Genetic approaches in comparative and evolutionary physiology

Jay F. Storz¹, Jamie T. Bridgham², Scott A. Kelly³, Theodore Garland, Jr.⁴

¹School of Biological Sciences, University of Nebraska, Lincoln, NE
²Institute of Ecology and Evolution, University of Oregon, Eugene, OR
³Department of Zoology, Ohio Wesleyan University, Delaware, OH
⁴Department of Biology, University of California, Riverside, Riverside, CA

Corresponding author:
J. F. Storz. School of Biological Sciences, University of Nebraska, Lincoln, NE, 68588.
E-mail: jstorz2@unl.edu

Running title: Genetic approaches in evolutionary physiology

Key words: complex traits, physiological genomics, QTL mapping, selection experiments, systems genetics
ABSTRACT

Whole-animal physiological performance is highly polygenic and highly plastic, and the same is generally true for the many subordinate traits that underlie performance capacities. Quantitative genetics therefore provides an appropriate framework for the analysis of physiological phenotypes and can be used to infer the microevolutionary processes that have shaped patterns of trait variation within and among species. In cases where specific genes are known to contribute to variation in physiological traits, analyses of intraspecific polymorphism and interspecific divergence can reveal molecular mechanisms of functional evolution and can provide insights into the possible adaptive significance of observed sequence changes. In this review we explain how the tools and theory of quantitative genetics, population genetics, and molecular evolution can inform our understanding of mechanism and process in physiological evolution. For example, lab-based studies of polygenic inheritance can be integrated with field-based studies of trait variation and survivorship to measure selection in the wild, thereby providing direct insights into the adaptive significance of physiological variation. Analyses of quantitative genetic variation in selection experiments can be used to probe interrelationships among traits and the genetic basis of physiological trade-offs and constraints. We review approaches for characterizing the genetic architecture of physiological traits, including linkage mapping and association mapping, and systems approaches for dissecting intermediary steps in the chain of causation between genotype and phenotype. We also discuss the promise and limitations of population genomic approaches for inferring adaptation at specific loci. We end by highlighting the role of organismal physiology in the functional synthesis of evolutionary biology.
INTRODUCTION

Within-species variation in whole-animal physiological performance is typically attributable to the combined effects of allelic variation at many loci and environmentally induced variation occurring across all stages of development and even prior to conception. That is, physiological performance can generally be regarded as a classic quantitative trait. The same is generally true for the many subordinate traits that underlie whole-animal performance capacities. For example, aerobic exercise performance in rodents is a function of the pulmonary, cardiovascular, muscular, and neural systems, each of which comprises numerous components. For these reasons, quantitative genetics is well-suited to the analysis of physiological phenotypes and can be used to infer the microevolutionary processes that have shaped patterns of trait variation within and among species.

In cases where it is possible to identify allelic polymorphisms at specific loci that contribute to variation in physiological traits, the tools and theory of population genetics can provide insights into the evolutionary processes that have shaped observed patterns of genetic variation within species (48, 185). In comparisons of homologous genes among species, analysis of sequence divergence and reconstruction of ancestral sequences permit inferences about historical pathways of molecular evolution (46). Moreover, ancestral sequence reconstructions provide a starting point for genetic engineering experiments that can provide detailed insights into the biochemical and biophysical mechanisms that are responsible for evolved changes in molecular function (76, 77).

Here we review how genetic approaches can inform our understanding of mechanism and process in physiological evolution. We also review approaches for identifying genetic loci that contribute to variation in physiological traits, including both forward and reverse genetics approaches. We do not discuss evolutionary analyses in general, such as the use of phylogenetic comparative methods or ancestral state reconstructions of phenotypes (63, 121, 154); instead, we focus more specifically on how the tools and theory of quantitative genetics, population genetics, and molecular evolution can be used to shed light on the evolution of physiological and biochemical phenotypes. Throughout, we maintain a primary focus on whole-organism physiological performance because it plays a pivotal role in both function and evolution.
Quantitative genetics

‘Physiology is an attractive field for quantitative genetic work for a number of reasons. First, the functional significance of many traits is well understood. The traits are not mere markers of an underlying genetic system, but play specific, known roles in adaptive processes. Second, comparative work is often available and can reveal which traits are evolutionarily labile. Third, many physiologists work with animals sampled directly from nature so that specific natural reference populations can be established for genetic work.’ - Arnold (4)(p.207)

Quantitative genetics was developed as a means of estimating the genetic and environmental components of variation in phenotypes that are characterized by more-or-less continuous distributions of trait values, such as body weight or basal metabolic rate (51, 125). Quantitative genetics makes use of variances and covariances in measured trait values among individuals with known pedigree relationships. The additive genetic component of trait variation (the fraction of variation that can be passed on from parents to offspring and that is used to calculate narrow-sense heritability) is of special relevance because it is the primary determinant of a trait’s potential for evolutionary change.

As stated by Lynch (124)(p.497): ‘For the physiological ecologist, [narrow-sense] heritability is the most useful piece of genetic information since it is both descriptive and predictive.’ It is predictive in the sense that resemblance between parents and offspring for a given trait enables us to forecast the short-term response to selection on the trait. At the phenotypic level, selection describes statistical change in the means of traits within the parental generation. Inheritance then determines how the effects of selection are transmitted from the parental generation to the offspring generation, i.e., the genetic response to selection. Both narrow-sense heritability and selection have now been measured for many physiological traits (84, 200), even in human populations (175).

