

# Numbering the hairs on our heads: The shared challenge and promise of phenomics

David Houle<sup>1</sup>

Department of Biological Science, Florida State University, Tallahassee, FL 32306-4295

Edited by Diddahally R. Govindaraju, Boston University School of Medicine, Boston, MA, and accepted by the Editorial Board September 21, 2009 (received for review July 22, 2009)

**Evolution and medicine share a dependence on the genotype–phenotype map. Although genotypes exist and are inherited in a discrete space convenient for many sorts of analyses, the causation of key phenomena such as natural selection and disease takes place in a continuous phenotype space whose relationship to the genotype space is only dimly grasped. Direct study of genotypes with minimal reference to phenotypes is clearly insufficient to elucidate these phenomena. Phenomics, the comprehensive study of phenotypes, is therefore essential to understanding biology. For all of the advances in knowledge that a genomic approach to biology has brought, awareness is growing that many phenotypes are highly polygenic and susceptible to genetic interactions. Prime examples are common human diseases. Phenomic thinking is starting to take hold and yield results that reveal why it is so critical. The dimensionality of phenotypic data are often extremely high, suggesting that attempts to characterize phenotypes with a few key measurements are unlikely to be completely successful. However, once phenotypic data are obtained, causation can turn out to be unexpectedly simple. Phenotypic data can be informative about the past history of selection and unexpectedly predictive of long-term evolution. Comprehensive efforts to increase the throughput and range of phenotyping are an urgent priority.**

disease | genotype–phenotype map | natural selection | G matrix | dimensionality

Medicine emphasizes proximal cause, for example, in the case of infectious disease, exposure to disease-causing microbes, environmental and genetic factors that have shaped the properties of exposure, antimicrobial therapy, and treatment of symptoms. Evolutionary biology approaches these same factors retrospectively in terms of the evolutionary history of the microbe and human host, that is the factors that have shaped the niche of microbes and humans, the evolutionary factors that allow or promote the existence of genetic variants, and prospectively the potential for the microbe to evolve in response to our therapies. The two approaches are reciprocally illuminating.

I want to point out another point of contact between medicine and evolutionary biology that is less appreciated: they both depend on our knowledge of the relationship between genotype and phenotype, the genotype–phenotype (G-P) map. This concept has a long history in evolutionary thinking. An early and influential statement of the importance of the G-P map in evolutionary biology is that of Lewontin (1), whose map is redrawn in Fig. 1A. The evolutionary process takes place in two “spaces.” The first is the genotype space (G space), which consists of all possible genotypes. Populations move in this space over generations in response to natural selection and genetic processes. Natural selection, however, takes place in continuous phenotype space (P space), the space of all possible phenotypes. The genotype of an individual strongly influences the location in P space through the process of epigenesis, the totality of interactions of genes and environment, including all aspects of development (2). The properties of the phenotype produced influence its probability of survival and success at reproducing its genotype. This process of weighting genotypes by phenotypic

success (and potentially epigenetic inheritance) then indirectly changes the mean genotype. Finally, transmission influences the mean of the next generation through the processes of segregation, mutation, and recombination. This process is repeated over many generations.

Medical genetics seeks to understand the genetic causes of variation in human morbidity and mortality. As represented in Fig. 1B, doing so involves unraveling the same transformations in and between G and P spaces as does understanding the process of evolution. Epigenesis and transmission are the same in both realms; the determination of disease state from phenotypes is precisely analogous to natural selection, although disease state may or may not influence reproductive success. Proximal causation of disease state takes place in P space and must ultimately be studied there. Current methods for explaining G-P relationships are, however, based almost entirely on determining the positions of subpopulations in G space, bypassing P space except as a classifier. For disease genetics, individuals are rather crudely sorted into diseased and healthy subpopulations so that their genetic differences can be compared. Analogous approaches are commonly used for simple continuous phenotypes, such as human height. The techniques of Mendelian analysis, candidate gene studies, and association studies are in this sense all association studies.

Thanks to genomics, we now have, or can readily obtain, abundant population data on genotypes. In addition, efforts to extend high-throughput techniques to aspects of the epigenetic process relatively close to the genome, such as gene expression, protein interactions, and metabolism, have greatly increased our ability to detect genetic influences of these subcellular phenotypes. If we consider, however, multicellular phenotypes, such as morphology, physiology, and behavior, our capabilities have remained relatively unchanged over the last 20 years. Commercially available gene chips now allow the simultaneous assay of the expression of an entire genome, but the average investigator of variation in whole organism phenotypes is not far removed from previous generations who took out the calipers, made a single measurement, and wrote it down in a notebook with a pencil. As a result, the depth of our knowledge of genomes is approaching completeness, whereas our knowledge of phenotypes remains, by comparison, minimal. Part of the explanation for this strong imbalance is certainly that P space is vastly more vast than G space.

All biologists need the problem of G-P relationships to be solved, or at least thoroughly described, but the need in evolu-

