UNCOVERING CRYPTIC GENETIC VARIATION

Greg Gibson and Ian Dworkin

Abstract | Cryptic genetic variation is the dark matter of biology: it is variation that is not normally seen, but that might be an essential source of physiological and evolutionary potential. It is uncovered by environmental or genetic perturbations, and is thought to modify the penetrance of common diseases, the response of livestock and crops to artificial selection and the capacity of populations to respond to the emergence of a potentially advantageous macro-mutation. We argue in this review that cryptic genetic variation is pervasive but under-appreciated, we highlight recent progress in determining the nature and identity of genes that underlie cryptic genetic effects and we outline future research directions.

MODERN SYNTHESIS
The theoretical synthesis of evolutionary biology with genetics, which occurred in the 1940s and included the recognition of the roles of genetic drift, genic selection and speciation in micro-evolutionary change.

CANALIZATION
The tendency of traits to evolve a reduction in variability — namely, resistance to mutational or environmental perturbation. Canalization can be revealed by comparing the variance of a trait under normal and perturbed circumstances, or by measuring the robustness of a trait to new mutations in two different species.

Department of Genetics, Gardner Hall, North Carolina State University, Raleigh, North Carolina 27695–7614, USA.
Correspondence to G.G.
e-mail: ggibson@unity.ncsu.edu
doi:10.1038/nrg1426

Cryptic genetic variation (CGV) is defined as standing genetic variation that does not contribute to the normal range of phenotypes observed in a population, but that is available to modify a phenotype that arises after environmental change or the introduction of novel alleles. For example, the severity of cystic fibrosis is probably modified by genetic factors, the influence of which is seen only in the presence of a Mendelian mutation in the *CFTR* locus[1]. The basic contention that the idea of CGV poses is that the visible genetic diversity observed in populations is only a fraction of the potential diversity that standing genetic variation can produce. If visible variation is just the tip of the genetic iceberg, the question arises as to whether occasional expression of the hidden component is large enough to affect evolutionary or physiological trajectories in a meaningful manner.

The study of CGV dates back to the MODERN SYNTHESIS, when it became popular to document the high levels of variation associated with novel phenotypes that arise as a result of perturbation. At this time, terms such as CANALIZATION, phenocopy induction, genetic assimilation and homeostasis were commonplace[2] (see BOX 1 for an explanation of these terms and their relationship to CGV). Such concepts are now generally regarded as too esoteric for mainstream consideration, but CGV could nevertheless reclaim a central place in biomedical research, applied breeding and basic genetic theory as modern analytical methods make inroads into the analysis of CGV. Indeed, as one of the objectives outlined in the new strategy for the Human Genome Project is to understand the genetic basis of good health[3], as opposed to disease, CGV is increasingly relevant because it can be argued that it is the unmasking of this variation under abnormal circumstances that contributes to illness.

The traditional model organism used to analyse CGV has been *Drosophila melanogaster* (see BOX 2 for three example studies), although yeast, *Arabidopsis thaliana* and mice are also well represented in the literature. The mapping of loci that are responsible for cryptic effects is of broad interest as it sets the stage for addressing several long-standing issues in quantitative genetics. For example, what is the biochemical function of cryptic alleles? Do the same genes regulate the expression and stability of traits? And how does the GENETIC ARCHITECTURE of cryptic variation compare with that of visible variation? Fisher[4] proposed that visible quantitative variation results from the ADDITIVE EFFECTS of many loci that have small contributions and relatively minor interactions. A more modern view recognizes the contribution of QUANTITATIVE TRAIT LOCI (QTLs) of moderate effect — namely loci that explain several percent of the total phenotypic variance[5]. As geneticists begin to characterize the contribution of genetic and environmental interactions to QTL effects, we can begin to ask how cryptic factors influence physiology and evolution. It is too early

Box 1 | **Canalization and cryptic genetic variation**

The concept of cryptic genetic variation (CGV) is intimately linked with that of a related process — canalization. The distinction was not apparent when Waddington first developed his popular metaphor for canalization[62]. He saw development as the channeling of initially totipotent cells along successively narrower bifurcating canals until they arrived at their final determined fate. The depth of the canals was seen as an indicator of the strength of buffering of these fates, and genes were proposed both to determine the pathways taken by the canals (pattern formation) and their depth (the degree of canalization). He further argued that if an environmental shift caused a change in developmental trajectory, selection would favour genes that promoted the new trajectory; eventually these would become fixed and the new trait would develop even in the old environment. This idea was termed 'genetic assimilation'[63,64], but there has never been any consensus that it is a key evolutionary process.

So, in modern parlance, CGV is a phenomenon, whereas canalization is a process that has as one consequence the production of CGV. Homeostasis is simply the buffering of development, whereas canalization is the evolution of a more buffered state owing to the suppression of the expression of genetic variation. Release of this CGV can be brought about by environmental disturbance (including phenocopy induction, in which the environmental change produces a phenotype that mimics a mutation) or genetic perturbation.

Over the past 50 years, the concept of canalization has been rigorously defined as a reduction in genetic variability[8,10], and the canalization of many traits has since been examined in detail. These include the HOMEOTIC TRANSFORMATION of HALTERE into wing[27,28,65], scutellar and sternopleural bristle numbers[66] (see the figure in BOX 2), corolla number and other floral traits[67,68]. No consensus has been reached on the key evolutionary mechanisms that produce canalization. Some important work has distinguished between environmental and genetic canalization[10], predicting that canalization that is induced by changes in the environment can evolve under a broad range of circumstances, but that genetic canalization can only evolve under a more limited set of conditions. Although the review focuses on the release of cryptic variation, it is equally important to note that canalized traits have an increased capacity to absorb MUTATIONAL VARIANCE, and it is this resistance to mutations that might provide the strongest impetus for the evolution of canalization. So, perhaps paradoxically, the traits that have the most CGV might be those that present the largest mutational target for the production of visible genetic variation.

Several excellent recent reviews deal with different facets of canalization. These include an historical overview[2], discussion of concepts pertaining to buffering in general[69], definitional issues pertaining to canalization, phenotypic plasticity and developmental stability[70–72], methodology and analysis[26], the evolution of canalization[8], and the putative role of the heat shock proteins[73,74] in promoting genetic stability.

to provide any definitive answers, so in this review, we begin by outlining why this question is worth asking, then discuss how quantitative geneticists model CGV and address some of the theoretical questions that CGV raises. We conclude with a brief survey of the empirical approaches used to investigate CGV.

Why study CGV?

CGV might not be part of the standard curriculum, but there are at least four reasons why it is worth studying. First, CGV is an important determinant of how organisms respond to mutational perturbation, in the context of both disease and evolution. Many diseases — most notably cancer, but arguably also those that result from the disruption of physiological homeostasis (including diabetes, autoimmunity and addictive behaviour) — are probably threshold-dependent. The severity of the disease depends on the nature of the genetic factors that operate beyond the threshold, whether or not the genes contributed to the initial perturbation. So, the initial sensitization that leads to the disease state (such as mutation or a change in nutritional status) releases CGV[6]. Perhaps more intuitively, when a potentially beneficial new allele arises in a population, either as a result of mutation or ADMIXTURE, the likelihood that it will increase in frequency will be affected by CGV that suppresses any deleterious side effects.

Second, CGV might contribute to the marked responses to artificial selection produced by animal and plant breeders, and/or help to stabilize new phenotypes[7]. Changes in the frequency of specific combinations of alleles can shift trait means more than would be predicted by summing the effects of shifting each allele independently. Such synergistic interactions cannot be selected for directly, but their accumulation might contribute to jumps in trait values beyond the normal range. Resolution of the mechanisms that are responsible for synergistic effects is clearly important for understanding the origin of domesticated plants and animals, and for designing future breeding programmes that involve the introduction of transgenes or genotypes from wild strains into standard varieties. Similarly, the effect of introgressing transgenes into a range of varieties might be modified by CGV.

Third, the concept of CGV is intimately related to that of canalization[8] (see BOX 1), which seeks to explain the evolution of homeostasis. The following example illustrates this point. The number of whiskers on the snout of a mouse is normally 19, with some individuals having 18 or 20. However, in the presence of a mutant such as *Tabby*, the mean number of whiskers shifts to just 12, and some individuals have as few as 8 or as many as 15 or 16 whiskers[9]. This indicates that as a perturbing mutation sweeps into a population, or as a population adapts to a new environment in which expression of the trait is perturbed, the genetic system adjusts such that the variance of the trait is reduced. This canalization then hides the effect of standing variation, and makes the system more resistant to the input of new mutations than it would be in the absence of canalization[10].