The genetic property of parents that accounts for phenotypic resemblance with their offspring is additive genetic value: the sum across loci of the average effects of alleles on a given trait. For a trait that is under polygenic control, an individual’s phenotypic value has an additive genetic component (which is transmitted from parents to offspring), a nonadditive genetic component due to dominance and epistasis (which is not transmitted to offspring because the
allelic combinations found in parents are broken up when gametes are produced), and an environmental component.

The narrow-sense heritability of a trait is the ratio of variance in additive genetic values (additive genetic variance) divided by variance in phenotypic values (phenotypic variance). This ratio can be calculated as the slope of a regression of additive genetic values (the average trait values for adult offspring) on the measured phenotypic values of parents (assuming the absence of non-genetic parental effects, as might occur through maternal care). The central axiom of quantitative genetics states that – from one generation to the next – the response to selection on a single trait is equal to the product of the narrow-sense heritability of the trait and the *directional selection differential*, a measure of the strength of selection. The selection differential is the difference in mean trait values between individuals who successfully reproduce in a given generation (actual parents) and the larger, more inclusive set of all potential parents. In the context of a selection experiment, the heritability of a trait is the fraction of the selection differential that is transmitted to the next generation, manifest as a change in the average value of the trait (123).

**Multivariate quantitative genetics**

Artificial selection as practiced by animal breeders is often applied to measurements of a single trait (e.g., growth rate, milk yield), but selection in nature inevitably acts on multiple traits. The desire to predict responses to selection that acts simultaneously on multiple traits motivated the development of a multivariate theory of quantitative inheritance and selection (3, 6, 14, 109). To predict responses to selection that acts simultaneously on multiple traits, we need to measure additive genetic covariances between traits. The measured covariances (standardized as genetic correlations) describe the extent to which pairs of traits may be genetically coupled due to pleiotropy (effects of a single gene on multiple traits) and/or linkage disequilibrium (LD, the nonrandom association between alleles at different loci in the gametic phase). In this multivariate theory, heritability is replaced by the additive genetic variance-covariance matrix (the ‘*G*-matrix’) as a descriptor of polygenic inheritance. Likewise, to predict the evolutionary response to selection acting on multiple traits, the single-trait selection differential is replaced by the *directional selection gradient*, which measures the direct force of selection on a given trait, controlling statistically for its correlations with other traits in the analysis. The selection gradient
can be measured as the partial regression of relative fitness on the trait of interest, holding all other traits constant. It predicts the change in relative fitness that would be produced by a unit change in the trait in the absence of changes in genetically correlated traits (111). In addition to predicting short-term responses to selection, quantitative genetic models of long-term phenotypic evolution are also potentially applicable to the analysis of physiological traits (7-9, 50, 82, 91, 107), although these models are premised on controversial assumptions about the constancy of genetic parameters over evolutionary time (14).

Quantitative genetic analysis in a multivariate framework can reveal genetic correlations between traits, which can determine whether directional selection on one trait will produce a correlated response in an unselected trait (3, 92, 111). Genetic correlations between traits can provide insights into the causes of physiological trade-offs and constraints that have always been of interest to both physiologists and evolutionary biologists (4, 28, 61, 64, 159). Documentation of genetic correlations between traits within populations can also suggest hypotheses about the nature of trait variation and co-variation among species. Information about the sign and magnitude of genetic correlations among traits suggests that certain trajectories of phenotypic change are more likely than others as species diverge by random genetic drift or in response to natural or sexual selection. In other words, genetic correlations may bias pathways of evolution along ‘genetic lines of least resistance’ (166, 167).

An important point to emphasize is that – as with narrow-sense heritabilities – elements of the $G$-matrix are not static and will change as a result of changes in allele frequencies that are brought about by selection or drift (91, 92). Moreover, correlated selection on multiple traits may favor particular combinations of alleles at multiple loci, thereby permitting the evolution of co-adapted gene complexes. Thus, traits may become genetically integrated as a partial step towards the evolution of modularity (131). At any given point in time, the form of the $G$-matrix may reflect not only functional relations among traits but also their past history of correlated selection.

Quantitative genetic analysis can also help characterize the allowable dimensions of organismal form and function and can potentially shed light on why certain phenotypes do not exist. The nonexistence of a particular trait or trait combination can be explained in two general ways: (i) it was not adaptive in ancestral environments (it did not offer a solution to any pressing problem faced by ancestral organisms), so the trait never evolved; or (ii) the requisite genetic
variation never existed, so the trait could never evolve no matter how adaptive it might have been. The latter hypothesis can be tested by combining analyses of quantitative inheritance with selection experiments (e.g., (205, 206)).

**Epigenetic sources of phenotypic variation**

When considering environmentally induced changes in physiological traits, it is important to note that some environmental effects, especially those experienced early in life during ‘critical periods’, can be passed on to offspring. Early-life effects may occur through variation in maternal care, transmission of hormones and nutrients to offspring, developmental plasticity, fetal adaptation and maladaptation, changes in appetite and dietary preferences, energy availability, and compensatory growth. Certain early-life effects can amplify across successive generations (204). Moreover, they can interact with other sources of environmental and genetic variation, including sex (207), in ways that facilitate or constrain evolutionary change at multiple levels of biological organization (11, 26, 141). Much more work is required to fully elucidate the role of epigenetics in physiology and evolution (45, 56, 80, 87, 88).

