

On the Association Between Genes and Complex Traits

H. F. Nijhout

Department of Biology, Duke University, Durham, NC, USA

Most diseases are complex in that they have multiple causes that interact in nonintuitive ways. Complex traits can be inherited, but they do not follow Mendel's rules of inheritance because they are influenced by many independent genetic and environmental factors, and thus exhibit incomplete penetrance and variable expressivity. Empirical approaches to understanding the causes of complex diseases typically attempt to isolate one factor at a time, and proceed to ascribe the manifestation of disease to some kind of additive or combinatorial effect of these many independent factors.

Recent developments in our understanding of regulatory mechanisms in genetics and development reveal that almost all the processes that lead to normal as well as abnormal phenotypes are nonlinear. It is the nonlinearity of interacting effects, not their diversity, that makes complex traits difficult to analyze and understand. One of the consequences of this nonlinearity is that the correlation of a complex trait with any one gene, or with any given environmental factor, varies from population to population. Moreover, when any one of these factors is altered in a given population (due to mutation or a change in environment), the correlations of other factors with the trait will also change. Thus the effects of any one of the factors that affects a complex trait cannot be understood except in the context of all other factors.

A visualization of how this context-dependency arises can be obtained from the study of relatively simple biochemical mechanisms, such as a linear enzymatic pathway (**Fig 1**). The rate of product accumulation, or flux through such a pathway, is a function of the activity of the enzymes in the pathway. The exact quantitative effect of variation in any one of the enzymes depends on the length of the pathway (Kacser and Burns, 1981). For a one-step pathway the flux is a linear function of enzyme activity, but for a multistep pathway the effect of variation of any one enzyme is nonlinear, and this nonlinearity becomes progressively more severe as the length of the pathway increases (**Fig 2**).

The consequence of this nonlinearity is not impressive if only one enzyme varies. But if there is variation of two or more of the enzymes in a multistep pathway the situation becomes quite complicated. **Figure 3** illustrates the effect of simultaneous variation in two of the enzymes in a three-enzyme pathway on the flux through the pathway. The simultaneous effect can be represented as a surface, called the phenotypic surface. This surface is simply a graph of the joint effects of two independent variables, just as **Fig 2** describes the effect of a single variable. The independent variable axes in this figure describe the range of possible activities that each enzyme can have. In a natural population there may be allelic variation in the genes that code for these enzymes

that results in slight differences in the activity of the encoded enzymes. Thus the axes of **Fig 3** can also be taken to represent a range of allelic variation. Different individuals in a population will have their genotypes along the axes, and their phenotype at the corresponding coordinates on the phenotypic surface. Points X and Y in **Fig 3** represent two individuals with different genotypes for the two alleles. These two individuals have identical phenotypes. Although they look the same, the two individuals will respond very differently to mutation in gene A. A mutation in gene A changes its allelic value, and can thus be represented by a shift along the corresponding axis. It can be seen by inspection of **Fig 3** that such a shift will have little or no effect on the phenotype (flux) of individual X, but will have a great effect on the phenotype of individual Y. The exact reverse is true for a mutation in gene B. Thus an observer of individual A would call gene A 'major gene' for the trait, and gene B a modifier gene. By contrast, an observer of individual Y would call gene B the major gene for the trait and gene A a modifier.

Thus, even though both genes contribute equally to the phenotype (as indeed they must in a linear pathway), the effect variation in one gene has on the trait is conditioned by the particular alleles that exist at the other gene. This illustrates what is meant by the 'context-dependency' of the effect of a gene. This context-dependency would not exist if the effect of genetic variation on the trait were linear and additive, because then the phenotypic surface would effectively be a flat plane. Context-dependency arises from nonlinearities of the effect of individual genes and the resulting nonlinearity of their interactions (this phenomenon is also known as epistasis and results in a curvature of the phenotypic surface).

This relatively simple complex system provides a foundation for understanding the nonintuitive effects of non-linear interactions. A short biochemical pathway is a part of almost every cellular and physiological control mechanism, and thus the properties described in **Fig 3** are likely to be extremely common. More complex mechanisms, with more independent variables, will behave in exactly the same way (Nijhout and Paulsen, 1999; Gilchrist and Nijhout, 2001), except that they cannot be easily visualized on paper because each independent variable must be represented as an axis of variation that is orthogonal to all other axes. Thus with more than two independent variables, the space of the phenotypic surface is multidimensional (Nijhout, 2002). There is nothing magical or abstract about this, if we remember that the phenotypic surface is nothing more than a graph. It is a description of the mathematical relationships among the independent and dependent variables.

Because the surface is a description of a mechanism, we can only construct such surfaces for complex traits where the underlying generative mechanism is known. This is possible for most metabolic and biochemical systems, for a great many physiological systems, and for a number of cellular and molecular mechanisms. An exploration of the properties and implications of one such set of landscapes is given in Nijhout *et al*, 2003.

Manuscript received December 30, 2002; revised May 12, 2003; accepted for publication May 12, 2003

Correspondence to: H.F. Nijhout, Duke University Dept of Biology, Durham, NC 27708, USA.

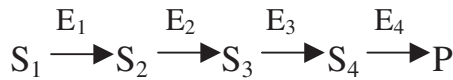


Figure 1. A simple linear enzymatic pathway, in which a substrate (S) is modified by successive enzymes (E) to produce a product (p).