GENETIC ARCHITECTURE
General features of standing quantitative genetic variation in relation to a particular trait, including parameters such as the number of loci, the degree of dominance and epistasis, and the frequency of quantitative trait locus alleles in a population.

ADDITIVE EFFECT
When two alleles have an effect on a trait, they are said to be additive if the average phenotype of heterozygotes is intermediate between those of the two classes of homozygote. Similarly, if the differences in phenotype among genotypes at two different loci are independent of one another, the effects at the two loci are said to be additive.

QUANTITATIVE TRAIT LOCUS
(QTL). One of several loci that contributes to quantitative genetic variation. QTLs are usually detected by mapping the association between anonymous genetic markers and a continuous or discrete phenotype in the progeny of a cross between two lines.

HOMEOTIC TRANSFORMATION
The conversion of one body part into another as a result of misexpression of a developmental regulatory gene. Another classic example is the antenna-to-leg transformations that result from misexpression of *Antennapedia* in *Drosophila melanogaster*.

HALTERE
In Diptera (true flies), the pair of club-like balancing organs that act as gyroscopes during flight. They are evolutionarily modified hind wings.

MUTATIONAL VARIANCE
Tendency towards dispersion of data about the mean due to new mutations that arise in each generation.

ADMIXTURE
The mixture of two genetically differentiated populations of a species, generally as a result of the migration or breakdown of a reproductive barrier, which leads to rapid changes of even common allele frequencies.

Fourth, CGV might be pervasive. Anyone working in the field has encountered the objection that CGV is of little practical relevance, and that studying modifiers and buffering mechanisms can wait until we know all of the genes that actually build traits or promote disease. However, quantitative genetics increasingly deals with threshold-dependent traits, the penetrance and expressivity of disease and GENOTYPE-BY-ENVIRONMENT INTERACTIONS[11]. To the extent that the mapping of genotype onto phenotype involves interactions among loci rather than purely additive effects[12], we will inevitably have to confront cryptic variation.

What causes CGV?

The total response of an organism to perturbation is mediated by all of the available genetic variation, including segregating rare alleles that individually have little phenotypic effect but that collectively might mediate a particularly strong response to selection. However, the following discussion of CGV is restricted to an operational definition of the phenomenon. What needs to be explained is how a genetic background that normally has little, if any, effect on a trait can significantly modify a new range of trait values immediately when shifted to a new environment or when the introduction of a new genotype by a few generations of introgression occurs. In formal quantitative genetic terms, the uncovering of CGV must involve the study of either genotype-by-environment (G×E) or genotype-by-genotype (G×G) interactions[13].

Genotype-by-environment interactions and epistasis. These two ideas are illustrated in FIG. 1. Suppose that two alleles at some locus have no differential effect on the distribution of trait values under 'wild-type' circumstances. When an environmental or genetic perturbation shifts that trait to a new mean, the two alleles now show a difference in their effect on the trait. The additive effect of

GENOTYPE-BY-ENVIRONMENT INTERACTION
In quantitative genetics, an interaction in which the degree of additivity or dominance at the locus is a function of the environment. Consequently, the difference in the mean value of each of the homozygote and heterozygote classes varies according to the environment.

EXPRESSIVITY
The degree to which a novel phenotype is aberrant. Not to be confused with penetrance, which is the proportion of individuals with a predisposing genotype that express the trait.

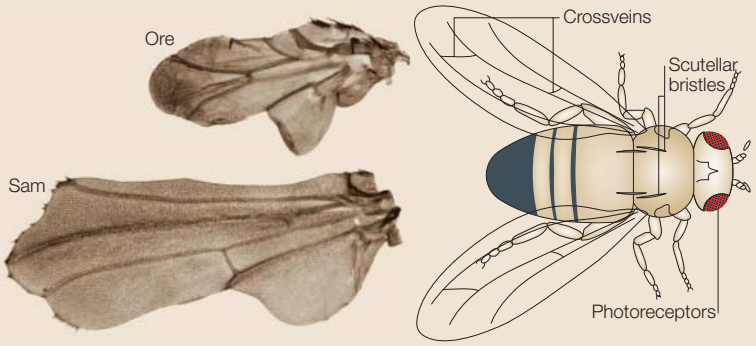
THRESHOLD-DEPENDENT
A response marked by a phase transition once the causal variable exceeds some threshold. An example is the switch in cell fate that occurs when receptor–ligand interactions exceed a threshold. In statistical terms, threshold-dependence refers to discrete trait states that are thought to arise at high or low levels of an underlying continuous variable.

SECOND-SITE MODIFIER SCREENS
Genetic screens designed to detect a mutation in a second locus that enhances or suppresses the effect of a dominant visible mutation.

INTROGRESSION
The deliberate movement of a chromosomal interval into a different genetic strain, generally by repeated backcrossing with selection for the allele in each generation.

ASSOCIATION STUDIES
An approach to identifying the genes that contribute to disease or other traits on the basis of a correlation between the genotype and some measure of the phenotype.

Box 2 | Examples of cryptic genetic variation from *Drosophila melanogaster*



Crossveinless phenotypes
A large body of work has confirmed and extended Waddington's initial observations that cryptic genetic variation (CGV) is polygenic. Bateman[75] demonstrated that there is CGV for several different wing-venation phenotypes (such as the presence of crossveins that bridge the longitudinal veins; labelled in the diagram on the right), and provided the first evidence that CGV arises from standing variation, rather than from newly arisen mutations. Milkman[76,77] showed that the expression of CGV is the result of selection for rare combinations of common alleles, rather than for individual rare alleles. Mohler[78,79] initiated genetic mapping of the crossveinless (*cv*) trait, further demonstrating the polygenic nature of the CGV, and hinting that only a few of the loci map near candidate genes. Another example of CGV in the wing is shown in the figure on the left, which illustrates the effect of the Oregon R (Ore) and Samarkand (Sam) wild-type genetic backgrounds on the EXPRESSIVITY of the scalloped (sd^{E3}) mutant phenotype.

Scutellar bristle number
Almost all the species in the Drosophilidae[80] have an invariant number of four scutellar bristles (see figure above). Nevertheless, Rendel demonstrated[30] that in the genetic background of the *scute* (*sc*) mutation, scutellar bristle number could be selected upon, providing evidence for standing CGV in the base population of *D. melanogaster*. He developed an elaborate 'probit' model for the THRESHOLD-DEPENDENCE of bristle determination[81] that assumed underlying liability distributions of gradients of regulatory molecules (in the pre-molecular era, Rendel called such gradients 'make'). Sheldon and Milton[82] demonstrated that there is an astonishing pool of CGV for scutellar bristles: after almost 60 generations of selection for increased bristle number, the mean number of scutellar bristles in selected populations was as high as 16.

Photoreceptor determination
The patterning of photoreceptors and support cells in each ommatidium (unit) of the compound eye (see figure on the right) is essentially invariant in wild-type flies, but is famously perturbed by mutations that relate to Ras-mediated signal transduction. SECOND-SITE MODIFIER SCREENS were extremely useful in investigating the biochemical pathway of photoreceptor determination[83], but it turns out that simple INTROGRESSION of a relatively weak mutation into a range of wild-type backgrounds results in even more extreme phenotypes. Polaczyk *et al.*[84] demonstrated that different alleles contribute to CGV when exposed by constitutively active mutations in two different receptor tyrosine kinases (Egfr and Sevenless) that initiate the signalling cascade. In 2003, we took the investigation of this CGV at the nucleotide level[47], as illustrated in FIG. 3, by performing an ASSOCIATION STUDY between nucleotide and phenotypic variation. This study indicates that even synonymous (non-replacement) substitutions in the *Egfr* locus can modify the phenotype of a dominant allele on the opposite chromosome.