- Of the various potential mechanisms that could be responsible for the persistent effects of early exposures (203), induced alterations in epigenetic regulation likely play an important role in developmental programming of adult behavioral and physiological traits (202). In particular, the epigenetic mechanism of DNA methylation is a prime candidate; developmental establishment of DNA methylation is affected by the environment (90) and, once established, is maintained with high fidelity (30), enabling the long-lasting stability that is the hallmark of developmental programming.

**Analysis of phenotypically plastic traits**

Under natural conditions, most physiological traits are characterized by a sizable environmental component of phenotypic variance, and this has important implications for interpreting patterns of trait variation within and among populations. For many physiological phenotypes in animals, it is generally a safe bet that a substantial fraction of the observed trait variation among conspecific populations will be attributable to reversible acclimatization (phenotypic plasticity) rather than genetic differentiation. As stated by Garland and Adolph (62)(p.194) ‘Recognizing problems introduced by rampant acclimatization, physiologists may have been inhibited from
searching for relatively small population differences. Consequently, the focus on proximate mechanisms of coping with environmental change (i.e., acclimatization) drew more attention than the possibility of genetic differentiation among populations. The lability of most physiological traits suggests that common-garden or reciprocal transplant experiments are essential for interpreting patterns of trait variation in an evolutionary context. Such experiments can be used to characterize the range of phenotypic variation that is achievable via acclimation/acclimatization, and can reveal the extent to which phenotypic plasticity might obviate the need for genetically based, evolutionary changes to cope with a changing environment. For example, a given high-altitude animal species may have a higher hematocrit than closely related lowland species when the trait is measured in wild-caught animals at their native elevation, but common-garden or reciprocal transplant experiments would be required to measure the environmental component of the observed trait variation among species. It may be that the elevated hematocrit in the high-altitude species is purely environmentally induced, reflecting an increased production of erythropoietin in response to hypoxemia.

Common-garden experiments are also critical in multi-species comparative studies because environmentally induced variation can confound the detection and interpretation of phylogenetic signal in trait values. If environmentally labile traits are not measured under common-garden conditions, then it may not be possible to determine the extent to which phenotypic similarity between a given pair of species is attributable to shared phylogenetic heritage vs. a shared, plastic response to a common environment factor. For example, a given pair of high-altitude species may have similarly high hematocrits when measured under the hypoxic conditions of their native environment. One possible explanation for the phenotypic similarity is that the two species inherited a constitutively elevated hematocrit from a common ancestor. But before drawing that conclusion, we would first need to rule out the alternative possibility that the two species have simply responded in similar fashion to the same hypoxic stimulus in their shared environment.

**Physiology, performance, behavior, and fitness: measuring selection in the wild**

Several different strategies are used to study the evolutionary adaptation of physiology...The one traditionally used by physiological ecologists involves examining comparative data on contemporaneous organisms (e.g., species in different environments) and then deriving post-hoc
reconstructions of patterns that have evolved over historical time…An alternative approach
involves studying selection on phenotypic traits in nature…’ - Bennett and Huey (22)(p. 265)

If a given performance trait is variable among individuals and if some fraction of that variation is
heritable, then it is possible to examine the direct and indirect effects of selection on the trait in
real time and to detect responses to multivariate selection over evolutionary time (5, 22). This
can be accomplished by measuring the statistical relationship between Darwinian fitness (or
estimated components of fitness, such as survival and fecundity) and multivariate phenotypes in
natural populations. The strategy is based on multivariate selection theory, which deals with the
effects of selection that acts simultaneously on multiple traits (107, 108, 110, 111). With
experimental measurements of inheritance and observational data on selection (shifts in trait
means within a generation), multivariate selection theory can be used to predict evolutionary
change in trait means from one generation to the next (the directional response to selection
(107)). The key parameter that predicts the directional response to selection on a given trait is the
selection gradient described above (Quantitative genetics). Arnold (5) described an approach
based on path analysis to factor the selection gradient into two parts: a ‘performance gradient’
(the measured effect of the trait on some ecologically relevant measure of organismal
performance) and a ‘fitness gradient’ (the effect of performance on fitness; Fig. 1). This
paradigm, which focused attention on the importance of measuring whole-organism animal
performance, has since been extended to include behavior (28, 68, 106)(Fig. 1).

This approach has been applied successfully in a number of vertebrates (84), including
studies of locomotor capacities of garter snakes (Thamnophis sirtalis (89)) and thermogenic
capacities of high-altitude deer mice (Peromyscus maniculatus (79)). These studies followed the
same basic formula: physiological performance was measured in wild-caught animals (known-
age cohorts in the garter snake study), and rates of survivorship were then estimated using mark-
release-recapture protocols. Both studies measured the direction and magnitude of selection on
physiological capacities in free-ranging animals, and therefore provided insights into the
adaptive significance of naturally occurring trait variation. As stated by Bennett and Huey
(22)(p.275): ‘…such studies have the promise of freeing physiological ecology from an
implicitly adaptationist program…by turning attention to the process of adaptation, rather than
its simple assumption. These protocols also open hosts of new and interesting questions and
firmly embed studies within the natural environments, demography and ecology of the organisms investigated.’