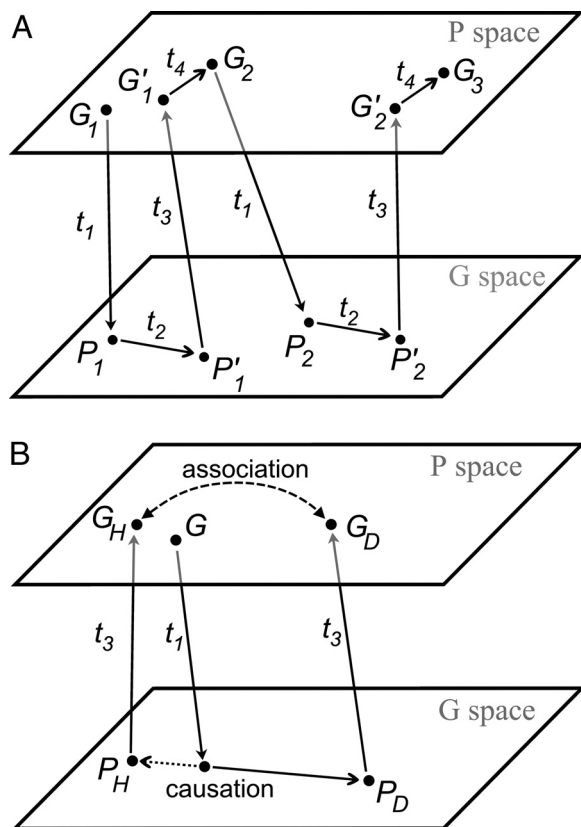
This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, “Evolution in Health and Medicine” held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at [www.nasonline.org/Sackler\\_Evolution\\_Health\\_Medicine](http://www.nasonline.org/Sackler_Evolution_Health_Medicine).

Author contributions: D.H. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission. D.R.G. is a guest editor invited by the Editorial Board.

<sup>1</sup>E-mail: [dhoule@bio.fsu.edu](mailto:dhoule@bio.fsu.edu).



**Fig. 1.** G-P maps in evolution and medicine. Circles represent population mean genotypes and phenotypes, and arrows indicate the processes by which genotypic and phenotypic means are interconverted. (A) In the evolutionary realm, epigenesis (transformation  $t_1$ ) transforms genomic information into the whole-organism phenotype. Natural selection (transformation  $t_2$ ) alters the proportions of types within the population of phenotypes, potentially changing the phenotypic mean. This process alters the frequency of genotypes by transformation  $t_3$ , the inverse of epigenesis. Finally, reproduction results in transmission ( $t_4$ ) of genotypes to the next generation, possibly again altering the mean genotype as a result of mutation and recombination. This process is repeated over many generations, moving both the population genotype and phenotype through their spaces. (B) The medical realm shares the process of epigenesis ( $t_1$ ). Any influence of the phenotypes on the likelihood that an individual will be healthy or diseased is reflected in the mean phenotype of healthy ( $P_H$ ) and diseased ( $P_D$ ) individuals, in a process precisely analogous to natural selection. This differential sorting of genotypes depending on the phenotype they produce affects the genotypic means of healthy and diseased individuals through  $t_3$ . Differences between healthy and diseased subpopulations in G space are detected in association studies. Proximal causation of disease is studied in the P space.

tionary biology is particularly acute, because no predictive science of evolutionary dynamics can emerge without such understanding. The study of natural selection is even more primitive than our knowledge of phenotypes, but only by combining a G-P map with detailed knowledge of natural selection can one predict what aspects of the genome can evolve in response. As of now, however, this effort is pinned to the type of association studies diagrammed in Fig. 1B, which rely on crude, simplistic phenotypic measures to categorize individuals, and conduct the remainder of the analysis in G space.

Does this relative lack of phenotypic information matter? There is a growing realization that phenotypically naive association studies are unlikely to explain more than a minority of genetic causation (3, 4). For some phenotypes, even driving an association project to complete description seems likely to give us a list of thousands of genes and perhaps millions of variants,

all with individually small effects. For these phenotypes we need alternative approaches to the G-P map.

Making sense of the evolutionary process requires that the phenotype as a whole be approached; understanding the causation of disease in human does as well. The large-scale study of high-dimensional phenotypes is phenomics; phenomics is the natural and inevitable complement to genomics. Implementation of a phenomic approach faces two critical challenges. One is obtaining comprehensive phenotypic data, and the second is learning how to use such data. The title of this article includes a quotation from Luke 12:7, where God is ascribed the power to evaluate the tiniest details of existence, not only to number the hairs of our heads, but to understand their meaning. Can we hope to do as well?

### Studies in G Space

The thrust of modern biology and much of medicine is that the most effective way to understand phenotypes, including disease state and mortality, is to understand how genes function. The assumption is that failures in gene function either directly cause failures of organismal function or mimic the effects of failures with other causes. For example, epidemiologists have explicitly turned to the concept of “Mendelian randomization” (5) to test hypotheses about even environmental causes of disease. This approach exploits genetic variants that manipulate the effective factor hypothesized to be responsible for altered risk. Because relatives that share exposure to environmental factors may nevertheless differ in their relevant genotypes, we may get a less confounded picture of causation. Furthermore, we know that virtually all common diseases show evidence of inheritance. As the human genome was being published, optimism that these genetic and genomic approaches would unravel the majority of disease causation ran high (e.g., ref. 6).

This G-space approach to disease has undoubtedly had great successes. As of June 2009, The Online Mendelian Inheritance in Man database (7) listed 2,908 disorders that can be traced to defects at particular human genes. More than 3,700 additional disorders show evidence for inheritance, although they have not yet been traced to a precise genomic location, so these numbers are very likely to continue to increase. If progress along these lines can continue, then maybe we do not need to worry about our relative inability to measure phenotypes.

The available methods for detecting genetic association have important limitations that give reason to doubt how far we can go in understanding phenotypic causation (8). For example, traditional Mendelian analyses of disease state have been supercharged by the availability of abundant markers, but require that the overall probability of disease in the random population be low, a set of candidate loci be in hand, and both the penetrance (probability of developing the disease state when the variant is present) and effect size (detectability of the disease phenotype) of an allele be very high. When these assumptions are met, first-degree relatives, such as siblings, of those with the disease are at a high relative risk of disease ( $\lambda$ ), and therefore have a high odds ratio (OR), a ratio of the probability of being affected when an allele is present to the probability of being affected when it is absent, which is typically approximately twice  $\lambda$ . Such analyses, therefore, only effectively uncover causation of rare syndromes that can be caused by single genes. Typical relative risk values for mapped Mendelian variants are well over 10. Some examples of Mendelian genetic diseases are cystic fibrosis ( $\lambda = 500$ ), phenylketonuria ( $\lambda = 500$ ), and sickle cell anemia ( $\lambda = 18$ ) (7).