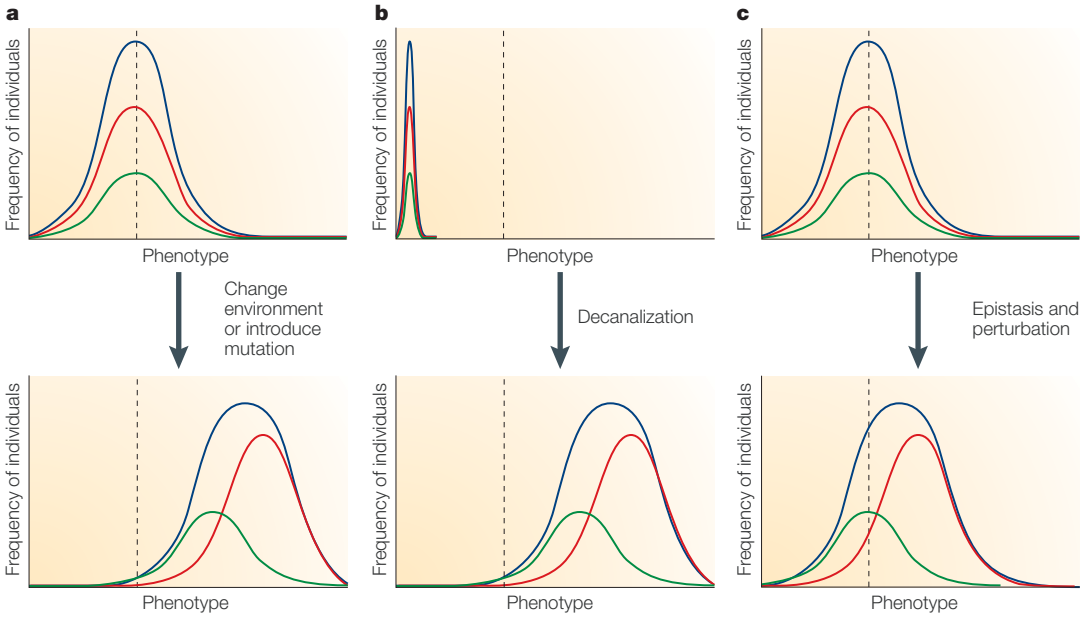


Figure 1 | **Quantitative genetic formulation of cryptic genetic variation. a** | The distribution of phenotypes observed in a population (blue curve) partly results from the sum of allelic effects at contributing loci. In this example, there is normally no difference between the contributing red and green alleles of one particular locus; however, after environmental or genetic perturbation, the more common (red) allele has a bigger effect on the trait than the minor (green) allele. This locus therefore contributes to cryptic genetic variation (CGV). The dashed line indicates the phenotype mean. **b** | Usually, but not necessarily, expression of CGV is accompanied by an increase in genetic variance, also known as decanalization, which in extreme cases can result in a normally invariant trait expressing variation in the perturbed circumstances. **c** | It is also possible that only one of the two alleles at a locus (in this case the red allele) affects the trait after perturbation, in which case the allele that contributes to CGV is also a causative allele in shifting the phenotype mean.

the perturbing locus on the trait has changed from zero to some measurable value, so that the genetic effect depends on the environment or on the genotype at the perturbing locus. It might also be the case that alleles that normally affect the trait no longer do so in the perturbed state, which would also imply a G×E or G×G interaction, but not one that results in the expression of CGV. If these two types of effect averaged over many loci offset one another, the trait variance will not change; however, expression of CGV is typically associated with an increase in the number of allelic effects and therefore trait variance. This process is referred to as decanalization (FIG. 1b), and in the extreme case, an invariant trait (such as the number of photoreceptors in a fly ommatidium; see BOX 2) becomes a variable one. Another special case of CGV is that in which just one of the two alleles affects the trait value in the perturbed state (FIG. 1c). In such cases, the locus not only modifies the new phenotype but can actually contribute to its emergence. In statistical genetics, the analysis of CGV focuses on detecting alleles that affect the trait in the perturbed but not the normal conditions, but sometimes such alleles can be revealed as ones that actually change the trait mean.

To some extent, then, the appreciation of the biological impact of CGV is reduced to establishing whether or not G×E and G×G interactions are prevalent in nature. G×G is also referred to by quantitative geneticists as EPISTASIS, which is most generally defined as the dependence of one genetic effect on another. Epistasis can be decomposed into additive and multiplicative components[13,14]. For example, if substitution of one allele for another changes the trait mean by 5%, under additive circumstances, two such substitutions should change the mean by 10%. If they actually change it by 20%, there is synergistic epistasis, or if they together change it in the opposite direction, there is antagonistic epistasis. Historically, epistasis has been assumed not to make an important contribution to standing genetic variation. This was thought to be the case for three main reasons: additive components of variation have been regarded as the principal source of response to selection[4]; epistasis is difficult to model analytically[15,16]; and epistasis is difficult to detect empirically with standard statistical methods (a recent article[17] calls for renewed attention to epistasis in quantitative genetic analyses). None of these reasons precludes the possibility that epistasis is actually prevalent, and in fact most molecular geneticists assume that genes and proteins in pathways and networks interact in complex ways — if anything, it is not clear from a biochemical standpoint why any variation should be predominately additive.

Epistasis and G×E are not sufficient to explain CGV, as these processes can, and probably most often do, modify phenotypes without influencing more marked cryptic phenotypes. The occurrence of high levels of epistasis among common segregating alleles is consistent with the presence of CGV, as any perturbation of the wild-type phenotype will change the matrix of genetic interactions and tend to modify phenotypes, and even

EPISTASIS
In quantitative genetics, an interaction between two or more genotypes, such that the degree of additivity or dominance at one locus is a function of the genotype at another locus.

SIGMOIDAL [RESPONSE]
A classic response in which the relationship between the dependent and independent variables shows a characteristic S-shaped curve, indicating a transition from slow-to-rapid response followed by a plateau.

produce new ones. Although we argue that CGV is simply the normal reaction of a genome that is subjected to abnormal circumstances, some might prefer a more stringent definition in which CGV is produced by novel genetic mechanisms. According to this view, there are at least two possible genetic mechanisms that explain CGV: threshold-dependence and the invocation of new genetic pathways (FIG. 2).

Mechanistic explanations for CGV. The concept of threshold-dependence is that some mechanisms allow the conversion of linear or other gradual inputs into sigmoidal or discrete responses. Classic threshold-dependent biochemical mechanisms include cooperative binding of transcription factors to regulatory DNA, as in the genetic switch[18] in phage lambda or *hunchback* (*hb*) induction by Bicoid in the *D. melanogaster* embryo[19], and phosphorylation cascades during signal transduction from the cell surface to the nucleus[20,21].

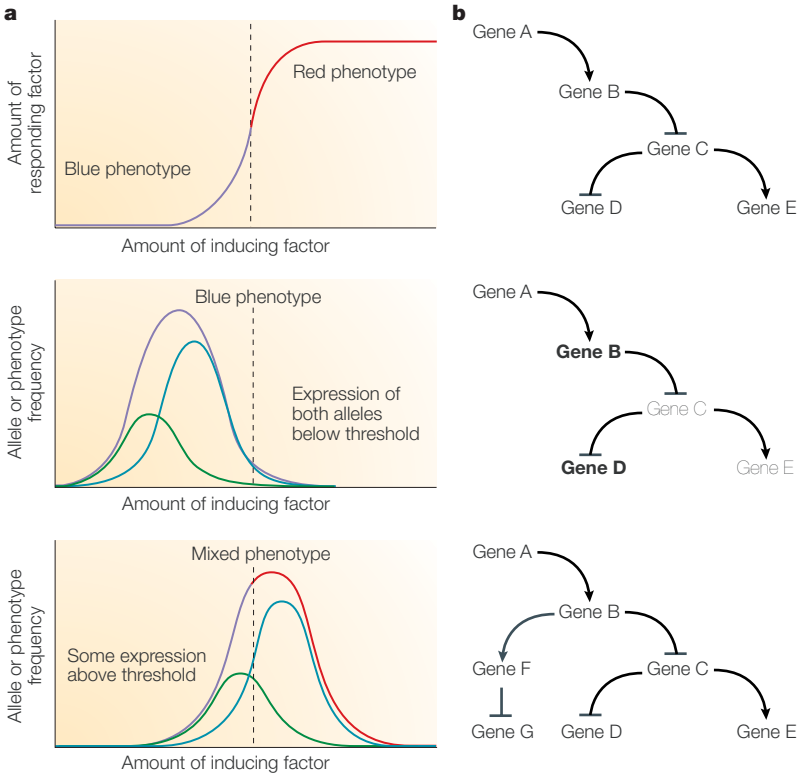


Figure 2 | **Mechanistic formulation of cryptic genetic variation. a** | Threshold-dependence. If the response of some factor such as the level of a protein or transcript shows SIGMOIDAL dependence on an inducing variable(s) (top), then an increase in the amount of the inducing factor can shift the population towards a new class of phenotype when the amount of inducing factor exceeds a threshold (dashed vertical line). As in the case shown, there need not be any difference in the relative contributions of two alleles (green and blue curves) to the production of the inducing factor before (middle) and after (bottom) perturbation to see an epistatic (threshold-dependent) response at the phenotype level. **b** | New genetic pathways. Given a genetic pathway that involves induction (arrowheads) and repression (bars), perturbation that results in overproduction, for example, of an upstream component (in this case of Gene B), can disrupt the concentration of responding factors, particularly in the presence of threshold-dependence (middle). Bold lettering represents increased expression, light lettering reduced expression. Alternatively, the perturbation might result in the induction of a new arm of the genetic pathway (bottom), such that genes that do not normally contribute to the trait now contribute to CGV.