**Selection experiments and experimental evolution**

Selection experiments may be viewed as the earliest form of ‘genetic engineering’. Humans were altering the genetic compositions of plant and animal populations through domestication thousands of years before we understood how heredity worked. The impressive nature of these feats was one major line of evidence used by Charles Darwin when he developed the theory of evolution by natural selection. Although he did not recognize them as such (69), intentional and unintentional selection experiments are an experimental way to study ‘evolution in action’. They are also a way to produce ‘useful’ organisms. The response to selective breeding constitutes the most direct and convincing test of whether a given trait harbors additive genetic variance. More recently, is has been noted that selection experiments are a modern corollary to the August Krogh Principle: if a suitable model does not exist, then create one (21)!

Selection experiments are an excellent way to probe the interrelations among traits, including complex traits, because one can atomize an organism to the desired level, impose selection at that level on a trait of interest, and observe correlated cross-generational changes in other traits (58, 60, 66, 187). They are a good way to test specific hypotheses about putative trade-offs (61) and constraints on the way organisms can evolve (e.g., (205, 206, 211)). In particular, they can be a powerful way to demonstrate mechanism, i.e., how organisms work (58, 60). For example, one could impose selection on some measure of whole-animal performance and then observe a correlated response in one or more lower-level traits that plausibly cause or permit the change in performance. This hypothesis of causality could then be tested by conducting a second experiment, selecting on the lower-level trait and determining whether performance changes as predicted. As a hypothetical example, one could breed for high VO$_{2}$max in rodents (e.g., (72, 73, 211)) and potentially observe correlated changes in hematocrit. A second experiment could select for high hematocrit and see if VO$_{2}$max increases as predicted. As another example, one might select for basal metabolic rate (BMR)(103, 105), observe a correlated change in circulating thyroid hormone levels, then conduct another experiment breeding for hormone levels.
Finally, as discussed in a separate section (Linkage mapping/association mapping), selection experiments can aid attempts to find the genes that underlie phenotypic variation. Specifically, genetic crosses between selected lines and non-selected or oppositely-selected lines can be used to produce a mapping population with sufficient levels of both genotypic and phenotypic variation for detecting quantitative trait loci (QTL).

As discussed elsewhere, various types of experiments can be considered under the general categories of selection experiments and experimental evolution (60, 69, 93, 189). The key attributes of these experiments include an emphasis on hypothesis testing, tracking of phenotypes (and genotypes) across generations, replication, and potential reproducibility. Beyond those fundamental principles, selection experiments encompass a tremendous range of empirical studies with a wide variety of organisms (69), often including microorganisms and even digital ‘organisms’ (221). Replication is crucial to allow strong inferences regarding both direct and correlated responses, because genetic drift can lead to divergence between selected and control lines for reasons unrelated to functional relationships or pleiotropic gene action (130). Replication also allows the discovery of multiple solutions to adaptive challenges (60, 67, 151, 160, 189).

At one end of the continuum, various organisms have been introduced to new habitats by both intentional and unintentional human activities. English house sparrows and *Drosophila* species are two familiar examples (86). Sparrows are represented in various museum collections in a way that offers some ability to track phenotypic changes across generations, and these introductions have been ‘replicated’ many times. Physiological ecologists have formulated specific hypotheses about how sparrows should evolve in response to altered environmental conditions, and these can be tested with measurements of individuals sampled from different populations and then raised under common-garden conditions (62, 117, 118, 134).

At the other end of the continuum lie laboratory-based artificial selection experiments, many of which have been conducted with rodents (156, 187, 211). Some of the examples we find most compelling involve replicated selection at the levels of behavior or whole-animal physiological performance. For eight generations, Hayes and colleagues selectively bred four replicate lines of laboratory house mice for high mass-independent maximal oxygen consumption (VO$_2$max) during forced treadmill exercise (211)(Fig. 2). The selection criterion included body mass as a covariate, so selection was independent of effects of body mass on
metabolic rate. In four other lines, they antagonistically selected for a combination of high mass-independent VO$_2$max and low mass-independent BMR. Four additional lines were maintained as non-selected controls. Compared with controls, VO$_2$max significantly increased by 11.2% in lines bred for VO$_2$max, while BMR did not change significantly (+2.5%). Compared with controls, VO$_2$max significantly increased by 5.3% in antagonistically selected lines, while BMR was not significantly reduced (-4.2%). Overall, analyses indicated a weak positive genetic correlation between VO$_2$max and BMR. These results offer weak support for the idea that selection to increase a mammal’s aerobic capacity might be accompanied by an energetic cost in terms of higher resting metabolic rate.

For more than two decades and 70 generations, Garland and colleagues have bred four replicate lines of laboratory house mice for high levels of voluntary wheel running, while also maintaining four non-selected control lines (60, 156, 187, 188). It is important to emphasize that in these sorts of experiments the line is the experimental unit. Therefore, statistical comparisons of the selected and control lines use so-called mixed models with line as a random effect nested within line-type, and degrees of freedom for testing the effect of selection are (in the wheel-running example) 1 and 6. Even if dozens or hundreds of individual animals are measured for a given comparison, the tests for effects of direct or correlated responses to selection will still be based on 1 and 6 degrees of freedom.