The relative risks of common, chronic, and late-onset diseases are typically much lower. The two leading causes of death in the United States are heart disease and cancer (9), but the average  $\lambda$  for 28 types of cancer was just 2.2 (10), whereas a typical number for heart disease is 3 (for the presence of coronary artery

disease in sibs of heart attack patients; ref. 11). Finding causation in such cases requires candidate gene or genomewide association (GWA) studies. GWA studies (e.g., ref. 12) are powerful when a causal allele is common enough to be present in multiple individuals in a sample and penetrance is as low as 10%, enabling alleles with small ORs of 1.1 or so to be detected, if sample size is very large. Common alleles, however, almost never have large effects (detected alleles usually have ORs between 1.2 and 1.5; ref. 8), and therefore explain little of the variance in disease. The candidate gene or “rare variant” approach intensively screens for variants in samples with and without the disease. Enrichment of rare variants in the diseased population, followed by verification that the function of the candidate gene is altered, indicates that altered function increases disease. Typical rare variants have ORs between 2 and 10, but individually explain little of the variance in disease susceptibility because of their rarity. The discovery that de novo copy-number variants are commonly associated with disease in a wide variety of syndromes (see, e.g., refs. 13 and 14) provides hope that these readily detected mutations will lead us to large numbers of new candidate genes. The GWA and candidate gene approaches are now being brought together by intensive resequencing of case-control populations (see, e.g., ref. 15).

This combination of available techniques is thus incapable of detecting variants with low penetrance or those at loci that are not yet candidates. These classes of variation appear to be quite common as only a small proportion of the total genetic causation can so far be assigned to a genomic location in most syndromes (3, 4), a result that suggests that low-penetrance alleles explain a substantial proportion of disease susceptibility. One response to this problem is to improve our methods, by explicitly addressing the mechanisms of low penetrance, such as genotype–environment interaction and epistasis (16).

Evidence is strong, however, that these holes in our understanding signal a deeper problem that cannot be fully addressed by better association studies. A particularly revealing set of GWA studies recently discovered multiple new regions with effects on height in human populations (17–19). The total sample size of genotyped individuals over those three studies was  $\approx 85,000$ ; each study accumulated large samples by combining data from many different GWA studies that incidentally recorded height. The studies collectively identified 52 loci that affect height, 40 of which were previously unknown. On one level, this result seems to confirm the usefulness of GWA studies (20), and each article proudly points out clustering of the associations near loci known to influence bone growth. The more important message, however, is the proportion of variation in height explained; the studies explain just 2.9%, 2.0%, and 3.7%, respectively of the variation in human height in populations of European ancestry. These figures suggest the distinct possibility that something approaching the entire genome is capable of influencing height (4), a conclusion supported by the finding that one-third of nonlethal mouse gene knockouts affect body weight (21).

Given these results, the goal of understanding GP relationships can probably advance only partway by association mapping. Many voices of caution have argued for a scaling back of the genomic rhetoric to match diminished expectations (8, 16). Others are ready to counsel that the entire enterprise of association mapping should be abandoned (e.g., refs. 3 and 22). To those focused on the overriding G–P problem the first response is eminently sensible, but unsatisfying, because it leaves no prospect of a solution. The second is defensible only if an alternative is available.

### A Phenomic Alternative?

Fig. 1 suggests a natural alternative to G space studies that incorporates some concepts from quantitative genetics and

evolutionary biology. These concepts are relatively straightforward if a single trait is involved, whose value can be symbolized  $p$ , with population mean  $P$ . First consider the study of natural selection during evolution, represented in Fig. 1A. Fitness is a function of trait value,  $f(p)$ . When this function is standardized to a value of 1 at the population mean [i.e.,  $f(P_1) = 1$ ] the derivative at the population mean is the selection gradient,  $\beta$ . The gradient is approximated by the ratio between the covariance between  $f$  and  $p$  and the population variance in  $p$ ,  $\beta = COV_{f,p}/V_P$  (23), and gives the rate at which relative fitness changes for a unit change in the trait. It is readily estimated as the regression of relative fitness on trait value. The gradient allows the calculation of the change in mean phenotype after one round of natural selection ( $P_1$  and  $P_1'$  in Fig. 1A) as  $P_1' - P_1 = V_P\beta = COV_{f,p}$ .

A very similar approach could be used to examine the function expressing the change in disease (or health) probability as a function of phenotype  $f(p)$ . In this case, the means of healthy and diseased individuals ( $P_H$  and  $P_D$ ; Fig. 1B) are readily calculated, and the gradients that transform the population mean to either  $P_H$  or  $P_D$  can be obtained as e.g.,  $\beta_H = (P_H - P)/V_P$ . This procedure is equivalent to using an indicator of disease state as an analog for fitness. For example, to obtain the health gradient  $\beta_H$  we could use an indicator,  $x$ , that has a value of 1 for a healthy individual and 0 in a diseased one. If the proportion of healthy individuals is  $h$ , regressing  $x/h$  on  $p$  gives the disease gradient that expresses the change in relative probability of health for a unit change in trait value. When a large change in disease probability occurs in the range of the data, logistic regression will provide a better estimate of  $\beta_H$  (24).

This thinking is most useful when generalized to multiple traits (25), that is to P space, where the phenotype is a vector  $\mathbf{p}$ . The weighting function,  $f(\mathbf{p})$  (which can be either a fitness function or a disease function) is a multidimensional surface. The derivative of this function is a gradient vector  $\boldsymbol{\beta}$  with elements that summarize the direction in which  $f(\mathbf{p})$  fitness or disease state probability increases the most rapidly. Quadratic (or even higher-order) terms can also be fit, capturing the curvature of  $f(\mathbf{p})$  around the population mean. The resulting matrix of quadratic terms is called  $\boldsymbol{\gamma}$  (25).