For threshold-dependent traits, standing genetic variation is maintained as long as it does not push the trait too close to an aberrant switch point. If, for some reason, the amount of a ligand or transcription factor increases, more individuals will exceed the threshold, resulting in the activation of another gene product or protein, and leading to the emergence of novel phenotypes. Note that, in this case, the G×E or G×G interaction is at the level of the observed phenotype: there is no absolute requirement for a change in the additive effects of alleles on the production of the inducing ligand or transcription factor (that is, expression of both alleles might increase by the same amount — in fact, threshold responses are classically modelled in terms of the inheritance of a normally distributed and additive liability[13]). Although FIG. 2a illustrates the threshold response as a shift from one discrete phenotype to another, more typically the new phenotype will be a continuous one, and CGV will then modify the new distribution of trait values. For example, diabetes and asthma have variable expressivity that can be modelled as the consequence of CGV modifying a cryptic trait that is not observed in unaffected individuals, but that is induced by environmental exposure. A corollary of this view is that discrete disease states such as stroke or cardiac arrest can be investigated by identifying QTLs for underlying continuously distributed factors such as blood pressure or heartbeat intervals (arrhythmia)[22].

An alternative mechanistic explanation for CGV is the idea that the alleles that affect the distribution of novel phenotypes are unrelated to those that regulate the trait under normal circumstances (FIG. 2b). For example, susceptibility to AIDS is affected by allelic status at the CCR5 cytokine receptor: this polymorphism might not have any impact on immune function in most people, but clearly affects the genetic response to the presence of the human immunodeficiency virus (HIV) in a population[23]. Similar arguments can be made for novel phenotypes that arise as a result of a change in the environment or the introduction of a new mutation. The thrifty genes hypothesis[24] supposes that diabetes can be attributed in part to the incidence of genetic variants that arose in the human population at a time when food availability was inconsistent. These alleles have little effect on metabolism in consumers of a healthy diet, but have an important effect on disease progression in consumers of an unhealthy diet because the G×E interaction brings them into play. More generally, because so many regulatory genes have pleiotropic effects, it can be argued that any mutation that shifts a phenotype away from the optimum might cause loci that normally have no effect on the development of the trait to now contribute to its variability. The cryptic variation in this case would be normal variation for some other phenotype that is unlocked as a result of the perturbation.

Empirical analysis of CGV

Investigating polygenic factors at the nucleotide level has proved to be difficult enough for continuous traits, in which the aim is to show that there is a phenotypic

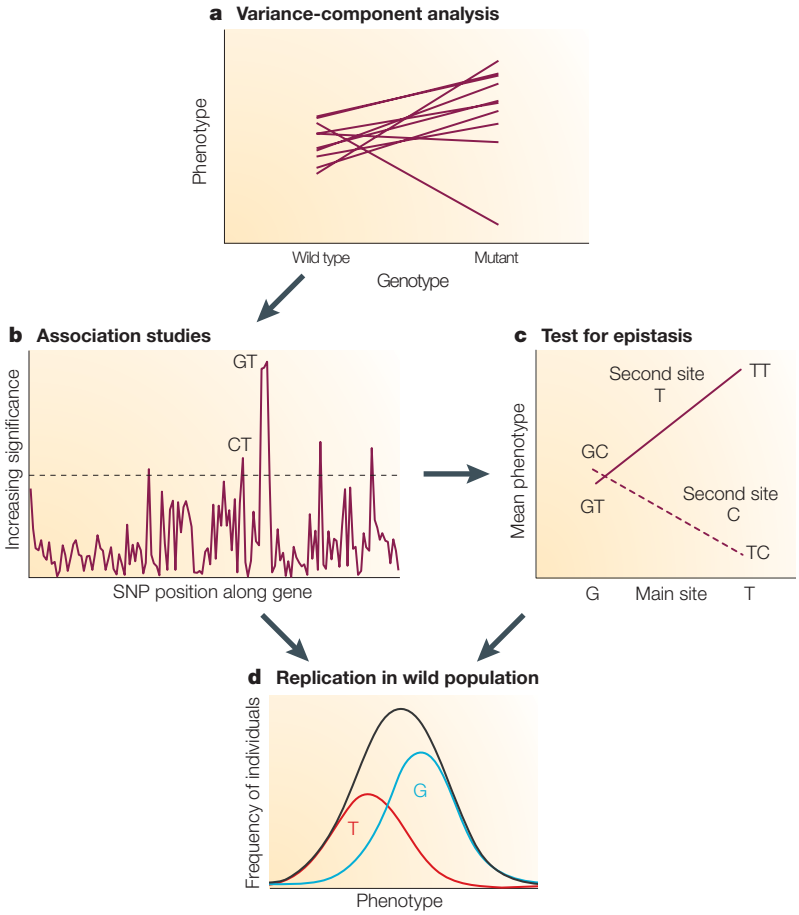


Figure 3 | **Experimental investigation of cryptic genetic variation. a** | An analysis of genetic variation usually starts with a description of the variation among lines. Cryptic genetic variation (CGV) is implicated when the phenotype of a series of lines differs between normal and perturbed circumstances; this is shown by the crossing of line means and/or an increase in variance among lines. Linkage mapping can be used to detect quantitative trait loci (QTLs; not shown), from which a candidate gene is selected. **b** | An association study involves genotyping SNPs throughout the candidate locus and performing a test of significance of the correlation between allele identity and phenotype: in the example illustrated, five SNPs exceed the experimental threshold (dashed line) for significance. From this, molecular evolutionary comparisons can be made to assess whether there is any departure from neutrality, and statistical tests of epistasis can be performed (**c**) to demonstrate whether the effect at one SNP is affected by the genotype at another SNP. In this case, the G allele at the main site is unaffected by the allele at the second site, whereas the combination of a T at both sites produces a large trait value, and a T and a C produces a small trait value. **d** | Once candidate SNPs have been identified, follow-up replication in a wild population is used to confirm the association. Wild populations can be further used to establish the allele frequencies, and to assess the fraction of the phenotypic variance that is explained by the cryptic variation as a function of the allele frequencies and the effect of the alleles on the trait. See REF. 47 for a more detailed description of an actual example of this type of analysis.

difference between two classes of allele[25]. For cryptic variation, the additional burden is to demonstrate allelic differences under perturbed circumstances that require extra genetic or environmental manipulation. The goal is to contrast the effects of each genetic factor under normal and abnormal circumstances, so as to demonstrate that they show a condition-dependent interaction that results in a substantial phenotypic shift[26]. This sounds prohibitively difficult, but, ironically, it might actually be easier to map quantitative effects at the nucleotide level (that is, to identify QUANTITATIVE TRAIT NUCLEOTIDES; QTNs) for cryptic rather than visible variation. This should be the case whenever the ratio of the additive genetic variance to the environmental variance is increased in mutant backgrounds, as will generally be the case for CGV.

The experimental investigation of cryptic variation has been pursued using a combination of genetic strategies, including mutagenesis, QTL mapping, introgression and, most recently, association studies, as summarized in FIG. 3. So far, most studies have focused on detecting an interaction between a single locus and an artificially applied perturbation. Consequently, they prove the potential for CGV, but do not yet establish its actual contribution either to phenotypic evolution or to the penetrance of disease.

Response to selection. One approach to establishing that a particular locus contributes to cryptic variation is to follow the frequency of an allele or HAPLOTYPE in response to artificial selection. This has been performed in relation to a classic case of canalization involving the homeotic transformation of haltere into wing tissue that occurs in a fraction of fruit flies after embryonic exposure to ether vapour[27]. This phenocopy mimics the effect of the *Ultrabithorax* (*Ubx*) mutants in *D. melanogaster*, and genotyping a set of SNP markers distributed across 100 kb of the *Ubx* locus was used to demonstrate that regulatory polymorphism(s) in this candidate homeotic gene must contribute to the cryptic variation[28]. A correlation between loss of UBX protein expression and phenocopy induction was also observed, which confirms the contribution of the candidate gene. Although it has not been possible to map the genetic contribution to a specific site[29], or to map other loci that might underlie the response (G.G., unpublished observations), this result established that variation that does not have any detectable effect on normal haltere development can affect the perturbed trait.

