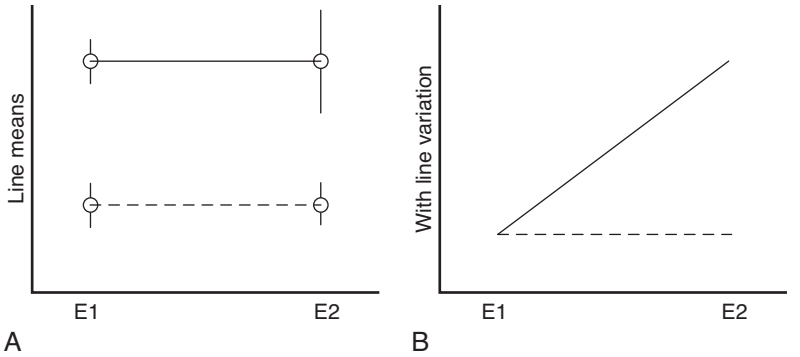

FIGURE 8-2. Variation approach to canalization. (A) Traditional reaction norm graph for two lines, demonstrating that, although the trait means do not change across environments, their within-line variance does, with the dashed line apparently being canalized (no change in within-line variation), while the solid line shows a significant increase in variation in E2 compared with E1. (B) Another view of the same phenomenon, but using a measure of within-line variation on the Y axis.

for a review of these definitions). Alternatively we can use an analogous approach to that for the RxNM, where we are not so much interested in the levels of within-line variation in a given environment, but how it changes across environments. The metric for this approach would be:

$$Vj = |CV_{j,\,E1} - CV_{j,\,E2}|$$

Where Vj is the measure of canalization for line j, $CV_{j,\,E1}$ is the coefficient of variation (or some other measure of within-line variation; see section on statistical inference) of j in E1.

VI. PARTITIONING SOURCES OF VARIATION

In an effort to make the distinctions between these definitions of canalization clearer, it may help to consider briefly the different sources of variation. Assume that we have a number of genetically distinct isogenic lines, where individuals within a line (who are all genetically identical) are all measured in a common set of environments. I will not demonstrate the statistical procedures of how to partition the variation (see Falconer and Mackay, 1996; Lynch and Walsh, 1998; and Palmer and Strobeck, 2003, for issues relating to V_{WI}). However, it is important to distinguish between the variation that is actually being measured and the sources of variation that are inferred (environmental or genetic).

A. Variation within Individual (V_{WI})

This is most commonly measured either on a serially repeating trait or more often using the two sides of a bilaterally symmetrical individual (i.e., measuring wing length on both the left and the right sides of a fly). The deviation of left versus right sides is a measure of asymmetry (often fluctuating asymmetry), which has a long and distinguished literature dealt with in Chapter 10 of this volume. There are a number of important biological issues that should be considered with respect to V_{WI}. First, there is no genetic variation (unless the somatic mutation rate is high), and second, there is no environmental variation (but see Nijhout and Davidowitz, 2003). It is generally assumed that V_{WI} is a proxy for the developmental noise of the organism (i.e., random developmental differences between left and right sides of the individual that cannot be controlled for).

B. Variation between Individuals, within Genotype (V_{BI})

The *CVj* discussed in the previous section is one particular metric of this. Some authors have suggested that this is an equivalent metric to V_{WI} if all individuals are raised in a common environment, with little or no uncontrolled variation, and there is some evidence to support this (Clarke, 1998).

C. Between-Line (Genetic) Variation (V_G)

Assuming all individuals (of all lines) have been raised in a common environments, this component represents the between-line variation. This is essentially what we are interested in for the RxNM metric of canalization (and how it changes across environments) when we infer canalization by its breakdown.

VII. INFERRING CANALIZATION: WHEN IS A TRAIT CANALIZED?

Given that the vast majority of studies of canalization have utilized (implicitly) the RxNM approach, we will spend some time dealing with inferences of canalization for the metric derived from it. When we perturb a system to cause decanalization (and infer canalization from this), we are actually measuring two related processes. First is the buffering component (canalization) of the system/trait. However, if there were no available "noise" in the system (i.e., genetic or environmental variation), then decanalization could not be observed (Nijhout and Davidowitz, 2003). Thus we are often measuring the release of the cryptic genetic variation (in the perturbed state) and inferring canalization (in the unperturbed state).

Let us begin with a hypothetical trait that shows "ideal" canalization (Figure 8-3). As can be seen in Figure 8-3, we have measured a number (four) of lines, in four environments (of which we can measure the magnitude of environmental effects). The between-line variation is greater in E1 and E4 than in either E2 or E3. This is an "ideal" trait for studies of canalization because we have environments in which we know *a priori* the environmental region for which the trait is in the "zone of canalization" where it varies little. The environmental effects of E1 and E4 are sufficient to perturb trait expression (decanalization), and we observe an increase in between-line variation in E1 or E4 (i.e., a release of cryptic genetic variation) relative to E2 or E3. If we observed this pattern of effects, then it would be relatively easy to make an argument about the canalization of the trait.

However, in the majority of studies dealing with canalization, only two environments are used. Usually one environment is considered "normal," while the second environment is "stressful." If *a priori* there were no information as to how a given environment affected the trait, then it would be possible that the experiment could run into some difficulties. If by chance the two environments compared were E1 and E2, E1 and E3, E2 and E4, or E3 and E4, then we would observe the increase in between-line variance and could make an inference about canalization from there. However, in comparisons of either E2 and E3 or E1 and E4, we would not observe any significant difference across environments for the between-line variance. From this, it would be possible to make two conclusions: The trait is canalized or it is not. This is not very useful information. With the comparison between E2 and E3, we are in fact observing a trait in its zone of canalization, but this is obviously not the case for E1 and E4. The reverse argument also holds.

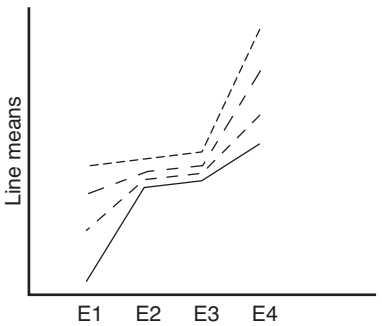

FIGURE 8-3. An idealized example of a canalized trait. For this trait, there is apparently a "zone of canalization" between environments E2 and E3, while the two extreme environments seem to cause a perturbation in trait expression resulting in an increase of between-line variation (a release of cryptic genetic variation). We assume that the environments E1 and E4 form some type of gradient of effect, even though we are measuring the traits in "discrete" environments.

Therefore, if only two environments are going to be used, it is important that there is prior information to inform which environments are chosen.