Estimation of  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}$  is by multiple regression and has the same advantages in this context as any other: if the function is well-behaved and the traits that actually cause the dependent variable to vary are in the analysis, the elements of  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}$  will reveal the relative importance of each phenotype in determining the outcome. This could, for example, reveal which of a large number of possible phenotypes are most predictive of disease. If, however, some or all of the causal traits are missing from the analysis, because they are almost certain to be when only one trait is analyzed, the estimated  $\boldsymbol{\beta}$  can underestimate or overestimate the importance of each trait, perhaps giving a misleading picture of which phenotypes matter (25, 26). Neither medical researchers nor evolutionary biologists currently have access to anything approaching complete phenotypic data, so full use of this approach awaits widespread implementation of phenomic-scale measurements.

In the evolutionary realm, the results of natural selection are transmitted back to the genotypes passed on to the next generation. The process of epigenesis ( $t_1$  in Fig. 1) is to some degree indeterminate so individuals with the same genotype produce a variety of phenotypes. This variation itself may be partly predictable from the study of interactions between genotype and environment. As a result, some of the change in phenotype brought about by applying function  $f(\mathbf{p})$  is caused by the deviations of an individual from the average phenotype that its genotype would produce, and only some of the weighting function is transmitted back to G space in transformation  $t_3$ . The amount that is transmitted for a single trait is captured by the

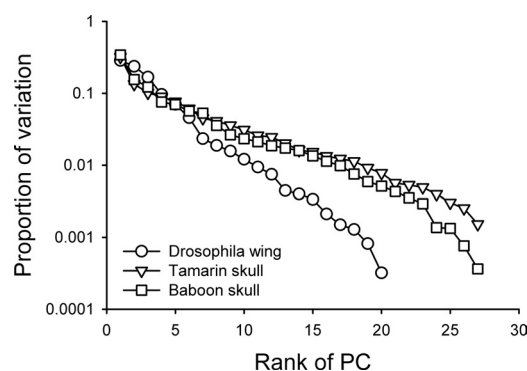
additive genetic variance,  $V_A$ , which is that part of  $V_P$  that causes offspring to resemble their parents. The expected change in mean phenotype in a single trait between generations (neglecting the transformations caused by mutation and recombination, which are usually small) is then  $P_2 - P_1 = V_A\beta$ , which can be rearranged to give the more familiar form  $P_2 - P_1 = V_A/V_P \cdot COV_{f,P} = h^2S$ . In the multivariate case, inheritance is captured by a matrix of variances and covariances,  $\mathbf{G}$ , where the diagonal elements are the additive genetic variances for each trait, and the off-diagonal elements are the additive genetic covariances between the traits. The resulting transformation across generations is then  $\mathbf{P}_2 - \mathbf{P}_1 = \mathbf{G}\beta$ .

### Steps Toward Phenomics

Assessment of variation at a few locations in the genome was not enough to characterize location in  $\mathbf{G}$  space, so we have turned to genomics. Similarly, the logic of natural selection and disease causation in  $\mathbf{P}$  space makes clear that studying a few traits cannot be enough. In the last 10 years, calls for enhanced phenotyping have become increasingly common, although the logic behind these arguments has been varied and not always explicit (27–33). These calls have increasingly been taken up and led to concrete increases in our phenotyping ability (e.g., refs. 32 and 34–38) of differing scale and complexity.

Clearly, phenomic measurements must be extensive, covering many different aspects of the phenotype, such as morphology, behavior, physiology, etc. Less obviously, phenomics must also be intensive; that is, it must lead to detailed characterization of each major aspect of the phenotype. For example, the genetic variants that affect function of the human heart are very likely to have pleiotropic effects on other body parts and functions, calling for extensive measurements of other systems. In addition, the heart itself cannot be adequately characterized by a small number of summary parameters like cardiac output, but must be approached in terms of the full complexities of physiological capacities, morphology, etc., calling for intensive measurements of the heart. Phenomic efforts are rising to both challenges. For example, the mouse research community is adopting a standard set of protocols for extensive measurement covering many different aspects of the phenotype (37, 38), and intensive measurements of mouse morphology are being pursued by other groups (e.g., ref. 39). Most important, the mouse community is focused on associating this detailed phenomic data with particular genotypes and their recombinants.

An easy objection to putting resources into phenomics is that most of what we might measure may prove irrelevant. Although the genotype has a finite extent and discrete content and can therefore be measured exhaustively, the phenotype is both continuous in multiple dimensions and infinitely divisible in some dimensions. For example the state of the phenotype can be measured at an infinitely great number of time points. If the goal is exhaustive measurement of the phenotype, it will forever remain beyond our reach. Rather, the goal must be defined in terms of understanding. How intensively we need to measure the phenotype to achieve goals like understanding the proximate causation of natural selection or disease is an open question that must be addressed with respect to a particular goal, such as predicting susceptibility to a particular disease, or response to a particular selection pressure. Both the genotype and especially the phenotype are immensely complex; our hope must be that any particular problem becomes simpler when viewed from a favorable perspective. Buchanan et al. (22) nicely summarized this hope with the metaphor of an hourglass with the full genotype at one end and the full description of the phenotype at the other. In between, we hope, is the waist of the hour glass, where measurement of just a few key aspects of the organism (which could be any combination of genetic, environmental, and phenotypic measurements) are maximally informative about the



**Fig. 2.** Proportion of additive genetic variation on each principal component (PC) axis of the  $\mathbf{G}$  matrix for three sets of morphological data: 30 measurements of baboon skulls (ref. 57 and C. Roseman, personal communication), 39 skull measurements of a combined estimate from two tamarin species, *Saguinus fuscicollis* and *Saguinus oedipus* (refs. 58 and 59 and J. M. Cheverud, personal communication), and 12 *Drosophila melanogaster* wing vein intersections (42).

problem we want to address, for example, fitness or disease susceptibility. This task is to increase the range of data that can be applied to a problem in hopes that, once the key pieces are in hand, we can build simplified, but powerful, models of causation.