Conditional phenotypes. A systematic approach to establishing that particular loci have the potential to contribute to cryptic variation is to screen for conditional phenotypes. Much of the early literature on bristle and crossvein determination in *D. melanogaster* (see BOX 2) took this approach: for example, the penetrance of *scute* and *crossveinless* mutations were both shown to be temperature-dependent[30,31]. Influential studies by Dykhuisen and Hartl[32] established that many metabolic loci in *Escherichia coli* only show a growth phenotype under nutrient limitation, establishing the idea that alleles might be maintained in populations not so much for any function under normal conditions, but because of their cryptic contribution to survival under stressful conditions. So accepted is this idea that it now forms the basis of efforts to functionally annotate the hundreds of bacterial genes that are not normally required for viability[33]. In yeast, condition-dependence has given rise to screens for SYNTHETIC LETHAL MUTANTS, and systematic procedures for finding pairs of viable mutations that together result in lethality, known as synthetic genetic arrays, are now being used to draw interaction networks

QUANTITATIVE TRAIT NUCLEOTIDE (QTN). A nucleotide that associates with a quantitative trait, but that is detected in an outbred population or a set of unrelated inbred lines.

HAPLOTYPE
An experimentally determined profile of genetic markers that is present on a single chromosome of any given individual.

on a genome scale[34]. Second-site modifier screens are more technically demanding in multicellular organisms, but numerous examples of deleterious or advantageous interactions have been documented in flies. A famous example is *killer of prune*[35], an otherwise benign mutation that kills *prune* (*pn*) animals. Similarly, targeted gene knockouts in mice often show strain-dependence of their effects, a good example being *Egfr* mutants, which are viable in certain laboratory strains but lethal in others[36].

QTL mapping. Conditional phenotypes have also been documented for a wide range of naturally occurring genotypes. An excellent recent study from maize illustrates the practical relevance of cryptic variation. Two subspecies of teosinte, *Zea mays mexicana* and *parviglumis*, show an invariant inflorescence architecture, but segregate variants of large effect when crossed to modern maize[7]. Lauter and Doebley argue that such cryptic variation probably contributed both to the selection for novel phenotypes in the domesticated species as well as to the stabilization of it. In flies, QTLs for lifespan that have been detected in a panel of RECOMBINANT INBRED LINES are conditionally neutral in the sense that they only contribute to longevity under particular nutrient or temperature regimes[37], and similar results were reported for bristle number[38]. Harsh environments also elicit fitness differences among mutation-accumulation lines that do not differ significantly from one another under standard laboratory conditions[39]; however, the mutations in this study were not mapped.

Introgression. A promising strategy for accentuating the expression of CGV involves introgression of a mutant allele into a set of wild-type lines, contrasting the phenotypic differences, and then, if possible, mapping the factors that are responsible for the cryptic phenotypes using standard QTL-mapping procedures. The application of this strategy to the aforementioned *Ubx* mutation demonstrated that a single major-effect locus is responsible for as much as 70% of the enlargement of the haltere in one genetic background[40]. This result implies that cryptic genetic variants can indeed have large and epistatic effects. The capacity of *Hsp90* to uncover a highly pleiotropic range of phenotypes in both flies[41] and *A. thaliana*[42] has been well publicized, and Rutherford and co-workers have presented evidence that different factors are responsible for particular phenotypes such as eye and leg malformation[41]. The implication of such studies is that genomes are full of cryptic variants that are available to modify the effects of new mutations, and although some recent mathematical modelling indicates that this property should be expected of complex gene networks[43], we need to learn more about the frequencies of cryptic alleles and the magnitude and directionality of their effects.

Association studies. Association mapping (see FIG. 3) refers to the attempt to identify QTNs, and is generally performed using LINKAGE DISEQUILIBRIUM MAPPING, the genetic strategy that is now commonly used to map disease-susceptibility loci in humans[44]. The effectiveness of this approach, and details of how it is applied, depend on features of the study organism, such as the amount of polymorphism and extent of haplotype structure, the feasibility of establishing inbred lines and the presence or absence of population stratification. Flies are so polymorphic and have so little linkage disequilibrium in most regions of the genome that complete scanning of a gene requires high-density genotyping and, if possible, complete sequencing of hundreds of alleles[45,46]. Nevertheless, we have been able to associate variation in the *Egfr* locus with the severity of eye roughening caused by a gain-of-function allele of *Egfr* that increases the number of photoreceptors in each ommatidium[47]. Wild-type eyes have an invariant ommatidial structure, so this result demonstrates that cryptic variation in the *Egfr* gene affects the expressivity of the mutant phenotype of the same locus. In this experiment, 7 out of 250 common SNPs were found to be associated with the phenotype beyond the experiment-wise significance threshold, consistent with their contributing several percent to the cryptic variance. Replication of the strongest association with a synonymous nucleotide substitution was achieved using modified TRANSMISSION DISEQUILIBRIUM and case–control designs[47]. Furthermore, there was some evidence for epistatic interaction between two linked SNPs, and the frequency of the most disruptive haplotype was significantly under-represented in the population. A literal interpretation of these data is that the frequencies of the cryptic variants are affected by selection, presumably on some other pleiotropic trait, although no association between these sites and wing shape or dorsal appendage spacing has been detected (REF. 46 and L. Goering and G.G., unpublished observations). Clearly, more examples of association mapping of cryptic variants will be needed before any general claims can be made about the relationship between cryptic and visible QTNs within and among loci.

Evolutionary and biomedical significance of CGV

The issues raised by the occurrence of CGV are generally of a polemical nature, so they will probably not be resolved in any conclusive manner. Here, we outline four hypotheses that concern the evolutionary and biomedical significance of CGV, and suggest the principal avenues of research that will probably shed light on the issues. In BOX 3, we briefly mention three empirical directions that we envision for research on CGV in the next several years.

A proportion of nucleotide polymorphisms are conditionally non-neutral. Polymorphisms are classically categorized as adaptive, neutral or deleterious. There is little consensus on the relative frequencies of these classes, as well as the effects of nearly neutral and slightly deleterious polymorphisms[48,49]. The occurrence of CGV further complicates the picture as it postulates that some fraction of genetic variation that is effectively neutral in populations at equilibrium becomes adaptive or deleterious when genetically or environmentally perturbed. There is no evidence to indicate what that fraction

SYNTHETIC LETHAL MUTANTS
Pairs of mutations that are individually viable, but in combination result in lethality.

RECOMBINANT INBRED LINES
A set of lines used in linkage mapping that consists of alternating patches of the genomes of two different parental lines.

LINKAGE DISEQUILIBRIUM MAPPING
A type of association study in which a locus is inferred to contribute to a trait on the basis of a correlation between one or more SNPs that are in linkage disequilibrium with the true causal site(s), which might not actually be genotyped.

TRANSMISSION DISEQUILIBRIUM
A distortion from the expected ratio of 1:1 in the transmission of 2 alleles from a heterozygous parent to offspring of a particular class, typically affected individuals.

might be, but as little as 0.1% would probably be evolutionarily significant, and it seems improbable that it would exceed 10%. In addition, it is still unclear whether these variants that are responsible for CGV have an effect when an organism is not perturbed, and therefore, whether they might also contribute to the maintenance of variation for visible traits[50,51]. Numerous cases of detailed empirical association of SNPs with perturbed traits will be required to shed light on this hypothesis, although differences in allele frequencies across populations might also indicate the presence of CONDITIONAL SELECTION.

Adaptive walks are shaped by CGV. The most highly cited recent theory of adaptive walks, formulated by A. Orr[52,53], suggests that when a population of organisms encounters a new environment, adaptation proceeds by the stepwise fixation of new mutations. His theory indicates that the probability of fixation is related to the likelihood that the new mutation is adaptive, as well as the probability that an adaptive allele becomes fixed. The former is expected to be a function of the extent of deleterious side effects of an allele, and the latter is modelled assuming a constant selection coefficient drawn from some reasonable distribution. CGV could affect adaptive walks in at least three ways: by modifying the nature and extent of deleterious effects; by causing the selection coefficient that is associated with an adaptive polymorphism to fluctuate during a sweep; and by providing raw material for adaptation with alleles at intermediate frequencies. J. Hermisson (personal communication) has dubbed this last process 'soft sweeps' and is developing mathematical theory that describes how the footprint of this process on molecular variation is different to that of hard sweeps from new mutations[54,55]. There is ample scope for further development of population-genetic theory in this area.