Although two environments are sufficient to make an inference about canalization, as wishfully concluded from the preceding argument, it is valuable to include more than just two environments. For external environmental effects (temperature, nutrients, density), this is usually a possibility. However, when the different environments being examined are in fact genetic in nature (i.e., different mutations), this can require additional introgressions (see the preceding text) that can be considerable work. As well, it is not always clear what sort of "gradient" of effects the mutations have. For future studies of canalization, it will be important to examine the shapes of the reaction norms within and outside of the "zone of canalization," and we recommend this type of approach for those interested.

VIII. WHAT ARE THE APPROPRIATE TESTS FOR MAKING STATISTICAL INFERENCES ABOUT CANALIZATION?

Clearly, given the multiple definitions of canalization, we are confronted by at least as many possible metrics. However, even within a given definition of canalization, it is not clear what is the best possible method for statistical inference. Although another chapter in this volume (Chapter 2) deals specifically with the statistics of variation, I will highlight a few approaches that have been used to address questions of canalization. We should point out at the outset that we are not in fact advocating any particular method, and a thorough analysis of the comparative properties of the various tests of variation is still required.

If we assume that canalization is best measured with respect to within-line (genotypic) variation, in a given environment, then perhaps an initial approach would be to use the coefficient of variation (*CV*) as a measure. The coefficient of variation is a standardized, dimensionless quantity measured by dividing the sample standard deviation by the sample mean (σ/μ, often multiplied by 100%). This approach has been used by Stearns and Kawecki (1994). However, there are certain issues with respect to the use of *CV*. Unlike an estimate of the mean for a trait, the sample size required for an accurate estimate can be quite large, and for small sample sizes, it is biased (Lande, 1977; Sokal and Braumann, 1980; Sokal and Rohlf, 1998), although there are corrections for this metric (Sokal and Braumann, 1980). If we simply compare individuals of a single genotype in two environments, then there are a number of options for how to make a statistical inference (and to date it is not clear which option is most appropriate for a given situation). Lewontin (1966) advocated computing an *F*-test statistic on the sample standard deviations computed from the log-transformed data, or on the CV^2 when

CV < 30%. Sokal and Braumann (1980) suggested the use of a *t*-test using the standard errors of the corrected *CV*s. Zar (1999) advocates the use of an asymptotic test statistic (Miller, 1991). Finally, Schultz (1985) advocated the median form of the Levene's test. Thus we are left with a rather large set of possible tests to do. Clearly, it is not appropriate to perform all of these tests.

Schultz (1985) compared a number of these tests via simulation and advocated the median form of the Levene's statistic. However, an examination of the results from these simulations clearly shows that the median form of Levene's statistic can be overly conservative for two groups, given normally distributed data. In this example (i.e., normally distributed *CV*s), the *F*-test statistic on the CV^2 appears to be the most appropriate. Van Valen (1978) has argued that even when the data appears normal, one has to be very cautious about the *F* test because of its sensitivities to nonnormal distributions. However, until further work has been done to determine which of these tests are most appropriate, I would suggest referring to the results in Schultz (1985) to determine the best recourse depending upon the data at hand. In my experience, for the majority of cases, the different tests do not make wildly different inferences, unless the data is far from normal.

If trying to decide on the appropriate test for a two-group situation seems baroque, then for $k > 2$ group case, it can seem downright daunting. The majority of these tests are labeled as tests for the homogeneity of variances (one of the assumptions of analysis of variance [ANOVA] models). Sokal and Braumann (1980) suggest using either the Bartlett's or Levene's test on the log-transformed variates. An alternative is an extension of the asymptotic test for the equality of *CV*s discussed earlier (Feltz and Miller, 1996). Schultz (1985) demonstrates that the median form of Levene's test is again the most appropriate test for the $k = 3$ example, as compared with Bartlett's test, which is quite sensitive to nonnormality of the data (Zar, 1999). Unfortunately, to date there has been no comparison of each of these methods. One advantage of using the Levene's statistic (either the median or the mean form) is that as a metric it is easily cast in the format of an ANOVA, which allows complex models with interaction terms to be explored. For studies of canalization, where there is likely a line component and an environmental component, being able to examine this interaction term can be quite useful, if not imperative (depending upon the definition of canalization). Given that we infer canalization through a release in cryptic genetic variation, we are fundamentally interested in changes in the genetic variation. Within the context of an ANOVA, there has been some work that has specifically explored this issue (Aitkin, 1987; Foulley *et al.*, 1994; Sancristobal-Gaudy *et al.*, 1998; Sorensen and Waagepetersen, 2003). It is important to the field that all of the methods described in the preceding text are explored and compared to determine the most appropriate methods for future empirical studies, although it is likely that different methods will suit different designs.

IX. IN THE INTERIM …

However, until the properties of the various tests have been more fully explored, I provisionally suggest a course of action. Given a study with k lines, and j environments (where k and $j > 1$), I suggest that the Levene's statistic on the log-transformed data be used, for both its relative simplicity and how readily it can be used for complex models. Even though the distribution of Levene's statistic will be a truncated normal, given that it is the unsigned deviation, it generally has the appropriate type I error when comparing groups with no differences (Schultz, 1985; Palmer and Strobeck, 1992; Palmer, 1994). Furthermore, permutation tests can be performed on the data if there are specific concerns about the distributions of Levene's statistic.

$$LS_{ijk} = \left| \text{Log}(x_{ijk}) - E[\text{Log}(x_{jk})] \right|$$

Where LS_{ijk} is the Levene's statistic for individual i, for line j in environment k. $E[\text{Log}(x_{jk})]$ is the mean of the log-transformed data for all individuals for line j in environment k. Schultz (1985) suggests that the mean approach also tends to be anticonservative when the data is not normally distributed and suggests the use of the median $\{Md[\text{Log}(x_{jk})]\}$ as an alternative measure of central tendency. It is worth computing and comparing, although they tend to give similar answers under many empirical circumstances. If Levene's statistic is used, then the measure of canalization across environments becomes (by substituting *LS* for *CV*):

$$Vj = \left| LS_{j,\,\text{E1}} - LS_{j,\,\text{E2}} \right|$$

As discussed earlier, it is not clear for this definition what the correct measure of canalization is. Therefore, I suggest following the ANOVA framework as per the RxNM approach described in the following text (using *LS* as the metric as opposed to the trait value for each individual).

X. ANALYSIS FOR THE RxNM APPROACH

When canalization is viewed as the "opposite" to phenotypic plasticity, then the framework for analysis is somewhat clearer. Given k lines (L) and j environments (E), we can start out within the framework of an ANOVA. With some variant of the model

$$Y_{ijk} = \mu + E + L + E \times L + \varepsilon$$

As with the analysis of phenotypic plasticity, we can begin by examining the significance of these model terms. Evidence for genetic variation for "plasticity" can be

inferred if there is a significant $E \times L$ term for the model. If this term is not significant, but both E and L are, then there is evidence for plasticity of the trait and genetic variation for the trait itself, but not for genetic variation for plasticity of the trait (Figure 8-4).