### Early Lessons from Phenomic Data

Evolutionary biologists have been increasingly dealing with intensive datasets where a single aspect of the phenotype is subjected to detailed characterization, and the outcomes of these studies can give a taste of what we might learn by expanding such studies to more be more comprehensive. I discuss four areas where we can draw tentative conclusions not accessible from a genomic viewpoint alone.

**Phenotypic Datasets Have High Dimensionality.** Perhaps the largest class of highly multivariate datasets are obtained from studies of morphological form, assessed by means of the spatial locations of landmark points that are recognizable across a series of specimens or the curves that connect such points. In a relative handful of instances, such data have been subjected to a genetic analysis, in which the variation was partitioned into additive genetic variance and all other sources of variation. This process yields a  $\mathbf{G}$  matrix with as many rows as the number of measurements, minus a few degrees of freedom for estimating the spatial orientation of each specimen relative to the others (40). When each  $\mathbf{G}$  matrix is subjected to a principal components analysis (PCA), we can see just how much variation will be missed in studies that only measure a handful of traits. PCA rotates the  $\mathbf{G}$  matrix to a new set of directions (the eigenvectors). Each direction has an amount of variation associated with it, its eigenvalue. The eigenvectors are chosen to maximize the range of the eigenvalues, so PCA allows us to measure the amount of variation in the least variable combinations of the original traits.

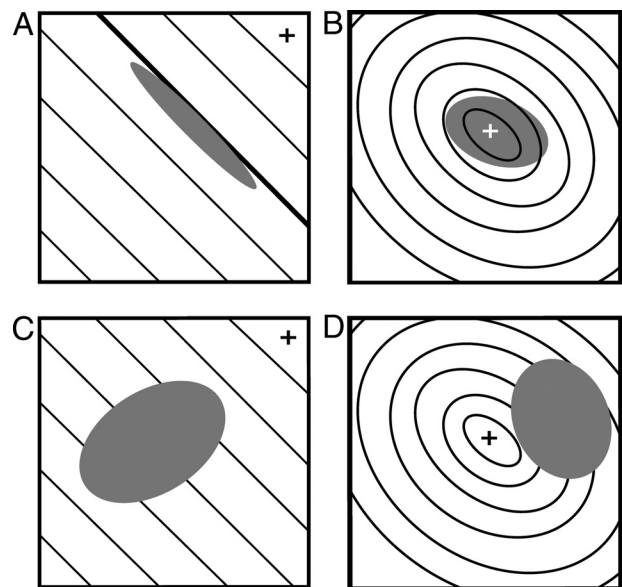
Three eigenvalue distributions are shown in Fig. 2, for measurements of baboon and tamarin and for *Drosophila* wing vein intersections. The analyses of these data remove size, leaving only variation in shape (41). On a log scale, the decrease in the amount of genetic variation in shape explained is approximately linear, suggesting an exponential distribution. What is most remarkable is how slowly the amount of variation falls from the  $k$ th most variable direction to the  $k + 1$ th direction. For the tamarin and baboon skulls, the variation falls by an average of 18% with each dimension, whereas for the flies it falls at the rate of 29% per dimension. No single summary measure can capture even 50% of the variation in the shape of these structures.

A second question is just how many aspects of form must be measured to characterize the genetic variation fully, that is, what is the dimensionality of the genetic variation. The fly-wing study used a particularly large number of families (800) and individuals (17,000), and demonstrated that at least 17 of the 20 possible dimensions had significant genetic variation (42). Further analysis using a restricted maximum-likelihood approach (43) revealed significant variation in all 20 directions. No similar analyses of the primate skull datasets have been done, and the sample size in each of those studies was substantially smaller. Nevertheless, the overall pattern of decrease in variation is quite similar and suggests that the genetic dimensionality in each of these species is also quite large. Full characterization of the genetics of these phenotypes cannot be undertaken from a small sample of measurements.

**Studies of Selection Can Reveal That Only Some Combinations of Traits Are Important.** Although the dimensionality of genetic variation is high, the selection-gradient analysis described above may well turn up some low-dimensional combination of traits that predicts an important outcome, either fitness or disease. Such a result could potentially indicate the narrow waist of a causal hourglass (22). The ability to attract a mate is an important component of fitness, and attractiveness can be readily assayed by directly observing matings, allowing selection gradient analyses to be performed. Blows and colleagues (44, 45) characterized the relative abundances of nine cuticular hydrocarbons (CHCs) in a population of *Drosophila serrata*, then compared the compositions of males to their success in competitive mating trials. The standardized selection gradient was extremely strong (change in relative fitness of 76% > 1 SD change in the relative proportions of different CHCs), suggesting female preferences for higher proportions of several CHCs and antipathies toward others.

This logic of simplification extends to sets of phenotypes near a fitness optimum, where fitness decreases away from the population mean. In such cases, the important parameters are the quadratic coefficients in the  $\gamma$  matrix. These can be manipulated to allow interpretation, even in very high dimensional space (46). Brooks et al. (47) studied sexual selection on the calls of a cricket by synthesizing variation in five aspects of the call so that they could assess attractiveness of phenotypes not actually found in nature. They found that mate choice favored an intermediate optimum phenotype and that females paid strong attention to just two of the possible directions in P space. These cases suggest that a phenomic approach that begins with extensive and intensive measurements can then turn around to indicate some low-dimensional subset of these that is actually important in a particular context. The advantage of passing through a phenomic phase is that which combinations of traits are actually important is not apparent at the start.