An under-appreciated proportion of developmental evolution is non-adaptive. Standard evolutionary theory assumes that macro-evolutionary change results from the extrapolation of micro-evolutionary change — that is, Darwinian selection at the population level. However, we and others[56,57] have emphasized the possibility that adaptation is neither necessary to explain long-term morphological evolution, nor can it explain some of the most complex novel phenotypes. As a result of CGV, genetic variation can turn over, much as the contents of bottles of cellared wine change over time without any obvious change until they are sampled. The intuitive idea is that in complex genetic networks, compensatory substitutions occur frequently without any discernable effect on the average phenotype of the population. A similar process is well accepted as a general explanation for the evolution of HYBRID INCOMPATIBILITY[58], but it is less well appreciated that it can also contribute to morphological evolution by facilitating the emergence of novel genetic interactions. It can be argued that complex patterning pathways (such as embryonic patterning or nervous-system wiring) cannot evolve directionally by

CONDITIONAL SELECTION
Selection at the molecular level that operates only under a subset of environmental conditions or in the presence of particular modifying genotypes.

HYBRID INCOMPATIBILITY
The inviability, infertility or infirmity of the progeny of a cross between two species, generally attributed to incompatibility between allelic combinations that never occurred in either lineage as they diverged.

SOFT SELECTION
Natural selection that acts on an allele that is initially at an intermediate frequency, as opposed to hard selection, which acts on a newly arising mutation. Selection on cryptic variation will generally entail soft selection, as the perturbation exposes intermediate-frequency alleles that were previously neutral.

F_{ST}
A statistic that compares the level of genetic variation within two or more subpopulations relative to all subpopulations combined (that is, the total population).

Box 3 | **Outstanding questions in cryptic genetic variation and future research directions**

The greatest impediments to research on cryptic genetic variation (CGV) are the relative paucity of detailed studies of the phenomenon and the perceived esoteric nature of it. However, we have argued that CGV is both pervasive and central to a large range of biological problems and, furthermore, that the genetic investigation of CGV need not be any more difficult than that of normal quantitative variation. Several of the key theoretical questions are discussed in the last section of the text; here we suggest three directions for empirical inquiry.

More case studies in more organisms

Much of the literature on CGV deals with *Drosophila melanogaster*, and concentrates on half a dozen morphological traits. Recent examples that pertain to floral architecture in crosses between wild progenitors of maize[7] and work towards a molecular description of skeletal robustness in stickleback fish[85,86] point the way forward for the isolation of CGV components in organisms of agronomic and evolutionary importance.

More direct comparison of quantitative and cryptic variation

Several investigations have recently addressed the condition-dependence of quantitative trait locus (QTL) effects. These include the effect of environmental factors such as diet and temperature on longevity in *D. melanogaster*[37–39] and the effect of genetic background on the viability of knockout strains of mice[36]. Optimally, genome scans for QTLs should be done on the same crosses under different conditions, but, short of this, it is straightforward to test markers associated with QTL peaks in one cross for their effect under perturbed conditions. Synthetic lethal screens[34] will be laborious in multicellular organisms, but would provide a direct demonstration of the potential levels of CGV for different traits and organisms. The direct demonstration of the potential evolutionary impact of CGV might be approached by comparing selection responses under normal and perturbed circumstances[87].

Detection of the population-genetic signature of CGV

SOFT SELECTION acting on standing variation is not expected to leave the same molecular signature of locally reduced nucleotide variation caused by hard sweeps from new mutations. More theoretical work on the population-genetic effects of non-neutrality that depends on environmental conditions or genetic background is urgently needed[88,89]. One strong possibility is that locally elevated population structure will often reflect the operation of natural selection in populations that are exposed to different environmental factors, or will perhaps have altered genetic backgrounds owing to admixture from a nearby isolated population[90]. The demonstration of elevated F_{ST} for single nucleotides or short haplotypes will require high-throughput genotyping across broadly sampled populations.

the normal process of successive substitution of new mutations because they require multiple parallel substitutions. Both empirical and theoretical research, particularly in relation to the effect of epistasis on the relationship between genotype, phenotype, and transcript and protein abundance, will continue to contribute to the debate on the degree of coupling of macro- to micro-evolution.

Expression of CGV contributes to the high incidence of genetic disease in humans. Humans are far from being at genetic equilibrium, owing to marked changes in population size, admixture and environment in the past few thousand years. As a species, we suffer from a series of complex genetic diseases, from diabetes and autoimmune disorders to heart disease, depression and other psychological disorders. We are not aware of any systematic comparative analysis of the incidence of complex genetic diseases in wild species (although much is known about other domesticated species), but the fact that more than half of all humans suffer from one of these diseases calls for explanation. We propose that the uncovering of CGV as a result of nonspecific genetic and environmental perturbation is a key contributing factor to human disease. This argument is subtly different from theory that attributes disease to deleterious effects of alleles that previously had adaptive function at some stage of recent human evolution. Rather, the idea is that perturbation increases phenotypic variance as a result of the uncovering of hidden genetic variation that might not even affect health under normal circumstances. There is no general way to test this hypothesis, and nor is it likely to affect the methods used to identify disease-susceptibility loci; however, modelling the evolution of disease will undoubtedly benefit from understanding the reasons for the occurrence and maintenance of genetic variation.

Conclusions

CGV is a natural property of complex genetic systems, and its influence on physiology and evolution is just beginning to be explored. New genomic technologies will accelerate analyses of CGV through their impact on efforts to describe the mapping of genotype onto phenotype. The ability to map eQTLs — which are QTLs that explain the variance for gene expression that is measured with microarrays[59,60] — indicates that additive contributions to transcription are ubiquitous and surprisingly large. Sobering results from crosses between two isogenic lines of *D. melanogaster* show that F_1 progeny can be more different at the transcriptional level from their parents than are the parents from one another[61]. Quantitative measures of protein abundance are also needed, but the possibility that the inheritance of transcription and translation is extensively non-additive is a direct challenge to mathematical models that assume an orderly mapping of genotypic onto phenotypic variation. Within the next decade, it is probable that genome-wide scans for association between SNPs and multiple phenotypes will be possible in several model organisms, and together, these methods will tell us just how prevalent epistasis is, and just how great a contribution it makes to genetic responses to perturbation in ontogeny and evolution.