If there is a significant $E \times L$ term, then we need to determine whether there is in fact evidence for canalization. It is important to recognize that a significant $E \times L$ term can arise from different processes (Robertson, 1959; Gibson and van Helden, 1997; Lynch and Walsh, 1998; Gibson and Wagner, 2000), but not all are evidence for canalization. Specifically, we want to separate out cases where the $E \times L$ term arose because of significant line crossing (change in relative ranks) across environment (Figure 8-1A), as opposed to a change in the scaling of the line means across environments (Figure 8-1B). For this latter case (i.e., Figure 8-1B), this will result in a perfect correlation across environments for the line means. For studies on phenotypic plasticity, the $E \times L$ term is usually partitioned into the two components (Robertson, 1959; see Lynch and Walsh, 1998, for review). However, knowing what proportion of the variation results from line crossing versus scaling effects is not itself of interest for canalization. We are specifically interested in whether this scaling effect (i.e., the increase in between-line variance) is significant.

One line of evidence that is consistent with canalization of a trait is a release of cryptic genetic variation in the "stressful" environment. This is inferred by the increase in between-line variation in the stressful environment (Gibson and van Helden, 1997; Gibson and Wagner, 2000). Thus, *a priori,* we recognize the need for some supplementary tests to determine whether there is a release of cryptic genetic variation. Here we return to many of the same statistical issues (and variety of tests) that we used to examine patterns of variation in the preceding section.

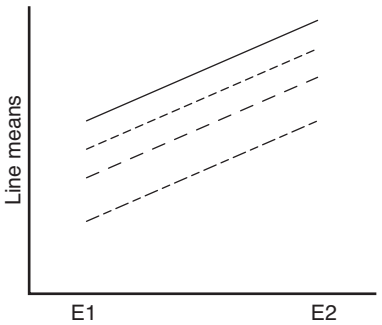

FIGURE 8-4. A reaction norm plot showing no evidence for a genotype by environment (genotype–environment interactions [GEI] or $E \times L$ effect in the model in the text) contribution. Although there is evidence for line (genetic) effects because the line means differ, and for environmental effects (because the mean obviously differs across environments), all of the slopes are identical.

Gibson and van Helden (1997) used the *F*-test approach, by comparing phenotypic variance across environments, as well as comparing the variances from between-line means (an estimate of genetic variation, but see Lynch and Walsh, 1998, for concerns with this approach). If using this approach, then we suggest the use of the CV^2 as opposed to the variances to cast the *F*-tests (Lewontin, 1966; Schultz, 1985). If this test is significant, then there is evidence for an increase in phenotypic variation in one environment over the other. The *F*-test used in this context suffers from all of the same inadequacies as discussed earlier. Demonstrating an increase in phenotypic variation is not sufficient to conclude canalization for the trait. It is the test of the variation of between-line means (again using CV^2) that will (if significant) infer a release in cryptic genetic variation (and thus provide evidence for canalization). It should be pointed out that, unless a large number of lines are used, this criterion may be hard to meet as formal significance will be unlikely (even if there is indeed cryptic genetic variation). Therefore, one of the other tests described in the preceding section may in fact be preferred (such as Levene's test). Of course all the tests can be employed, but formal significance must be adjusted (by Bonferroni or other such approaches) to control for multiple tests. In the interim, we again suggest the use of Levene's test to test for an increase in the between-line variation. In this case, the test is perhaps best framed as follows:

$$LS_{jk} = \left| \mathrm{Log}(\mu_{jk}) - E[\mathrm{Log}(\mu_k)] \right|$$

where LS_{jk} is the Levene's statistic for line j in environment k. $\mathrm{Log}(\mu_{jk})$ is the log-transformed line mean for j in environment k, and $E[\mathrm{Log}(\mu_k)]$ is the mean of the log-transformed line means in environment k. If this approach is used, and there are two environments (k = E1, E2), then a paired t test may in fact be the most logical test using the pairs ($LS_{j,\,\mathrm{E1}}$ and $LS_{j,\,\mathrm{E2}}$).

XI. THE ANALYSIS OF CRYPTIC GENETIC VARIATION

Implicit from the first studies of canalization was the realization that there was a significant amount of standing hidden (cryptic) genetic variation for traits. Waddington (1953) demonstrated that the wing venation phenocopies produced via a temperature heat shock could be selected upon to increase or decrease their frequencies (and ultimately the phenotypes could be expressed without the environmental stimulus). Of some consequence, Waddington (1953) demonstrated that some of the cryptic genetic variation for this trait was allelic to known developmental mutants (whose phenotype was being phenocopied). Subsequent work demonstrated that this was also the case for other phenotypes and genes

(Waddington, 1956; Gibson and Hogness, 1996). Given that there is evident abundance of cryptic genetic variation, we are left with an interest in the genetic architecture of cryptic genetic variation for traits. Several questions immediately arise from this.

Some of the standard questions pertaining to the genetic architecture of a trait include (see Mackay, 2001, for a complete list):

- How many genes are involved with trait expression?
- What is the distribution of the (magnitude of) effects of each of the genes?
- How do the genes interact with each other and the environment with respect to trait expression?
- What is the mutation rate of the genes involved with trait expression?

In some sense, these questions (and ones that derive from them) are addressed because they inform us of the evolutionary potential of the trait (evolvability).

However, with regards to cryptic genetic variation for a trait, there are a few more questions that we must ask. First and foremost, we are interested in whether the genetic architecture is in any way different than that for trait expression itself (i.e., is the genetic architecture of bristle number different than the genetic architecture for the canalization of bristle number?). I will provide some evidence here that in fact there is no evidence for difference in genetic architecture (or molecular polymorphisms within genes) to suggest this. Second, we are interested in whether all cryptic genetic variation (not just that caused by genetic perturbations) is caused by epistasis. If we can in fact examine the actual polymorphisms within genes that are responsible for the cryptic genetic variation, we can try to understand the evolutionary forces that have been maintaining them (i.e., do those regions of the gene seem to be under specific evolutionary forces, or are they evolving in a more or less neutral fashion, are they common or rare alleles?).

As evolutionary biologists, we are interested in cryptic genetic variation for a number of reasons. Does the existence of cryptic genetic variation allow for an increased rate of evolutionary response (i.e., can a trait evolve faster with the presence of cryptic genetic variation than without it)? Second, why do we not observe the effects of cryptic genetic variation under most environmentally relevant circumstances?