**Studies of Selection Can Suggest Past History of Trait Evolution.** Both of the sexual-selection studies cited above went further and compared the pattern of selection on phenotypes to the pattern of genetic variation in the population studied. In each case, those aspects of the phenotype that were most strongly selected also had little genetic variation (45, 48). This pattern suggests a persistent mismatch between the phenotypes that females prefer and the ability of the males to produce them. The result is that the genetic consequences of female choice are very small; successful males are those that happen differ from normal in the favored direction, perhaps because they have experienced a favorable environment. Fig. 3A represents this relationship between population variation and selection schematically. The gray ellipse represents the expected phenotype (averaged over environmental factors) of the genotypes in the population. There



**Fig. 3.** Relationship between genetic variation in P space and the probability of disease or of fitness. Gray ellipses represent the distribution of genotypic variation in phenotype. + indicates the phenotype with the lowest probability of disease (or highest fitness), and the black lines are iso-lines that mark a particular level of probability of disease. (A) Constraints on the possible genotypes. The heavy diagonal line represents the constraint. A genotype above and to the right of this line cannot evolve. (B) Population mean is near the optimum, but mutation creates variation around that optimum. (C) A population in a novel environment, evolving toward a new optimum. (D) An aging population, where deterministic changes in phenotype caused by senescence drive the population away from the optimum.

is plenty of variation in the genetic basis of the phenotype, but it is oriented orthogonal to the direction of selection.

This finding suggests that the comparison of the pattern of genetic variation for phenotype with the probability of disease could be very informative about the nature of the genetic variation in human disease. If the contours in Fig. 3 are now taken to represent disease probabilities instead of fitnesses (with + indicating a healthy phenotype with low disease probability), several possible scenarios might be found. Fig. 3A would represent an outcome with low  $\lambda$ , such as our inevitable demise caused by aging. Mutation-selection balance might produce a distribution like that in Fig. 3B, where the population generally matches the optimum state, but individuals with extreme phenotypes have increased probability of disease. Note that this disease may not be the same in each direction. The emerging hypothesis that many psychiatric disorders represent overexpression or under-expression of continuous personality traits provides a possible example, in which deviation in one direction leads to autism and deviation in another leads to schizophrenia (49). Diseases of civilization might lead to a pattern like that in Fig. 3C, where there is ample variation that has not yet been removed by a long history of natural selection. In the environment of evolutionary adaptedness, the selective pattern on the same variation might have been like that in Fig. 3A or B. The shift from those patterns to that shown in Fig. 3C would be caused by genotype-environment interactions that alter the relative consequences of genetic variation. A second kind of alteration in the probability landscape might occur with age. In this case, Fig. 3B might represent the probability landscape during the reproductive years, and Fig. 3D might represent the landscape in the same population at an advanced age.

**Genetic Variation Predicts Long-Term Evolution.** There are many reasons to believe that the genetic variation that segregates

within a population might be irrelevant to long-term evolution. For example, most variation could be in the form of unconditionally deleterious mutations destined for quick elimination from the population, and conversely those rare mutations that will lead to major phenotypic changes might not be polymorphic for long. Contrary to this expectation, comparison of standing genetic variation in phenotype with patterns of among-species divergence suggests the relationship can be reasonably strong. Most work along these lines has relied on the relationship between the direction with the most genetic variance in P space, the first eigenvector of  $\mathbf{G}$ , called  $\mathbf{g}_{\max}$ , and the direction of evolutionary change (50). In most cases, the angle between these directions is less than expected under random models of change. Recent work has widened the scope of such comparisons to include all possible directions in multivariate P space, and here again those directions that show evolutionary change tend to have the most variation (51–53). The existence of relationships between variation and evolution suggests that the variation present in populations reflects deep conserved properties of the G-P map in ways that are not fully understood.

### Phenomics: What Needs To Be Done

The foremost reason that G space is the favored locale for G-P studies is clear: “Collecting phenotypic data . . . is expensive and time consuming . . .” (54). Fifty years ago few could have imagined how our ability to obtain molecular data would increase; 20 years ago few could have imagined the scale at which we can now collect genomic information; 10 years ago few anticipated that genome-scale data could become as cheap as it now is. A key to this set of transformations was the vision of the Human Genome Project, which brought intellectual, technical, and financial resources to bear on genomes. Now is the time for a phenome project bringing the same kinds of gains in throughput and economic efficiency to the study of the phenotype.

Many biologists share my enthusiasm for the prospects of phenomics. There are increasing numbers of self-described phenome projects that should be wholeheartedly supported. The most useful of these take advantage of species where differentiated genotypes already exist as a scaffold onto which phenotype information can be added (37–38, 55). Inspection of the details of these projects, however, reveals that they are makeshift, shoestring operations compared with the magnitude of the challenges. We are pursuing phenomics as a piecemeal, small-science endeavor.

The need for a bigger-science approach is most apparent in the development of high-throughput approaches to phenotyping. To take one example, imaging is an extremely promising source of phenotypic data. The analysis of images should be generalizable across many different organisms and many different sorts of phenotypes (morphology of course, but also flows, spatial locations of metabolites, etc.). To maximize throughput, one would obviously optimize hardware for rapid, repeatable imaging, but also optimize specimen handling, automate phenotyping in software, and solve database issues to allow the handling of the massive amount of data that would result, among other challenges. The efforts of biologists who exploit imaging for pheno-

typing always far short of these ideals, however. Biologists use a huge variety of different, often ad hoc techniques for dealing with such data. In many cases, automation is restricted to use of a computer mouse. Sophisticated approaches are often applied to reduce the complexity of the phenotype measured to just a few dimensions, rather than to acquire intensive phenotypic data. My own approach to *Drosophila* wing measurement (34) reduces handling time of a live specimen to about a minute for all operations, but could readily be improved in various ways. For example, the low resolution and depth of field in the images prevents us from characterizing the cells and hairs clearly visible on the wing; the software we depend on was written to recover the locations, but not the thicknesses of veins. Because we already have an immobilized specimen, why not characterize body parts other than the wing?