1. Merlo, C. A. & Boyle, M. P. Modifier genes in cystic fibrosis lung disease. *J. Lab. Clin. Med.* **141**, 237–241 (2003).
2. Scharloo, W. Canalization — genetic and developmental aspects. *Annu. Rev. Ecol. Syst.* **22**, 65–93 (1991).
3. Collins, F. S., Green, E. D., Guttmacher, A. E. & Guyer, M. S. A vision for the future of genomics research. *Nature* **422**, 835–847 (2003).
4. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
5. Mackay, T. F. The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**, 303–339 (2001).
6. Nadeau, J. H. Modifier genes in mice and humans. *Nature Rev. Genet.* **2**, 165–174 (2001).
7. Lauter, N. & Doebley, J. Genetic variation for phenotypically invariant traits detected in teosinte: implications for the evolution of novel forms. *Genetics* **160**, 333–342 (2002).
An empirical study of cryptic genetic variation in which quantitative trait locus mapping was used to identify modifiers of traits that emerge in the cross between teosinte and maize, but that are invariant in teosinte.
8. Gibson, G. & Wagner, G. Canalization in evolutionary genetics: a stabilizing theory? *Bioessays* **22**, 372–380 (2000).
9. Dun, R. B. & Fraser, A. S. Selection for an invariant character; vibrissa number in the house mouse. *Nature* **181**, 1018–1019 (1958).
10. Wagner, G. P., Booth, G. & Bagheri-Chaichian, H. A population genetic theory of canalization. *Evolution* **51**, 329–347 (1997).
Uses a mathematical model and simulations to explore the theoretical consequences of additive-by-additive epistasis for genetic and environmental canalization, and, therefore, the hiding of cryptic genetic variation.
11. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, Sunderland, Massachusetts, 1998).
12. Lewontin, R. C. *The Triple Helix: Gene, Organism, and Environment* (Harvard University Press, Cambridge, Massachusetts, 2000).
13. Falconer, D. & Mackay, T. *Introduction to Quantitative Genetics* (Longman, Essex, 1996).
14. Cheverud, J. M. & Routman, E. J. Epistasis and its contribution to genetic variance components. *Genetics* **139**, 1455–1461 (1995).
15. Falconer, D. Selection for large and small size in mice. *J. Genetics* **51**, 470–501 (1953).
16. Zhivotovsky, L. A. & Feldman, M. W. On models of quantitative genetic variability: a stabilizing selection-balance model. *Genetics* **130**, 947–955 (1992).
17. Carlborg, S. & Haley, C. S. Epistasis: too often neglected in complex trait studies? *Nature Rev. Genet.* **5**, 618–625 (2004).
Summarizes empirical approaches used to detect epistasis, and makes a strong case for more attention to epistasis in agricultural, medical and evolutionary genetics.
18. Ptashne, M. *A Genetic Switch: Phage Lambda and Higher Organisms* (Blackwell Scientific, Oxford, 1992).
19. Driever, W., Thoma, G. & Nusslein-Volhard, C. Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the bicoid morphogen. *Nature* **340**, 363–367 (1989).
20. Nijhout, H. F., Berg, A. M. & Gibson, W. T. A mechanistic study of evolvability using the mitogen-activated protein kinase cascade. *Evol. Dev.* **5**, 281–294 (2003).
21. Ferrell, J. E. How responses get more switch-like as you move down a protein kinase cascade. *Trends Biochem. Sci.* **22**, 288–289 (1997).
22. Rapp, J. P. Genetic analysis of inherited hypertension in the rat. *Physiol. Rev.* **80**, 135–172 (2000).
23. Michael, N. L. *et al.* The role of CCR5 and CCR2 polymorphisms in HIV-1 transmission and disease progression. *Nature Med.* **3**, 1160–1162 (1997).
24. Turner, R. C., Levy, J. C. & Clark, A. Complex genetics of type 2 diabetes: thrifty genes and previously neutral polymorphisms. *Quart. J. Med.* **86**, 413–417 (1993).
25. Mackay, T. F. Quantitative trait loci in *Drosophila*. *Nature Rev. Genet.* **2**, 11–20 (2001).
26. Dworkin, I. in *Variation* (eds Hallgrimsson, B. & Hall, B. K.) Ch. 14 (Academic, San Francisco, 2004).
27. Waddington, C. H. Genetic assimilation of the bithorax phenotype. *Evolution* **10**, 1–13 (1956).
28. Gibson, G. & Hogness, D. S. Effect of polymorphism in the *Drosophila* regulatory gene *Ultrabithorax* on homeotic stability. *Science* **271**, 200–203 (1996).
This reconstitution of a classic experiment involving artificial selection on phenocopy induction by ether vapour was the first to pinpoint an actual gene that is responsible for cryptic genetic variation.
29. Phinchongsakuldit, J., MacArthur, S. & Brookfield, J. F. Evolution of developmental genes: molecular microevolution of enhancer sequences at the *Ubx* locus in *Drosophila* and its impact on developmental phenotypes. *Mol. Biol. Evol.* **21**, 348–363 (2004).
30. Rendel, J. M. Canalization of the scute phenotype of *Drosophila*. *Evolution* **13**, 425–439 (1959).
31. Waddington, C. H. Genetic assimilation of an acquired character. *Evolution* **7**, 118–126 (1953).
32. Dykhuizen, D. & Hartl, D. L. Selective neutrality of 6PGD allozymes in *E. coli* and the effects of genetic background. *Genetics* **96**, 801–817 (1980).
33. Hayes, F. Transposon-based strategies for microbial functional genomics and proteomics. *Annu. Rev. Genet.* **37**, 3–29 (2003).
34. Tong, A. H. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
The first systematic, synthetic lethal screen, demonstrating the enormous potential for two-locus interactions to produce novel phenotypes.
35. Lifschytz, E. & Falk, R. A genetic analysis of the *killer-prune* (*K-pn*) locus of *Drosophila melanogaster*. *Genetics* **62**, 353–358 (1969).
36. Threadgill, D. W. *et al.* Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype. *Science* **269**, 230–234 (1995).