XII. MAPPING CRYPTIC GENETIC VARIANTS

Although the early studies demonstrated some of the genetic basis of cryptic genetic variation, if we are interested in the evolutionary dynamics of the alleles responsible for this phenotypic variation, we must extend our analysis. Gibson and Hogness (1996) did just this by following single-stranded conformational polymorphism (SSCP) marker frequencies in the *ultrabithorax* (*Ubx*) region in a

D. melanogaster population that was selected for increased sensitivity to ether-induced haltere to wing homeotic transformations. They observed that the ether exposure was correlated with a decrease in the amount of *Ubx* transcript found in the haltere imaginal disc. In addition, selection for the homeotic phenotype was itself correlated with an increase in certain markers, which were likely in linkage disequilibrium with polymorphisms under strong selection. This was the first such demonstration of intermediate frequency polymorphisms being subject to selection for such a "cryptic" trait.

However, the design used in this study did not allow a test of whether those polymorphisms scored were in fact responsible for the phenotypic variation observed. A recent study has overcome this difficulty using another trait that demonstrates cryptic genetic variation, namely photoreceptor determination in *D. melanogaster*. Normally the number of photoreceptors per ommatidia is invariant at eight. However, when the system is perturbed genetically, the number of photoreceptors can be altered. Via the introgression of the dominant $Egfr^{E1}$ allele into a panel of isofemale lines, the effect of genetic background on this allele was studied (Polaczyk *et al.*, 1998). It was demonstrated that (1) there was a considerable amount of phenotypic variation for this trait and (2) a large portion of this variation was genetic in nature (74% of the variation was estimated to be genetic). This result suggests that a surprising amount of hidden cryptic genetic variation is available (and that there is a sufficient mechanism of canalization to buffer against the effects of this variation).

Based on these results, we asked whether we could in fact map (to the nucleotide level) the polymorphisms responsible for this cryptic genetic variation. Given that previous studies have suggested that some of the cryptic genetic variation for traits was allelic to known genes whose phenotypes were being phenocopied, we decided to investigate how natural genetic variation in *Egfr* itself modifies the effect of the $Egfr^{E1}$ allele (Dworkin *et al.*, 2003). In this study, we observed significant association between several single nucleotide polymorphisms and phenotypic variation for eye roughness (a proxy for photoreceptor number) and replicated the most significant associations with an independent sample and statistical approach. Interestingly, we also found some evidence to suggest that there was mutation–selection balance acting on this trait–gene association. Thus this provided evidence that at least one source of cryptic genetic variation for a trait was in fact in the locus whose function was perturbed. Interestingly, we have found no evidence for common (pleiotropic) polymorphisms involved with variation for eye roughness and two other candidate traits known to be affected by *Egfr* function, namely wing shape (Palsson and Gibson, 2004; Dworkin *et al.*, 2005) and the spacing between the dorsal appendages on the egg shell (Goering and Gibson, 2005). Unlike photoreceptor determination, both of these traits display natural genetic variation without sensitization. Does this suggest that the genetic basis for cryptic genetic variants is somewhat different than for "normal" varying traits?

XIII. IS THE GENETIC ARCHITECTURE OF CRYPTIC GENETIC VARIATION DIFFERENT FROM THAT OF OTHER GENETIC VARIATION INVOLVED WITH TRAIT EXPRESSION?

To address this question, we have employed a reanalysis of two studies that have examined the association between molecular polymorphisms in a number of candidate genes and sternopleural bristle number in *D. melanogaster*. These studies are ideal for a number of reasons. First, environmental variables such as "stressful" temperatures (greater than 29° C) and mutations have been shown to increase the phenotypic variance of sternopleural bristle number (Beardmore, 1960; Moreno, 1994; Lyman and Mackay, 1998; Bubliy *et al.*, 2000; Indrasamy *et al.*, 2000; Dworkin, 2005b). There is substantial evidence that the increase in the variation has a genetic component (Lyman and Mackay, 1998; Robin *et al.*, 2002; Dworkin, 2003b), at least for cases where the perturbation is genetic in nature. The second useful quality for our purpose is that two of the preceding studies also examined patterns of association between molecular polymorphisms in candidate genes and bristle number (Long *et al.*, 1998; Robin *et al.*, 2002). The particular design used in these studies allows us to address specific questions about the genetic architecture with regards to cryptic genetic variation. For both of these studies, chromosomes derived from wild populations of *D. melanogaster* were extracted and placed into a common laboratory background. More importantly, the "wild" alleles were then introgressed into the common laboratory background, so that except for the gene region of interest (and flanking sequence), there was no uncontrolled genetic variation. Each of these "wild" alleles were genotyped for a number of molecular markers for the genes in question (*Delta* for Long *et al.* [1998] and *hairy* for Robin *et al.* [2002]). What makes these experiments particularly suitable to address the question of the genetic architecture of cryptic genetic variation (and whether it is different from trait architecture in any particular way) is that well-characterized mutations in the candidate genes (*Delta* and *hairy*) were also independently introgressed into the common laboratory background (*Samarkand*). The "wild" alleles (introgressed chromosomal segments) were tested by crossing to both the laboratory stock allele (*Sam)* as well as being crossed to the mutant of interest (*Delta* or *hairy*), which was also congenic with *Sam*. Thus the ideal comparison of "congenic" chromosomes differing only in the mutant or wild-type (*Sam*) allele of interest were both used to test for association of bristle number with the molecular polymorphisms. For these results, a simple linear model was employed; both sexes were analyzed separately (no effect with respect to sternopleural bristles) with the molecular polymorphism as the independent variable and line means as dependent (conducted for each polymorphism).

Figure 8-5 illustrates the results for the association test for sternopleural bristles with polymorphisms in *Delta*, for both the *Sam* chromosome and the *Delta* (*Dl*) mutation. The strength of the association is monitored by the –log(*p*) value (thus a more significant value shows a higher peak). There are two important features to point out for this figure. First, the overall shape of the profiles for both the normal (*Sam*) and sensitized (*Dl*) backgrounds are similar, with both showing the same significant peak around the marker ha_8_6. Second, the associations from the sensitization crosses are generally stronger. Although the results are not presented here, the same qualitative results are found for the *hairy* gene region. Thus this evidence suggests that the same polymorphisms are responsible for both the natural and "cryptic" genetic variation for sternopleural bristle number. It generally appears that sensitization only amplifies the effects of the variation. Therefore for discontinuous traits (such as photoreceptor number) it simply lowers the threshold for the genetic variation in the trait that is otherwise suppressed, but does not reveal a qualitatively different type of variation. These results are consistent with the general quantitative genetic models for discontinuous traits, where trait expression is affected by an underlying normal distribution of effects, but the presence of a