As expected from numerous studies selecting for locomotor behavior or performance in rodents (e.g., see Table 1 in (53)), wheel running responded rapidly to selection in the High Runner (HR) lines before reaching apparent limits at generation 16-28, depending on line and sex (29, 67), and wheel running has not increased appreciably in the HR lines for an additional ~40 generations of selective breeding (T. Garland, Jr., unpublished results). Mice in the HR lines have mainly evolved to run faster, rather than running for more minutes per day (especially in the case of females), perhaps partly because it is more economical (155) or more rewarding (20, 70, 156) to run faster.

Many correlated responses to selection have been documented in the HR mice, including increases in VO$_2$max (101) and treadmill endurance capacity (133). HR mice are also smaller, leaner, reach higher maximum speeds and run more intermittently during wheel running, and have a reduced incremental cost of transport on a whole-animal basis, though not on a mass-corrected basis (155, 190, 191). Sub-organismal traits that have evolved in HR mice include
larger hearts (71), increased hindlimb symmetry, larger femoral heads, heavier foot bones, and altered semicircular canal shape (65, 96, 168). Although many of these correlated evolutionary responses would have been anticipated based on physiological knowledge, others might not have been, such as the increase in baseline circulating corticosterone levels or the reduced growth rate and smaller adult body size of HR mice (70, 156, 187). Several lines of evidence indicate altered brain reward systems in HR mice (70), including an increase in overall brain size and size of the midbrain (102). However, there are a number of evolutionary changes that could have been expected to occur in HR mice (e.g., changes in resting or basal metabolic rate, litter size, maternal care behavior, and general levels of aggression) that did not occur (70). As with any approach in science, well-designed selection experiments may raise more questions than they answer!

Koteja and colleagues used a wild population of bank voles (*Myodes glareolus*) to begin an ambitious selection experiment (16 total lines) designed to model the adaptive evolution of aerobic metabolism in relation to changes in locomotor activity, predatory behavior, and diet (164). All three selection treatments were successful: the replicate lines diverged from the non-selected control lines. Several important observations about correlated responses to selection (or lack thereof) have since been published. For example, after 13-14 generations, selection for high activity-related aerobic metabolism did not alter the capacity for non-shivering thermogenesis (174). In the lines selected for increased aerobic exercise capacities, an analysis of variation in tissue-specific transcriptomes from animals sampled at generation 13 revealed that changes in gene expression make a significant contribution to the selection response (104).

**Through a glass, darkly:** population genetic analysis of single-locus polymorphisms

Alternative alleles at a single gene may have detectable effects on proximal biochemical phenotypes, but the effects on fitness-related physiological performance may typically fall well below the resolving power of our experimental measurements. From an evolutionary perspective, the key question is whether they also fall below the resolving power of natural selection. Specifically, we want to know whether allelic variation in gene function translates into quantitative trait differences, which in turn translate into fitness differences of sufficient magnitude that the deterministic effects of selection are not overwhelmed by the stochastic effects of genetic drift (48, 185) (approximately, when $2N_e s > 1$, where $N_e$ is the effective
population size and $s$ is the selection coefficient of the mutant allele). Although it may seldom be possible to measure fitness variation among alternative single-locus genotypes under natural conditions, the effects of selection are expected to leave an historical imprint on levels and patterns of nucleotide variation at causative loci. Because DNA sequence data integrate the effects of fitness variation over long periods of time, neutrality tests based on DNA polymorphism and/or divergence hold the promise of providing indirect, retrospective insights into histories of positive selection at specific loci (185). These indirect inferences can complement direct, experimental measurements of allelic differences in gene function, and can provide suggestive evidence that the measured differences have adaptive significance. The integration of population-genetic neutrality tests with functional experiments holds the promise of elucidating each step in the so-called ‘adaptive recursion’ (54, 185). As stated by Feder et al. (52)(p.320): ‘the general paradigm is that genes encode the phenotype, the phenotype determines the performance of organisms in natural environments in response to ecological or evolutionary stimuli, the performance determines the evolutionary fitness of alternative genotypes, and the fitness determines the frequency of genotypes in the next generation, in recursive fashion…’ It is this last step in the recursion that is not generally open to empirical scrutiny in natural populations, as it is typically not possible to quantify how differences in net reproductive rates of alternative genotypes affect the allelic composition of the gamete pool in the following generation. The cumulative effects of fitness variation over many past generations can only be inferred indirectly, and population-genetic neutrality tests – despite their inherent limitations – provide the best available means of making such inferences.

As an example of how population genetic analyses can bolster inferences about physiological adaptation, a multilocus analysis of nucleotide polymorphism in high- and low-altitude populations of deer mice (*Peromyscus maniculatus*) revealed evidence for a history of spatially varying selection at duplicated genes that encode the subunit chains of the tetrameric $\alpha_2\beta_2$ hemoglobin protein. The population genetic evidence for spatially varying selection on two $\alpha$-globin gene duplicates and two $\beta$-globin gene duplicates, in combination with functional analyses of native and recombinant Hb variants (135, 136, 179-183), corroborated independent lines of experimental evidence which had demonstrated that allelic variation in hemoglobin-O$_2$ affinity contributes to adaptive variation in whole-animal aerobic performance under hypoxia (31, 32).
In addition to providing corroborative support for adaptive hypotheses, the integration of population-genetic neutrality tests and experimental tests of protein function can also serve a useful function by falsifying such hypotheses (38, 161), thereby providing a check on unbridled speculation and adaptive-storytelling. As stated by Cheviron et al. (38)(p.2954): ‘In the absence of experimental validation, analyses of variation in coding sequence can lead to facile inferences about the adaptive significance of observed amino acid replacements. This is especially true in studies of candidate genes in natural populations where there is some plausible a priori expectation about the adaptive significance of changes in protein function.’