The fundamental problem for phenomics is that the need for expertise is truly transdisciplinary (33). We need engineers, computer scientists, mathematicians, and statisticians as much as all flavors of expertise in biology (56). The time for the Human Genome Project did not arrive until fast and inexpensive methods were developed. Coordinated, large-scale efforts to develop such approaches are what is currently missing from phenome efforts. As in the case of images, general approaches to phenotyping applicable across many organisms are surely possible for groups with the right expertise.

Therefore, although biologists continue valuable piecemeal efforts toward phenomics, we need large-scale efforts with the following aims: (i) further development of robust, general high-throughput phenotyping techniques; (ii) combined sequencing and phenotyping efforts that expand from the handful of genotypically controlled model systems, such as mice, to encompass natural population variation; and (iii) further development of analytical approaches that can use high-dimensional genotypic, endo-phenotypic, and end-phenotypic data to generate well-supported hypotheses for further testing.

Short of the ideal project outlined above, humans are clearly the one outbreeding species where the prospects for informative phenomics are the greatest. We have the peculiar tendency to measure our own species obsessively; the biomedical community is the one best positioned to provide the most complete phenomic data. The ultimate reward is to understand the G-P maps needed to turn biology and medicine from descriptive to predictive sciences.

We did not begin to study genomes because we care about genotypes; we study genomes because we care about phenotypes, the health and well-being of humans and the diversity of life on Earth. Now is the time to begin to take the study of the phenotype as seriously as we take the study of the genotype. We must number, locate, and measure even the hairs of our heads, the details of the phenotype, so that we can understand which of those details matter.

**ACKNOWLEDGMENTS.** I thank the organizers Randy Nesse and Raju Govindaraju for the invitation to participate in the symposium, Charles Roseman and Jim Cheverud for sharing unpublished data, and Stevan J. Arnold and an anonymous reviewer for detailed comments. This work was supported by National Science Foundation Grants DEB-0344417 and DEB-0129219 and the National Institutes of Health through National Institutes of Health Roadmap for Medical Research Grant U54 RR021813.

- Lewontin RC (1974) *The Genetic Basis of Evolutionary Change* (Columbia Univ Press, New York).
- Waddington CH (1942) The epigenotype. *Endeavor* 1:18–20.
- Weiss KM (2008) Tilting at quixotic trait loci (QTL): An evolutionary perspective on genetic causation. *Genetics* 179:1741–1756.
- Goldstein DB (2009) Common genetic variation and human traits. *N Engl J Med* 360:1696–1698.
- Davey Smith G, Ebrahim S (2003) Mendelian randomization: Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 32:1–22.
- Blangero J (2004) Localization and identification of human quantitative trait loci: King Harvest has surely come. *Curr Opin Genet Dev* 14:233–240.
- McKusick-Nathans Institute of Genetic Medicine (2009) OMIM: Online Mendelian Inheritance in Man. Available at [www.ncbi.nlm.nih.gov/sites/entrez?db=omim](http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim). Accessed June 30, 2009.
- Bodmer W, Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 40:695–701.
- Kung H-C, Hoyert DL, Xu J, Murphy SL (2008) Deaths: Final data for 2005. *Nat Vital Stat Rep* 56:1–121.
- Goldgar DE, Easton DF, Cannonalbright LA, Skolnick MH (1994) Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Nat Cancer Inst* 86:1600–1608.
- Pohjola-Sintonen S, Rissanen A, Liskola P, Luomanmaki K (1998) Family history as a risk factor of coronary heart disease in patients under 60 years of age. *Eur Heart J* 19:235–239.