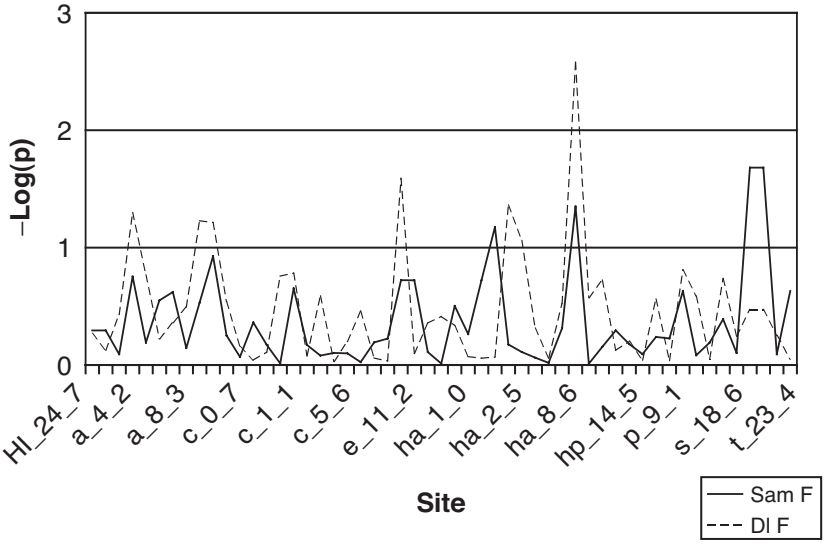

FIGURE 8-5. Association test for sternopleural bristle number with restriction site polymorphisms in the *Delta–Hairless* gene region. Y axis is the –log of the significance from the ANOVA. The strength of the association is stronger when the allele of interest is heterozygous over a mutation of *Delta* (*Dl F*) rather than over a "wild-type" *Samarkand* chromosome (that only differ genetically for the mutant allele). Although the magnitude of the effects are different, the two crosses show similar overall profiles.

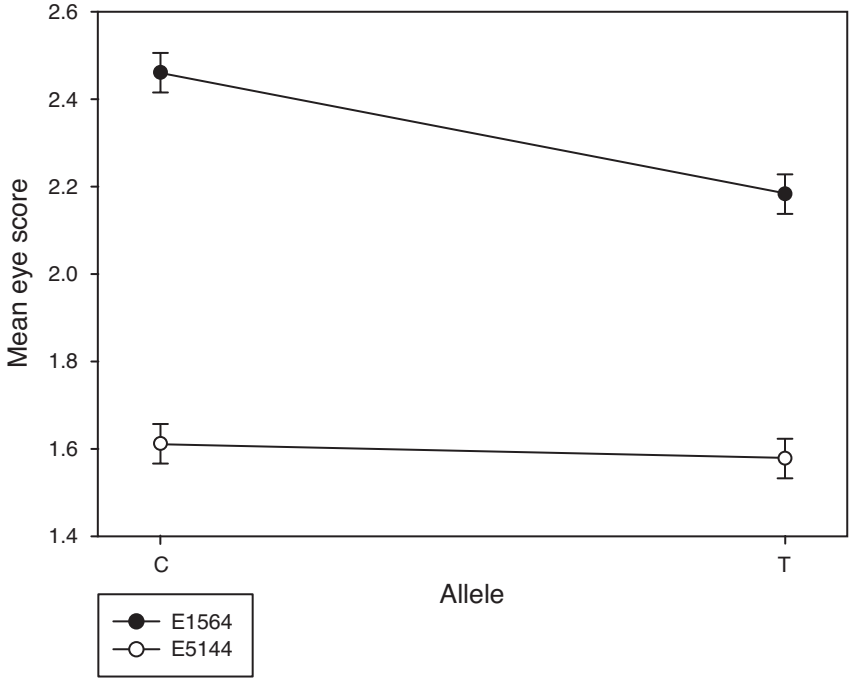

FIGURE 8-6. The effects of genetic background on single nucleotide polymorphism affects. The mean effect of substituting one allele for another depends heavily on genetic background. Not only is the mean effect on photoreceptor determination different for these two crosses, but there is evidence for a large scaling difference of the substitution of alleles.

threshold either hides the variation (for invariant traits) or sets up a situation where a quantitative gradient is interpreted digitally (dichotomous). If we return to the results for photoreceptor determination (Dworkin *et al*., 2003), we can see some results consistent with this as well. Figure 8-6 displays a reaction norm profile across two different alleles for mean eye roughness for the two different crosses employed in this study. Not only does the cross utilizing line 1564 have a greater mean trait value than 5144, but as can be seen, there is evidence for a cross-X polymorphism effect, given that the slopes of the two lines are different from one another. This also points out another important feature. Even though crosses with line 5144 are sufficient to reveal some cryptic genetic variation for eye roughness, in general it produced much lower levels of between-line (genetic) variation. Thus the breakdown of the canalization is not an all or nothing proposition, and (at least for some traits, such as photoreceptor determination) there is a degree to its effects.

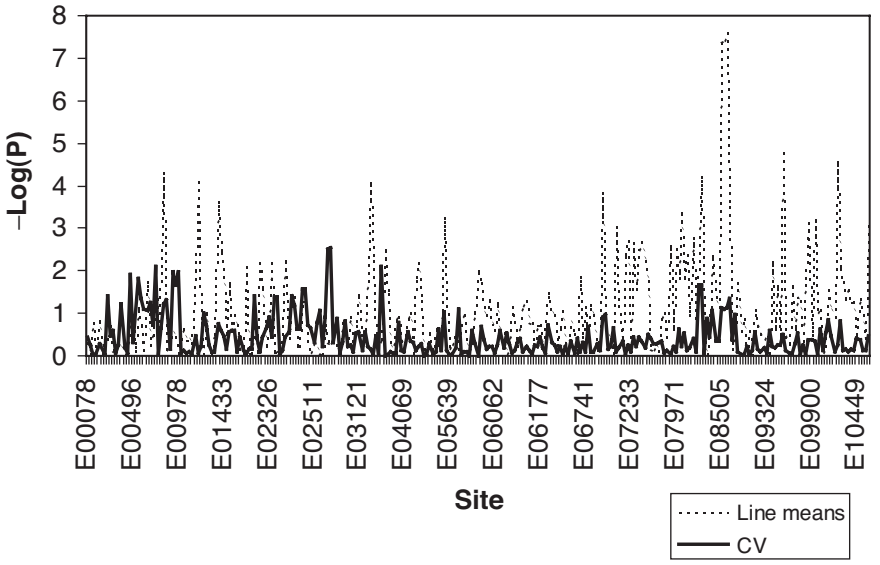

FIGURE 8-7. No evidence for a common mechanism for the two measures of canalization. Association between molecular polymorphisms in the *Egfr* gene in 210 inbred lines and line means or within-line coefficient of variation (*CV*). Not only is there no evidence for *Egfr* having an effect on the levels of within-line variation, but the profiles are quite different for each measure.