Molecular evolution: analysis of interspecific divergence at specific loci

In comparisons of homologous protein-coding genes among species, one popular approach for detecting a history of positive selection involves a comparison of relative rates of synonymous and nonsynonymous substitutions at interleaved codon positions (214). This approach has proven useful for identifying episodes of positive selection that involved recurrent amino acid substitutions, as might be expected for proteins involved in co-evolutionary arms race dynamics (e.g., immune defense or gamete recognition proteins), but there are good reasons to expect that it is generally less useful for detecting and identifying adaptive modifications of protein function that were caused by small numbers of substitutions (186). Nonetheless, the development of increasingly sophisticated codon-substitution models has spawned a cottage industry of comparative genomic studies that follow a highly formulaic script: (i) compare coding sequences of orthologous genes among multiple species; (ii) identify specific genes with a subset of sites in which the ratio of nonsynonymous-to-synonymous substitution rates ($d_N/d_S$) is significantly greater than unity (the neutral expectation) in a specific lineage; and (iii) tell a ‘just-so’ story about why the observed amino acid substitutions in those genes must have contributed to lineage-specific adaptations. Such analyses can be useful for generating hypotheses about the adaptive significance of observed sequence changes, but such hypotheses ultimately need to be experimentally tested, ideally with protein-engineering experiments that measure the functional effects of specific substitutions on evolutionarily relevant ancestral backgrounds. Studies that integrate evolutionary analyses of sequence divergence with functional analyses of native and/or recombinant proteins can obtain far more detailed and robust evolutionary insights than are
possible with analyses of sequence variation alone. Recent examples of this integrative approach include studies of vertebrate hemoglobins (74, 135, 151, 198) and opsins (25, 192, 217-220).

**Ancestral protein reconstruction**

In general, conclusive inferences about mechanisms of protein evolution (adaptive or not) require experimental insights into the functional effects of historical mutations on the ancestral genetic background in which they occurred. In recent years, especially refined insights into mechanisms of protein evolution have been obtained by using protein engineering experiments that involved the reconstruction and functional testing of resurrected ancestral proteins (76, 77, 196). Using an alignment of nucleotide sequences (or translated amino acid sequences) from homologous protein-coding genes of extant species, the first step in this approach is to estimate a phylogeny of the sampled sequences. Model-based statistical methods are then used to reconstruct ancestral amino acid sequences of proteins at internal nodes in the phylogeny (215). For each variable site in the alignment, maximum likelihood and/or Bayesian statistical methods are used to infer the most likely character state that existed at a particular node, given the character states in observed (extant) sequences, the estimated phylogeny, and the assumed substitution model. The reconstructed coding sequence of the ancestral gene can then be synthesized and cloned into a plasmid construct, followed by expression and purification of recombinant protein in cultured cells. The ‘resurrected’ ancestral protein can then be used for *in vitro* functional experiments and structural studies.

This approach has been applied with great success in studies designed to elucidate the molecular mechanisms by which vertebrate steroid receptors evolved to recognize their hormonal ligands. Steroid receptors are transcription factors that have evolved to recognize specific hormonal activators in order to regulate physiological responses. Following binding of specific steroid ligands, the individual receptors initiate specific downstream signaling cascades. Steroid receptors evolved from an ancestral receptor in the stem lineage of bilaterian animals. Repeated rounds of steroid receptor gene duplication and divergence during vertebrate evolution have produced multiple receptors with affinities for a variety of steroid hormones including estrogens, androgens, progestogens, corticosteroids and mineralocorticoids (197). By reconstructing ancestral receptor proteins at key nodes in the phylogeny of steroid receptors, it was possible to identify particular branches of the phylogenetic tree in which important shifts in
ligand-specificity evolved (Fig. 3). Protein engineering experiments based on site-directed mutagenesis were then used to introduce historical substitutions into ancestral backgrounds to identify which combinations of substitutions were sufficient to recapitulate the historical shifts in ligand specificities. Integrating the functional tests with crystallographic analyses of ancestral proteins provided insights into the biophysical mechanisms that were responsible for the evolved ligand-specificities.

The most ancient steroid receptor, ancestral to all family members (AncSR1) and the ancestor of all keto-steroid (progestogens, androgens and corticosteroids) receptors (AncSR2) were synthesized and functionally characterized using reporter gene assays to determine their hormone affinity (Fig. 3). These experiments led to the discovery that the ancestor of all vertebrate steroid receptors was sensitive to many chemicals, like estrogen, that have aromatic rings, and that this specificity has been retained in the estrogen receptor clade. AncSR2 lost sensitivity to estrogens, and gained specificity for non-aromatized steroids (49). Functional testing of ancestral proteins on the phylogeny bracketed the functional switch that occurred on the interval between AncSR1 and AncSR2 (Fig. 3). Structural analysis of the ancestral proteins revealed two amino acid substitutions that were predicted to alter the relative affinities for aromatic vs. non-aromatized steroids. When these two substitutions, Glu41Gln and Leu75Met, were introduced into the ancestral receptor, AncSR1, they decreased affinity for aromatic estrogens and increased affinity for non-aromatized steroids, resulting in the derived phenotype. Subsequent experiments reversing these two large-effect historical substitutions in the AncSR2 background convincingly demonstrated that they were causal determinants of the evolved ligand specificity. Additional analyses of these ancestral proteins determined the biophysical basis for this switch in hormone specificity (75).