12. Wellcome Trust Case Control Consortium (2007) Genomewide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–678.
13. Lupski JR (2007) Genomic rearrangements and sporadic disease. *Nat Genet* 39:543–547.
14. Stefansson H, et al. (2008) Large recurrent microdeletions associated with schizophrenia. *Nature* 455:232–236.
15. Tenesa A, Dunlop MG (2009) New insights into the aetiology of colorectal cancer from genomewide association studies. *Nat Rev Genet* 10:353–358.
16. Guerra S, Martinez FD (2008) Asthma genetics: From linear to multifactorial approaches. *Annu Rev Med* 59:327–341.
17. Weedon MN, et al. (2008) Genomewide association analysis identifies 20 loci that influence adult height. *Nat Genet* 40:575–583.
18. Lettre G, et al. (2008) Identification of 10 loci associated with height highlights new biological pathways in human growth. *Nat Genet* 40:584–591.
19. Gudbjartsson DF, et al. (2008) Many sequence variants affecting diversity of adult human height. *Nat Genet* 40:609–615.
20. Hirschhorn JN (2009) Genomewide association studies: Illuminating biologic pathways. *N Engl J Med* 360:1699–1701.
21. Reed D, Lawler M, Tordoff M (2008) Reduced body weight is a common effect of gene knockout in mice. *BMC Genetics* 9:4.
22. Buchanan AV, Weiss KM, Fullerton SM (2006) Dissecting complex disease: The quest for the philosopher's stone? *Int J Epidemiol* 35:562–571.
23. Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics* (Addison Wesley Longman, Essex, UK), 4th Ed.
24. Janzen FJ, Stern HS (1998) Logistic regression for empirical studies of multivariate selection. *Evolution (Lawrence, Kans)* 52:1564–1571.
25. Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution (Lawrence, Kans)* 37:1210–1226.
26. Mitchell-Olds T, Shaw RG (1987) Regression analysis of natural selection: Statistical inference and biological interpretation. *Evolution (Lawrence, Kans)* 41:1149–1161.
27. Weng G, Bhalla US, Iyengar R (1999) Complexity in biological signaling systems. *Science* 284:92–96.
28. Bassingthwaite JB (2000) Strategies for the physiome project. *Ann Biomed Eng* 28:1043–1058.
29. Paigen K, Eppig JT (2000) A mouse phenome project. *Mamm Genome* 11:715–717.
30. Houle D (2001) in *The Character Concept in Evolutionary Biology*, ed Wagner GP (Academic, New York), pp 109–140.
31. Freimer N, Sabatti C (2003) The human phenome project. *Nat Genet* 34:15–21.
32. Oti M, Huynen MA, Brunner HG (2008) Phenome connections. *Trends Genet* 24:103–106.
33. Bilder RM, et al. (2009) Phenomics: The systematic study of phenotypes on a genome-wide scale. *Neuroscience*, in press.
34. Houle D, Mezey J, Galpern P, Carter A (2003) Automated measurement of *Drosophila* wings. *BMC Evol Biol* 3:25.
35. Ohya Y, et al. (2005) High-dimensional and large-scale phenotyping of yeast mutants. *Proc Natl Acad Sci USA* 102:19015–19020.
36. Vizeacoumar FJ, Chong Y, Boone C, Andrews BJ (2009) A picture is worth a thousand words: Genomics to phenomics in the yeast *Saccharomyces cerevisiae*. *FEBS Lett* 583:1656–1661.
37. Beckers J, Wurst W, de Angelis MH (2009) Toward better mouse models: Enhanced genotypes, systemic phenotyping, and envirotypes modelling. *Nat Rev Genet* 10:371–380.
38. Grubb SC, Maddatu TP, Bult CJ, Bogue MA (2009) Mouse phenome database. *Nucleic Acids Res* 37:D720–D730.
39. Kristensen E, Parsons TE, Hallgrímsson B, Boyd SK (2008) A novel 3D image-based morphological method for phenotypic analysis. *IEEE Trans Biomed Eng* 55:2826–2831.
40. Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) *Geometric Morphometrics for Biologists: A Primer* (Elsevier, Amsterdam).
41. Mosimann JE (1970) Size allometry: Size and shape variables with characterizations of the lognormal and generalized gamma distributions. *J Am Stat Assoc* 65:930–945.
42. Mezey JG, Houle D (2005) The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. *Evolution (Lawrence, Kans)* 59:1027–1038.
43. Kirkpatrick M, Meyer K (2004) Direct estimation of genetic principal components: Simplified analysis of complex phenotypes. *Genetics* 168:2295–2306.
44. Hine E, Lachish S, Higgie M, Blows MW (2002) Positive genetic correlation between female preference and offspring fitness. *Proc R Soc London Ser B* 269:2215–2219.
45. Blows MW, Chenoweth SF, Hine E (2004) Orientation of the genetic variance-covariance matrix and the fitness surface for multiple male sexually selected traits. *Am Nat* 163:329–340.
46. Phillips PC, Arnold SJ (1989) Visualizing multivariate selection. *Evolution (Lawrence, Kans)* 43:1209–1222.
47. Brooks R, et al. (2005) Experimental evidence for multivariate stabilizing sexual selection. *Evolution (Lawrence, Kans)* 59:871–880.
48. Hunt J, Blows MW, Zajitschek F, Jennions MD, Brooks R (2007) Reconciling strong stabilizing selection with the maintenance of genetic variation in a natural population of black field crickets (*Teleogryllus commodus*). *Genetics* 177:875–880.
49. Crespi B, Summers K, Dorus S (2009) Genomic sister-disorders of neurodevelopment: An evolutionary approach. *Evol Appl* 2:81–100.
50. Schluter D (1996) Adaptive radiation along genetic lines of least resistance. *Evolution (Lawrence, Kans)* 50:1766–1774.
51. Hansen TF, Armbruster WS, Carlson ML, Pélabon C (2003) Evolvability and genetic constraint in *Dalechampia* blossoms: Genetic correlations and conditional evolvability. *J Exp Zool B Mol Dev Evol* 296:23–39.
52. Hansen TF, Houle D (2008) Measuring and comparing evolvability and constraint in multivariate characters. *J Evol Biol* 21:1201–1219.
53. Hunt G (2007) Evolutionary divergence in directions of high phenotypic variance in the ostracode genus *Poseidonamicus*. *Evolution (Lawrence, Kans)* 61:1560–1576.
54. Hancock AM, Di Rienzo A (2008) Detecting the genetic signature of natural selection in human populations: Models, methods, and data. *Annu Rev Anthropol* 37:197–217.
55. Canine Phenome Project (2009) The Canine Phenome Project. Available at [www.caninephenome.org](http://www.caninephenome.org). Accessed July 16, 2009.
56. Anonymous (2007) Geneticist seeks engineer; must like flies and worms. *Nat Methods* 4:463.
57. Willmore KE, Roseman CC, Rogers J, Richtsmeier JT, Cheverud JM (2009) Genetic variation in baboon craniofacial sexual dimorphism. *Evolution (Lawrence, Kans)* 63:799–806.
58. Cheverud JM (1995) Morphological integration in the saddle-backed tamarin (*Saguinus fuscicollis*) cranium. *Am Nat* 145:63–89.
59. Cheverud JM (1996) Quantitative genetic analysis of cranial morphology in the cotton-top (*Saguinus oedipus*) and saddle-back (*S. fuscicollis*) tamarins. *J Evol Biol* 9:5–42.