Finally, this sort of study can also be used to address some canalization-specific questions. For instance, what is the relationship between the two different measures of canalization (RxNM and variation)? If we assume that the degree of eye roughness in the crosses is inversely proportional to canalization (the greater the eye roughness score, the less canalization that line shows), then the association informs us about the RxNM approach to canalization. At the same time, we can use the coefficient of variation for each line to perform an association study for within-line variation. The results are displayed in Figure 8-7 (see Dworkin *et al.*, 2003, for more details about the primary analysis). As can be seen, the association using the RxNM metric of canalization shows several strong associations. However, the variation metric (*CV*) does not show any significant sites (after correction for multiple tests). Indeed, the profile of the associations is also quite distinct. This suggests that *Egfr* does not harbor natural genetic variation for the "variation" measure of canalization, unlike what we have observed for the RxNM approach. This is a relatively weak test of this idea. However, a recent study (Dworkin, 2005b) has explicitly addressed this question with sternopleural bristles and did not find evidence for a common genetic mechanism between these metrics of canalization.

XIV. NOW THAT I HAVE ALL OF THIS CRYPTIC GENETIC VARIATION, WHAT DO I DO WITH IT?

Considering all of the traits that have been observed to express cryptic genetic variation once sensitized, it is reasonable to ask whether or not this source of genetic variation has any evolutionary potential. To date this question has only been addressed in two empirical studies. Bubliy *et al.* (2000) investigated whether or not the increased phenotypic variation observed for sternopleural bristle number when flies are raised at high temperature (31° C) would increase the response to artificial selection as compared with 25° C controls. In this 10-generational experiment, they found no evidence for an increased response to selection. However, their lack of a "positive" result is in and of itself telling. Although the basic design of the experiment is sound, it was unlikely that the authors would have observed an effect (even if there is indeed a real effect). This is because of a conspiracy of factors that make observing a significant effect very difficult. First, as has been previously shown, the sampling variance on selection differentials is extremely large (Falconer and Mackay, 1996). For this study (Bubliy *et al.*, 2000), I estimated (from the available data) the sampling variance to be $\sim$0.19 (at least as large as the difference in phenotypic variation between treatments). In relation to this, the average increase in phenotypic variation from temperature stress is relatively small (approximately a 12% increase in *CV* under the stressful temperature in Bubliy *et al.*, 2000). Thus to observe a significant effect if there is in fact a real difference could require a large increase in sample size to see a significant effect. Although the approach of Bubliy *et al.* (2000) was the correct one (in principle), other techniques over mass selection may be required. Alternatively, using a sensitization procedure (such as a mutation) that increases the genetic variation by a larger degree may also help to increase the probability of seeing an effect (if it exists).

However, there is an alternative approach that has addressed the same question of evolutionary potential of cryptic genetic variation. Lauder and Doebley (2002) have examined patterns of genetic variation between hybrids of teosinte (the maize progenitor species) and an inbred line of maize. They investigated a number of traits that are invariant in maize, teosinte, or both (but differ from maize to teosinte) and have shown that lines of teosinte harbor considerable cryptic genetic variation for traits (invariant in teosinte) that appear to those that have been selected upon during maize evolution. Interestingly, many of the quantitative trait locus (QTL) regions responsible for teosinte–maize differences were also found to also harbor cryptic genetic variation in teosinte. To my knowledge, this is the first study demonstrating a plausible evolutionary role for cryptic genetic variation. I hope that further work will be done to narrow down the QTL to candidate loci, for linkage disequilibrium (LD) mapping. If candidate polymorphisms are found

that are responsible for some of the cryptic genetic variation, molecular population genetic analysis may help to reveal the evolutionary history of such alleles.

XV. THE FUTURE FOR STUDIES OF CANALIZATION

In this chapter, I have provided a quantitative framework for future studies of canalization. However, this does not mean that I am only advocating a traditional quantitative genetics framework for the study of canalization. In fact I suggest that a "quantitative developmental genetic" framework may be more revealing. What I suggest is that using the techniques of molecular and developmental genetics, in combination with the analytical framework of quantitative genetics, may provide clear and efficient routes to understanding canalization, especially from a mechanistic point of view. In the following text, I discuss a few studies that illustrate the potential for this approach.

Recent work has demonstrated that mutations in the *HSP90* gene reveal an extensive amount of phenotypic variation in a large number of seemingly developmentally unrelated (and otherwise invariant) traits in both *Drosophila* and *Arabidopsis* (Rutherford and Lindquist, 1998; Queltsch *et al*., 2001). In addition, these studies demonstrated that this variation could be selected upon in the same manner as Waddington (1953, 1956). Although there have been numerous critiques about inferences made from this work (Wagner *et al*., 1999, Meiklejohn and Hartl, 2002), it has provided stimuli for research into buffering mechanism in general. Perhaps one of the most important questions to come out of this work is whether or not there is evidence for "universal" mechanisms of canalization. Although this idea is contrary to previous evidence reporting little correlation between lines in their ability to buffer against the effects of different (but related) perturbations (Polaczyk *et al*., 1998), it is an interesting idea. To date the work done with the *HSP90* gene has focused on qualitative and not quantitative traits. The latter would be a welcome addition both in terms of the types of traits studied and allowing for a more rigorous analytical procedure.

Along these lines, I would suggest that one type of experiment that could prove extremely valuable would be a mutagenesis, preferably using a tagged mutagen, where the traits studied would in fact be the canalization of several traits, i.e., examining patterns of within-individual (fluctuating asymmetry) and between-individual variation, as well as across environment variation (RxNM). If this sort of design is used for a number of independent traits, then it may help to distinguish whether the effects of genes such as *HSP90* are at the extreme of a distribution of genes with pleiotropic effects or whether they are in some way qualitatively different. Perhaps more importantly, this could help identify genes involved with buffering and determine if they are in some manner independent of trait expression.