By revealing the molecular mechanisms underlying the evolution of hormone specificity in steroid receptors, this work provided insights into why estrogen receptors are susceptible to mis-regulation by environmental chemicals. The minimal recognition criterion of the oldest ancestor was that ligands must have an aromatized A-ring. Since aromatization is at the end of the enzymatic synthesis pathway leading to estrogens, this specificity prevented the receptors from being activated by biosynthetic intermediates, and this rather simple criterion apparently sufficed. After gene duplication, one duplicate receptor retained the ancestral specificity, while the other evolved a promiscuous response to other steroids in this pathway. This second group
further diversified with descendent lineages evolving specificity for different hormonal intermediates (195). The minimal specificity requirement for an aromatized ring, inherited from the ancestral receptor, predicts the capacity for estrogen receptors to be disrupted by exogenous xenobiotic phenolics. The ancestral receptors existed within physiological environments where they evolved to distinguish among the various endogenous and exogenous compounds to which they were exposed. Their minimal specificity meant that when novel hormones or new environmental conditions were encountered, their endocrine systems could be sensitive to misregulation.

Through recognition of the evolutionary importance of an aromatized ring to estrogen receptor specificity, we now understand why certain environmental chemicals - those that contain aromatized ring structures - can disrupt estrogen receptor signaling, leading to severe developmental, behavioral, and physiological disorders (47). By reconstructing and testing ancestral proteins, together with careful dissection of molecular mechanisms by which proteins distinguish among binding partners, this work revealed the historical causes of susceptibility to endocrine disruption in modern-day vertebrates and provided information about which types of exogenous compounds might interfere with physiological endocrine signaling (49). This case study represents how knowledge of historical context and molecular evolutionary mechanisms can shed light on current physiological phenomena (49, 75, 197).

INSIGHTS INTO THE GENETIC AND MECHANISTIC BASIS OF PHYSIOLOGICAL TRAITS

Linkage mapping/association mapping
Commuzie and Allison (40) introduced a review of the search for human obesity genes in 1998 with the following: “One of the greatest challenges in biomedical research today is the elucidation of the underlying genetic architecture of complex phenotypes such as obesity. At first glance, body weight seems exceptionally simple. It can be defined precisely and measured with great accuracy and reliability. However, recent research on obesity has revealed that body weight is in fact a truly complex phenotype.” The same applies to almost any behavioral, physiological or morphological trait, and to this day, efforts to characterize the genetic architecture of such traits face the same challenges. Here we briefly review mapping approaches for characterizing the genetic architecture of complex physiological traits.
Quantitative trait locus (QTL) mapping experiments attempt to identify specific chromosomal regions—and, ultimately, specific genes and mutations—that contribute to variation in a trait of interest. In the broader enterprise of evolutionary quantitative genetics, the goal of the QTL program is to explain genetic variation for quantitative traits in terms of the underlying genes, the effects of segregating alleles in the context of different environments and different genetic backgrounds, and the frequencies of causal variants in natural populations (13, 126, 127, 129). QTL are typically detected and localized by measuring statistical associations between molecular markers, such as single nucleotide polymorphisms, and the trait of interest in a ‘mapping population’ (e.g., the F2 progeny of a cross between inbred lines; (125)). The premise of QTL mapping is that marker alleles that are associated with differences in trait values are in LD with the causal variant(s) (in this context, LD describes the nonrandom association between alleles segregating at linked sites along a recombining chromosome).

In addition to the usual factors that affect statistical power (e.g., sample size, range of the independent variable, measurement accuracy), the power to detect QTL for a given trait is a function of the effect sizes and frequencies of causative alleles. The effect size of a given QTL is measured as the mean difference in phenotype between alternative marker genotypes, scaled by the phenotypic variance of the trait within each genotype class. The power to detect a given QTL is maximized when the effect size is large and frequencies of alternative marker alleles are equal (=0.50, as is the case for all segregating alleles in the F1 progeny of a cross between two inbred lines). The power to localize the chromosomal position of a given QTL depends on the number of recombination events that have occurred between causative variants and marker alleles at linked loci. QTL can be localized to smaller chromosomal intervals when there have been many cross-overs that uncouple genotype-phenotype associations between all marker loci except those that are tightly linked to the causative variant(s).

In contrast to traditional line-cross mapping populations, association mapping makes use of historical recombination between QTL and marker loci in natural populations (including genome-wide association studies [GWAS] in humans) or in reference panels of recombinant inbred lines, so overall levels of LD are typically much lower. Consequently, association mapping typically has much higher power to localize QTL, as causal variants will have been randomized against a larger number of genetic backgrounds. Moreover, association studies in natural populations simultaneously provide estimates of the frequencies of causal alleles, and can
therefore reveal whether segregating variation for the trait under study is mainly attributable to rare, deleterious alleles maintained at mutation-selection equilibrium, or intermediate-frequency alleles, which would be consistent with selective neutrality or maintenance of variation by some form of balancing selection (13, 127). For any given population, the utility of association mapping (and the number of marker loci that need to be genotyped) depends on the scale and pattern of LD. The ability to localize QTL to narrow intervals is greatest in species like \textit{Drosophila melanogaster}, where LD decays over short distances (i.e., a few tens of kilobases or less (27)). In natural populations of house mice, the rate at which LD decays with distance is similar to that in human populations, and is much higher than that in laboratory strains of mice (113), suggesting that wild-derived strains of house mice are well-suited to fine-scale association mapping.