37. Leips, J. & Mackay, T. F. Quantitative trait loci for life span in *Drosophila melanogaster*: interactions with genetic background and larval density. *Genetics* **155**, 1773–1788 (2000).
38. Dilda, C. L. & Mackay, T. F. The genetic architecture of *Drosophila* sensory bristle number. *Genetics* **162**, 1655–1674 (2002).
39. Kondrashov, A. S. & Houle, D. Genotype-environment interactions and the estimation of the genomic mutation rate in *Drosophila melanogaster*. *Proc. R. Soc. Lond. B* **258**, 221–227 (1994).
40. Gibson, G., Wemple, M. & van Helden, S. Potential variance affecting homeotic *Ultrabithorax* and *Antennapedia* phenotypes in *Drosophila melanogaster*. *Genetics* **151**, 1081–1091 (1999).
41. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
42. Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).
43. Bergman, A. & Siegal, M. L. Evolutionary capacitance as a general feature of complex gene networks. *Nature* **424**, 549–552 (2003).
44. Zondervan, K. T. & Cardon, L. R. The complex interplay among factors that influence allelic association. *Nature Rev. Genet.* **5**, 89–100 (2004).
45. De Luca, M. *et al.* Dopa decarboxylase (*Ddc*) affects variation in *Drosophila* longevity. *Nature Genet.* **34**, 429–433 (2003).
46. Palsson, A. & Gibson, G. Association between nucleotide variation in *Egfr* and wing shape in *Drosophila melanogaster*. *Genetics* (in the press).
47. Dworkin, I., Palsson, A., Birdsall, K. & Gibson, G. Evidence that *Egfr* contributes to cryptic genetic variation for photoreceptor determination in natural populations of *Drosophila melanogaster*. *Curr. Biol.* **13**, 1888–1893 (2003).
Describes the first example of mapping cryptic genetic variation at the nucleotide level using association studies in inbred lines followed by replication in a natural population.
48. Barton, N. H. & Turelli, M. Evolutionary quantitative genetics: how little do we know? *Annu. Rev. Genet.* **23**, 337–370 (1989).
49. Gillespie, J. H. *The Causes of Molecular Evolution* (Oxford Univ. Press, New York, 1991).
50. Whitlock, M. C. Neutral additive genetic variance in a metapopulation. *Genet. Res.* **74**, 215–221 (1999).
51. Turelli, M. & Barton, N. H. Polygenic variation maintained by balancing selection, pleiotropy, sex-dependent allelic effects and *G×E* interactions. *Genetics* **166**, 1053–1079 (2004).
52. Orr, H. A. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* **52**, 935–949 (1998).
A classic theoretical analysis of adaptation that argues that any walk towards a new adaptive peak that involves new mutations should include fixation of at least one factor of large effect.
53. Orr, H. A. The distribution of fitness effects among beneficial mutations. *Genetics* **163**, 1519–1526 (2003).
54. Schlotterer, C. A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* **160**, 753–763 (2002).
55. Galtier, N., Depaulis, F. & Barton, N. H. Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics* **155**, 981–987 (2000).
56. Gibson, G. *Hox* genes and the cellared wine principle. *Curr. Biol.* **10**, R452–R455 (2000).
57. Haag, E. S. & True, J. R. From mutants to mechanisms? Assessing the candidate gene paradigm in evolutionary biology. *Evolution* **55**, 1077–1084 (2001).
58. Orr, H. A. & Turelli, M. The evolution of postzygotic isolation: accumulating Dobzhansky–Muller incompatibilities. *Evolution* **55**, 1085–1094 (2001).
59. Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003).
60. Yvert, G. *et al. Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nature Genet.* **35**, 57–64 (2003).
61. Gibson, G. *et al.* Extensive non-additivity of gene expression in *Drosophila*. *Genetics* (in the press).
62. Waddington, C. H. *The Stratgey of the Genes* (Allen and Unwin, London, 1957).
63. Waddington, C. H. Canalization of development and genetic assimilation of acquired characters. *Nature* **183**, 1654–1655 (1959).
64. Eshel, I. & Matessi, C. Canalization, genetic assimilation and preadaptation: a quantitative genetic model. *Genetics* **149**, 2119–2133 (1998).
65. Gibson, G. & van Helden, S. Is function of the *Drosophila* homeotic gene *Ultrabithorax* canalized? *Genetics* **147**, 1155–1168 (1997).
66. Moreno, G. Genetic architecture, genetic behaviour, and character evolution. *Annu. Rev. Ecol. Syst.* **25**, 31–44 (1994).
67. Huether, C. A. Exposure of natural genetic variability underlying pentamerous corolla constancy in *Linanthus androsaceus* ssp. *androsaceus*. *Genetics* **60**, 123–146 (1968).
68. Pelabon, C., Carlson, M., Hansen, T., Yoccoz, N. & Armbruster, W. Consequences of inter-population crosses on developmental stability and canalization of floral traits in *Dalechampia scandens* (Euphorbiaceae). *J. Evol. Biol.* **17**, 19–32 (2004).
69. Hartman, J. L., Garvik, B. & Hartwell, L. Principles for the buffering of genetic variation. *Science* **291**, 1001–1004 (2001).
70. de Visser, J. A. *et al.* Evolution and detection of genetic robustness. *Evolution* **57**, 1959–1972 (2003).
71. Nijhout, H. F. & Davidowitz, G. in *Developmental Instability: Causes and Consequences* Vol. 1 (ed. Polak, M.) 1–12 (Oxford Univ. Press, New York, 2003).
72. Meiklejohn, C. D. & Hartl, D. L. A single mode of canalization. *Trends Ecol. Evol.* **17**, 468–473 (2002).
73. Wagner, G. P., Chiu, C. H. & Hansen, T. F. Is Hsp90 a regulator of evolvability? *J. Exp. Zool.* **285**, 116–118 (1999).
74. Rutherford, S. L. Between genotype and phenotype: protein chaperones and evolvability. *Nature Rev. Genet.* **4**, 263–274 (2003).
75. Bateman, K. G. The genetic assimilation of four venation phenocopies. *J. Genetics* **56**, 443–474 (1959).
76. Milkman, R. D. The genetic basis of natural variation II. Analysis of a polygenic system in *Drosophila melanogaster*. *Genetics* **45**, 377–391 (1960).
77. Milkman, R. D. Genetic basis of natural variation VI. Selection of a crossveinless strain of *Drosophila* by phenocopying at high temperature. *Genetics* **51**, 87–96 (1965).
78. Mohler, J. D. Preliminary genetic analysis of *crossveinless*-like strains of *Drosophila melanogaster*. *Genetics* **51**, 641–651 (1965).
79. Mohler, J. D. Influence of some *crossveinless*-like genes on crossveinless phenocopy sensitivity in *Drosophila melanogaster*. *Genetics* **51**, 329–340 (1965).
80. Wheeler, M. R. in *Manual of Nearctic Diptera* Vol. 2 (eds, McAlpine, J. F. *et al.*) 1011–1018 (Agriculture Canada, Ottawa, 1987).
81. Rendel, J. M. *Canalization and gene control*. (Logos, New York, 1967).
82. Sheldon, B. L. & Milton, M. K. Studies on the scutellar bristles of *Drosophila melanogaster*. II. Long-term selection for high bristle number in the Oregon RC strain and correlated responses in abdominal chaetae. *Genetics* **71**, 567–595 (1972).
83. Wassarman, D. A., Therrien, M. & Rubin, G. M. The Ras signaling pathway in *Drosophila*. *Curr. Opin. Genet. Dev.* **5**, 44–50 (1995).
84. Polaczyk, P. J., Gasperini, R. & Gibson, G. Naturally occurring genetic variation affects *Drosophila* photoreceptor determination. *Dev. Genes Evol.* **207**, 462–470 (1998).
85. Shapiro, M. D. *et al.* Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).
86. Bell, M. A., Aguirre, W. E. & Buck, N. J. Twelve years of contemporary armor evolution in a threespine stickleback population. *Evolution* **58**, 814–824 (2004).
87. Bubliy, O. A., Loeschke, V. & Imashevae, A. G. Effect of stressful and non-stressful growth temperatures on variation of sternopleural bristle number in *Drosophila melanogaster*. *Evolution* **54**, 1444–1449 (2000).
88. Comeron, J. M. & Kreitman, M. Population, evolutionary and genomic consequences of interference selection. *Genetics* **161**, 389–410 (2002).
89. Innan, H. & Kim, Y. Pattern of polymorphism after strong artificial selection in a domestication event. *Proc. Natl Acad. Sci. USA* **101**, 10667–10672 (2004).
90. Rockman, M. V., Hahn, M. W., Soranzo, N., Goldstein, D. B. & Wray, G. A. Positive selection on a human-specific transcription factor binding site regulating *IL4* expression. *Curr. Biol.* **13**, 2118–2123 (2003).

Acknowledgements

We would particularly like to express our gratitude to the many colleagues who have encouraged and fostered our interest in this topic. Conversations with B. Hill and J. Hermisson helped us in thinking about some of the specific concepts, and the comments of three anonymous reviewers were enormously useful. G.G.'s research on cryptic genetic variation has been supported by the David and Lucille Packard Foundation and by the National Institutes of Health.

Competing interests statement

The authors declare no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

Entrez: http://www.ncbi.nih.gov/Entrez

CFTR | *cv* | Egfr | *hb* | *pn* | *sc* | *sd* | Sevenless | *Ubx*

OMIM: http://www.ncbi.nlm.nih.gov/Omim

arrhythmia | cystic fibrosis | diabetes

FURTHER INFORMATION

Genetic Association Database:

http://geneticassociationdb.nih.org

Greg Gibson's laboratory: http://statgen.ncsu.edu/ggibson/

Access to this links box is available online.

Author biography

Greg Gibson received his Ph.D. and postdoctoral training in *Drosophila melanogaster* developmental genetics in the laboratories of Walter Gehring at the University of Basel, Switzerland, and David Hogness at Stanford University, USA. While at Stanford University, he fell under the influence of Ward Watt and Marcus Feldman, and initiated research on the quantitative and population genetics of homeotic genes. This led him to replicate some of the old research of C. H. Waddington on canalization, using modern molecular biology tools. After starting his academic career at the University of Michigan, USA, he has spent the past 5 years at North Carolina State University, USA, in the Department of Genetics and the Bioinformatics Research Center, where he is pursuing quantitative genomic studies, mainly with the fruit fly, but more recently with dogs as well. He has also just been appointed Associate Director for Extension and Education at the new National Science Foundation (NSF) Center for Synthesis in Biological Evolution, to be located in Durham, North Carolina, USA. For more information on the author, visit http://statgen.ncsu.edu/ggibson.

Ian Dworkin received his Ph.D. from the Department of Zoology at the University of Toronto, Canada, under the guidance of Ellen Larsen, and has been a postdoctoral fellow in Greg Gibson's laboratory for the past 2 years. His research encompasses the developmental evolution of appendages in *Drosophila melanogaster*, with particular interests in geometric morphometrics, statistical genetics and their application to the study of canalization.

Online Summary

- Cryptic genetic variation (CGV) is genetic variation that is not normally expressed, but that is available to modify abnormal phenotypes produced by environmental or genetic perturbation.
- CGV is relevant to understanding the expressivity of disease phenotypes, mechanisms of animal and plant breeding, and the relationship between macro- and micro-evolution.
- CGV is thought to arise as a result of unusually large genotype-by-environment and/or genotype-by-genotype (epistatic) interactions.
- Threshold-dependent effects might help to hide CGV in natural populations.
- Some alleles that contribute to CGV might modify standing variation for different (pleiotropic) traits, but in general, the factors that help to maintain CGV are unknown.
- The same tools that are used to examine visible complex phenotypes can be used to examine CGV, including complementation testing, quantitative trait locus mapping and association studies.

Online links

Entrez

CFTR
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=Retrieve&dopt=Graphics&list_uids=1080

cv
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=Retrieve&dopt=Graphics&list_uids=44510

Egfr
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?dopt=GenPept&cmd=Retrieve&db=protein&list_uids=17136534

hb
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=Retrieve&dopt=Graphics&list_uids=41032

pn
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=Retrieve&dopt=Graphics&list_uids=31194

sc
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=Retrieve&dopt=Graphics&list_uids=30982

sd
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=Retrieve&dopt=Graphics&list_uids=32536

Sevenless
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?dopt=GenPept&cmd=Retrieve&db=protein&list_uids=24641176

Ubx
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=Retrieve&dopt=Graphics&list_uids=42034

OMIM

arrhythmia
http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=600919

cystic fibrosis
http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=219700

diabetes
http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=222100

Further information

Genetic Association Database
http://geneticassociationdb.nih.org

Greg Gibson's laboratory
http://statgen.ncsu.edu/ggibson/