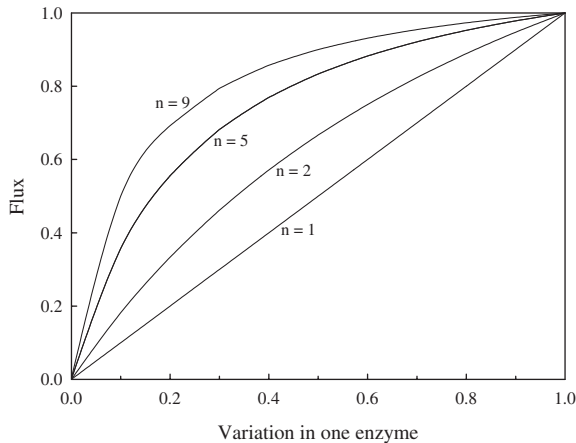


Figure 2. The rate of flux through a linear enzymatic pathway as a function of variation in the activity of one of the enzymes in the pathway. Each curve is the function for a pathway with different lengths. As the pathway becomes longer, the sensitivity to variation in any one of its enzymes becomes more nonlinear (after Kacser and Burns, 1981; Nijhout 2002).

If the mechanism is known, and the surface can be defined, it is then necessary to discover how individuals and populations are distributed on that surface. For strictly genetic mechanisms, we would need to know the range of allelic variation and the frequencies of the alleles, so that the distribution of a population on such a surface can be plotted and studied. There are various reasons for assuming that populations form rather compact clusters on a phenotypic surface (Nijhout, 2002). But it is likely that different populations are located on different regions of the surface, if for no other reason than they will have had different histories of mutation and selection for at least some of their genes.

Thus just as two individuals at different points of a phenotypic surface are differently sensitive to mutations (Fig 3), so too are populations that are dispersed on different regions of the surface. Thus if X and Y in Fig 3 represent population means (rather than individuals), then allelic variation in gene A would be more highly correlated with variation in the trait than the same amount of allelic variation in gene B, and, again, vice versa for population Y. The correlations of each gene with the trait can be deduced from the slopes of the phenotypic surface and the dispersal of the population on that surface (Nijhout, 2002). Thus, just as with the effect of mutation, the correlation of a gene with the trait is not a property of the gene itself so much as it is a property of the interactive system in which the gene is embedded. These correlations will change when a population moves to a different region of the phenotypic surface (through mutation and selection, for instance), where the slopes are different.

A particularly appealing feature of an explicit mathematical formulation of a complex trait is that it allows one to explicitly incorporate the effects of environment. If the effect of environmental variation (in temperature, vitamins, or sunlight, for example) on the mechanism that produces a trait can be modeled, then this effect can be represented as an additional independent axis of

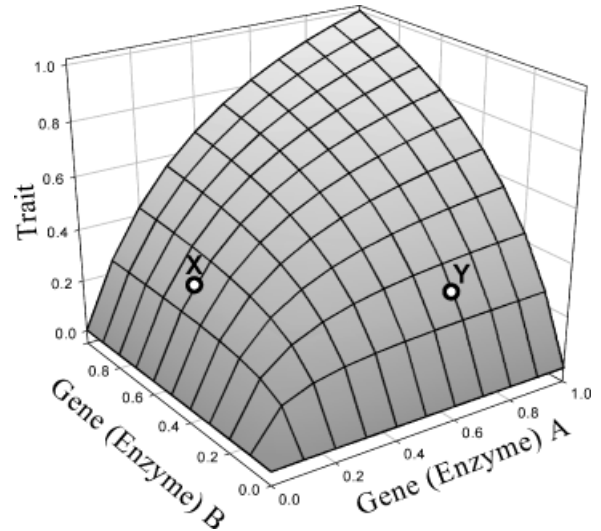


Figure 3. This graph represents the value of a trait, in this case the flux through a short three-enzyme linear biochemical pathway as a function of simultaneous variation in the activity of two of the enzymes in the pathway (A and B). The graph of this bivariate function is a curved two-dimensional surface. The points labeled X and Y represent two individuals that possess different alleles for the two enzymes represented on the x and y axes, and thus have different activities for the respective enzymes. These two individuals, however, have identical phenotypes (after Nijhout, 2002).

variation. In this way, the effects of environmental variation are put on the same footing as the effects of genetic variation. This implies that changing an environment can move a population to a different region of the phenotypic surface (again, remembering that this surface is just a graph of the mechanism), and if in this region the slopes are different, then the correlation between the other variables and the trait will change accordingly.

The nonlinearities of biochemical and physiological mechanisms ensure that the many factors that contribute to a trait are not all equally correlated with the trait. One of the consequences of this is that complex traits are typically robust to some genetic and environmental variables and highly sensitive to others (Nijhout, 2002). Moreover, if two populations occur in different regions of a landscape, they will experience different correlation patterns between the same genetic and environmental variables and the trait. This observation has obvious implications for our ability to understand and control the properties of complex traits and complex diseases. The nonlinearity of the underlying mechanisms inevitably results in a context-dependency of the association of particular genetic or environmental variables and the complex trait. The search for causes of complex diseases should take this context-dependency of causes into account.

REFERENCES

- Gilchrist MA, Nijhout HF: Non-linear developmental processes as sources of dominance. *Genetics* 159:423-432, 2001
- Kacser H, Burns JA: The molecular basis of dominance. *Genetics* 97:639-666, 1981
- Klingenberg CP, Nijhout HF: Genetics of fluctuating asymmetry: A developmental model of developmental instability. *Evolution* 53:358-375, 1999
- Nijhout HF: The nature of robustness in development. *Bioessays* 24:553-563, 2002
- Nijhout HF, Berg AM, Gibson WT: A Mechanistic Study of Evolvability Using the Mitogen Activated Protein Kinase Cascade. *Evolution and Development* 5:281-294, 2003
- Nijhout HF, Paulsen SM: Developmental models and polygenic characters. *Am Nat* 149:394-405, 1997