There is at least one other approach that is worth considering. By quantitatively examining the spatial distribution of proteins for determinants of early embryonic

polarity, Houchmandazdeh *et al.* (2002) observed a most interesting phenomenon. The maternally deposited *Bicoid* (*Bcd*) showed a considerable amount of phenotypic variation in its spatial distribution, specifically in the amount of protein observed at 50% of embryo length. A direct downstream target of *Bcd*, *Hunchback* (*Hb*) showed significantly less variation at this point along the embryo than *Bcd* (although a considerable amount of variation at other points along the embryo). This raises a number of questions. First, how is it that the noise in the *Bcd* signal is apparently filtered out with respect to *Hb* expression? Why is it that *Hb* only shows this effect at certain locations along the embryo? This type of study posits some suggestive ideas for the mechanism of buffering of this particular interaction, but we believe the implications for this study are much wider. First, it would be very interesting to have a handle on what sort of natural genetic variation (if any) is found for the canalization of *Hb* expression? Second, what are the phenotypic consequences of this buffering (or the lack there of)? How general a phenomenon is this? Neither this study nor the one of Rutherford and Lindquist (1998) used the type of controls as I outlined earlier in this chapter. So, for instance, it is unknown whether some of the variation observed in the *Bcd* gradient is genetic in nature or whether in fact it is all environmentally induced. Not only are these studies interesting in their own right, but within the framework of a rigorous approach for analysis as suggested in this chapter, it is likely that as a community we will be able to answer many of the outstanding questions in canalization research.

ACKNOWLEDGMENTS

I would like to thank the editors of this volume, B. Hallgrimsson and B. K. Hall, for allowing me to express my views on the study of canalization. Thanks to E. Larsen for initial discussions of many of these ideas and to A. Palsson, L. Goering, and J. Moser for comments on an earlier draft. Special thanks to Greg Gibson for extended discussion on canalization and how to measure it, which has lead to this chapter. Thanks also to R. Lyman and T. F. C. Mackay for providing data.

REFERENCES

Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using GLIM. *Applied Statistics* **36**(3), 332–339.

Alonso-Blanco, C., El-Din el-Assal, S., Coupland, G., and Koornneef, M. (1998). Analysis of natural allelic variation at flowering time loci in the *Landsberg erecta* and Cape Verde Islands ecotypes of *Arabidopsis thaliana*. *Genetics* **149**, 749–764.

Ancel, L. W., and Fontana, W. (2000). Plasticity, evolvability and modularity in RNA. *Journal of Experimental Zoology (Molecular and Developmental Evolution)* **288**, 242–283.

Bateman, K. G. (1959). The genetic assimilation of four venation phenocopies. *Journal of Genetics* **56**, 443–474.

Bateson, W. (1894). *Materials for the Study of Variation Treated with Especial Regard to Discontinuity on the Origin of Species*. Baltimore: Johns Hopkins University Press (reprinted 1992).

Beardmore, J. A. (1960). Developmental stability in constant and fluctuating temperatures. *Heredity* **14**, 411–422.

Bolivar, V. J., Cook, M. N., and Flaherty, L. (2001). Mapping of quantitative trait loci with knockout/congenic strains. *Genome Research* **11**, 1549–1552.

Bourget, D. (2000). Fluctuating asymmetry and fitness in *Drosophila melanogaster*. *Journal of Evolutionary Biology* **13**, 515–521.

Bubliy, O. A., Loeschcke, V., and Imashevae, A. G. (2000). Effect of stressful and nonstressful growth temperatures on variation of sternopleural bristle number in *Drosophila melanogaster*. *Evolution* **54**(4), 1444–1449.

Clarke, G. M. (1998). The genetic basis of developmental stability. IV. Inter- and intra-individual character variation. *Heredity* **80**, 562–567.

Darwin, C. (1859). *The Origin of Species by Means of Natural Selection*. London: Murray.

Debat, V., and David, P. (2001). Mapping phenotypes: Canalization, plasticity and developmental stability. *Trends in Ecology and Evolution* **16**(10), 555–561.

Debat, V., Alibert, P., David, P., Paradis, E., and Auffray, J. C. (2000). Independence between developmental stability and canalization in the skull of the house mouse. *Proceedings of the Royal Society London Series B* **267**, 423–430.

de Visser, J. A. G. M., Hermisson, J., Wagner, G. P., Meyers, L. A., Bagheri-Chaichian, H.,. Blanchard, J. L., Chao, L., Cheverud, J. M., Elena, S. F., Fontana, W., Gibson, G., Hansen, T. F., Krakauer, D., Lewontin, R. C., Ofria, C., Rice, S. H., von Dassow, G., Wagner, A., and Whitlock, M. C. (2003). Perspective: Evolution and detection of genetic robustness. *Evolution* **57**(9), 1959–1972.

Dworkin, I. (2005). A study of canalization and developmental stability in the sternopleural bristle system of *Drosophila melanogaster*. *Evolution*. Acceptance pending revision.

Dworkin, I. (2005). Evidence for canalization of *distal-less* function in the leg of *Drosphilia melanogaster*. *Evolution and Development* **7**(2), 89–100.

Dworkin, I., Palsson, A., Birdsall, K., and Gibson, G. (2003). Evidence that *Egfr* contributes to cryptic genetic variation for photoreceptor determination in natural populations of *Drosophila melanogaster*. *Current Biology* **13**, 1888–1893.

Dworkin, I., Palsson, A., and Gibson, G. (2005). Replication of an *Egfr*-wing shape association in a wild-caught cohort of *Drosophilia melanogaster*. *Genetics*, in press.

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. 4th ed. New York: Longman.

Feltz, C., and Miller, G. E. (1996). An asymptotic test for the equality of coefficients of variation from *k* populations. *Statistics in Medicine* **15**, 647–658.

Foulley, J. L., Hebert, D., and Quass, R. L. (1994). Inferences on homogeneity of between-family components of variance and covariance among environments in balanced cross-classified designs. *Genetics Selection Evolution* **26**, 117–136.

Gibson, G., and Hogness, D. S. (1996). Effect of polymorphism in the *Drosophila* regulatory gene *Ultrabithorax* on homeotic stability. *Science* **271**, 200–203.

Gibson, G., and van Helden, S. (1997). Is function of the *Drosophila* homeotic gene *Ultrabithorax* canalized. *Genetics* **147**, 1155–1168.

Gibson, G., and Wagner, G. (2000). Canalization in evolutionary genetics. A stabilizing theory? *Bioessays* **22**, 372–380.

Gibson, G., Wemple, M., and van Helden, S. (1999). Potential variance affecting homeotic *Ultrabithorax* and *Antennapedia* phenotypes in *Drosophila melanogaster*. *Genetics* **151**, 1081–1091.

Gottlieb, T. M., Wade, M. J., and Rutherford, S. L. (2002). Potential genetic variance and domestication of maize. *Bioessays* **24**, 685–689.

Hartl, D. L., and Clark, A. G. (1997). *Principles of Population Genetics*. Sunderland, MA: Sinauer.

Hartman, J. L., Garvik, B., and Hartwell, L. (2001). Principles for the buffering of genetic variation. *Science* **291**, 1001–1004.

Houchmandzadeh, B., Weischaus, E., and Leibler, S. (2002). Establishment of developmental precision and proportions in the early *Drosophila* embryo. *Nature* **415**, 798–802.

Indrasamy, H., Woods, R. E., McKenzie, J. A., and Batterham, P. (2000). Fluctuating asymmetry for specific bristle characters in *Notch* mutants of *Drosophila melanogaster*. *Genetica* **109**, 151–159.

Lande, R. (1977). On comparing coefficients of variation. *Systematic Zoology* **26**, 214–216.

Lauter, N., and Doeley, J. (2002). Genetic variation for phenotypically invariant traits detected in teosinte: Implications for the evolution of novel forms. *Genetics* **160**, 333–342.

Lewontin, R. C. (1966). On the measurement of relative variability. *Systematic Zoology* **15**, 141–142.

Long, A. D., Lyman, R. F., Langley, C. H., and Mackay, T. F. C. (1998). Two sites in the *Delta* gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics* **149**, 999–1017.

Lyman, R. F., and Mackay, T. F. C. (1998). Candidate quantitative trait loci and naturally occurring phenotypic variation for bristle number in *Drosophila melanogaster*: The *Delta-hairless* gene region. *Genetics* **149**, 983–998.

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

Mackay, T. F. C. (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics* **35**, 303–339.

McLaren, A. (1999). Too late for the midwife toad: Stress, variability and Hsp90. *Trends in Genetics* **15**(5), 169–171.

Meiklejohn, C. D., and Hartl, D. H. (2002). A single mode of canalization. *Trends in Ecology and Evolution* **17**(10), 468–473.

Miller, G. E. (1991). Asymptotic test statistics for coefficients of variation. *Communication in Statistics: Theory and Methods* **20**(10), 3351–3363.

Moreno, G. (1994). Genetic architecture: Genetic behaviour and character evolution. *Annual Review of Ecological Systematics* **25**, 31–44.

Nijhout, H. F., and Davidowitz, G. (2003). Developmental perspectives on phenotypic instability, canalization, and fluctuating asymmetry. In *Developmental Instability: Causes and Consequences* (M. Polak, ed.). New York: Oxford.

Palmer, A. R. (1994). Fluctuating asymmetry analysis: A primer. In *Developmental Instability: Its Origins and Evolutionary Implications* (T. A. Markow, ed.). Dordrecht, Netherlands: Kluwer.

Palmer, A. R., and Strobeck, C. (1986). Fluctuating asymmetry: Measurement, analysis, patterns. *Annual Review of Ecological Systematics* **17**, 391–421.

Palmer, A. R., and Strobeck, C. (1992). Fluctuating asymmetry as a measure of developmental stability: Implications of non-normal distributions and power of statistical tests. *Acta Zoologica Fennica* **191**, 55–70.

Palmer, A. R., and Strobeck, C. (2003). Fluctuating asymmetry analyses revisited. In *Developmental Instability: Causes and Consequences* (M. Polak, ed.). New York: Oxford.

Palsson, A., and Gibson, G. (2004). Association between nucleotide variation in *Egfr* and wing shape in *Drosophila melanogaster*. Genetics **167**(3), 1187–1198.

Polaczyk, P. J., Gasperini, R., and Gibson, G. (1998). Naturally occurring genetic variation affects *Drosophila* photoreceptor determination. *Development Genes and Evolution* **207**, 462–470.

Queltsch, C., Sangster, T. A., and Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624.

Rendel, J. M. (1959). Canalization of the scute phenotype in *Drosophila*. *Evolution* **13**, 425–439.

Robertson, A. (1959). The sampling variance of the genetic correlation coefficient. *Biometrics* **15**, 469–485.

Robin, C., Lyman, R. F., Long, A. D., Langley, C. H., and Mackay, T. F. C. (2002). Hairy: A quantitative trait locus for *Drosophila* sensory bristle number. *Genetics* **162**, 155–164.

Roff, D. A. (1997). *Evolutionary Quantitative Genetics*. New York: Chapman & Hall.

Rutherford, S. L. (2000). From genotype to phenotype: Buffering mechanisms and storage of genetic information. *Bioessays* **22**(12), 1095–1105.

Rutherford, S. L., and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342.

SanCristobal-Gaudy, M., Elsen, J. M., Bodin, L., and Chevale, C. (1998). Prediction of the response to a selection for canalization of a continuous trait in animal breeding. *Genetics Selection Evolution* **30**, 423–451.

Scharloo, W. (1991). Canalization: Genetic and developmental aspects. *Annual Review of Ecology and Systematics* **22**, 65–93.

Schultz, B. B. (1985). Levene's test for relative variation. *Systematic Zoology* **34**(4), 449–456.

Sershen, H., Hashim, A., and Vadasz, C. (2002). Strain and sex differences in repeated ethanol treatment-induced motor activity in quasi-congenic mice. *Genes, Brain, and Behavior* **1**, 156–165.

Sokal, R. R., and Braumann, C. A. (1980). Significance tests for coefficients of variation and variability profiles. *Systematic Zoology* **29**, 50–66.

Sokal, R. R., and Rohlf, F. J. (1998). *Biometry.* 3rd ed. New York: Freeman.

Sorensen, D., and Waagepetersen, R. (2003). *Normal Linear Models with Genetically Structured Residual Variance Heterogeneity: A Case Study of Litter Size in Pigs*. Aalborg, Denmark: Aalborg University.

Stearns, S. C., and Kawecki, T. J. (1994). Fitness sensitivity and the canalization of life-history traits. *Evolution* **48**(5), 1438–1450.

Stern, C. (1958). Selection for sub-threshold differences and the origin of pseudoexogenous adaptations. *American Naturalist* **92**, 313–316.

Van Valen, L. (1978). The statistics of variation. *Evolutionary Theory* **4**, 33–43.

Waddington, C. H. (1952). Selection for the basis of an acquired character. *Nature* **169**, 278.

Waddington, C. H. (1953). Genetic assimilation of an acquired character. *Evolution* **7**, 118–126.

Waddington, C. H. (1956). Genetic assimilation of the bithorax phenotype. *Evolution* **10**, 1–13.

Wagner, G. P., Booth, G., and Bagheri-Chaichian, H. (1997). A population genetic theory of canalization. *Evolution* **51**(2), 329–347.

Wagner, G. P., Chiu, C., and Hansen, T. F. (1999). Is HSP90 a regulator of evolvability. *Journal of Experimental Zoology (Molecular and Developmental Evolution)* **285**, 116–118.

Woods, R. E., Sgro, C. M., Hercus, M. J., and Hoffman, A. A. (1999). The association between fluctuating asymmetry, trait variability, trait heritability, and stress: A multiply replicated experiment on combined stresses in *Drosophila melanogaster*. *Evolution* **53**(2), 493–505.

Zar, J. H. (1999). *Biostatistical Analysis*. 4th ed. Upper Saddle River, New Jersey: Prentice Hall.