High-resolution mapping in model organisms has revealed that single QTL often fractionate into multiple, closely linked QTLs, some of which segregate alleles with opposing effects on the focal trait (55). Moreover, allelic effects are often conditional on sex, genetic background (i.e., epistasis, or gene × gene interaction), and/or environment (gene × environment interaction). For these reasons, it has proven extremely difficult to move beyond the delimitation of broad QTL intervals (which may contain hundreds of candidate genes for the trait of interest) to the ultimate goal of positional cloning and identification of causative mutations (i.e., quantitative trait nucleotides, or ‘QTN’). As stated by Mackay (127) (p.1233): ‘[Studies of quantitative genetic variation in \textit{Drosophila}]…highlight how little we know about ‘candidate genes’ affecting quantitative traits: the majority of the genome is uncharted territory with respect to phenotypic effects of naturally segregating alleles affecting even extensively studied phenotypes in a genetic model organism’. However, for model organisms like \textit{Drosophila} and mice, the development of community mapping resources (83, 119, 128, 132) and systems genetics approaches (see below) should greatly facilitate future advances.

As mentioned above (\textit{Selection Experiments and Experimental Evolution}), one way to create mapping populations that harbor high levels of variation for the trait of interest is to cross lines derived from artificial selection experiments. In such cases, intercrosses are typically carried out to the second (F$_2$) generation or backcrosses are performed, yielding a high level of variation in the selected trait (137). However, F$_2$ and backcross populations yield minimal recombination events and the identified QTLs are typically located in broad intervals that may
contain numerous candidate genes (44). Through random intercrossing over multiple generations
beyond the F2, advanced intercross lines (AILs) accumulate recombination events and provide
increased mapping resolution (narrowing confidence intervals) (43). Kelly et al. (99) utilized one
of four replicate lines of house mice that were selectively bred for high voluntary wheel-running
(discussed above) and the inbred mouse strain C57BL6/J to create an AIL suitable for mapping
voluntary wheel-running traits. Animals were genotyped and phenotyped at generations 4 (G4)
\( n=800 \text{ mice} \) and 10 (G10) \( n=473 \text{ mice} \) (99, 115). Analyses of the G4 yielded 32 significant and
13 suggestive QTL for running distance, duration, average speed, and maximum speed. The
largest QTL effect accounted for 6.6% of the total phenotypic variation, indicating that variation
in the measured trait is mainly attributable to allelic differences at many loci that have
individually small effects (99). In an effort to prioritize candidate genes, Kelly et al. (97, 98)
examined global gene expression profiles in brain and muscle tissue. These two tissues were
chosen for the purpose of investigating the motivational aspects (brain) as well as the physical
abilities (muscle) that contribute to exercise performance. Examination of QTL for transcript
abundance (expression quantitative trait loci, eQTL) that co-localized with QTL for the
performance trait implicated hundreds of potential candidate genes that may be functionally
important in the control of voluntary wheel running behavior. This is because co-localization of
QTLs for a given trait and the expression level of a particular gene suggest that \textit{cis}-regulatory
changes in the gene may contribute to the observed variation in trait values (\textit{cis}-regulatory
elements are non-coding loci that regulate the transcription of nearby genes). A backcross
mapping population derived from a different HR line (compared with the AIL above) was also
used to identify a causative mutation (autosomal recessive) in a gene of major effect that is
primarily characterized by a major reduction in hindlimb muscle mass (78, 95).

These findings represent an important initial step in understanding the genetic
architecture of voluntary wheel running behavior in the artificial selection experiment (see Kelly
et al. (99) for a discussion of limitations). However, this approach is limited by the variation
segregating in the two parental strains used to map QTL for voluntary activity. This problem of
restricted genetic sampling can be circumvented by using reference panels of recombinant inbred
lines, such as the Collaborative Cross or the Diversity Outbred mouse population (119, 132).
Finally, we note that it is important to consider the evolutionary ‘relevance’ of the mapping
population in addition to considerations of experimental tractability or statistical power. If the
goal of a QTL study is to identify genes that can affect a particular trait in a particular species, then choices based purely on statistical power, tractability, and cost may reasonably take precedence. These considerations have generally driven development of community mapping resources. However, if the goal is to determine which loci have contributed to phenotypic changes in particular species or populations, then results obtained from ‘generalized’ mapping populations may not be evolutionarily relevant.

**Systems genetics**

‘Understanding the relationship between DNA sequence variation, transcriptional, protein and metabolite networks and organismal-level phenotypes is the main challenge for the future and will add the missing biological context to genotype-phenotype associations’ – Mackay (127) (p.1235)

‘…[N]atural genetic variation has the character of an ideal multifactorial perturbation, providing a natural experimental design that is helping researchers to uncover the mechanistic basis of the map that connects genotype to phenotype. As the molecular catalogues of genomics yield to the integrated models of systems biology, natural genetic variation will have an increasingly central role.’ – Rockman (158)(p.743)

Mechanistic insights into trait formation require a consideration of intermediate steps in the chain of causation that links genotype to phenotype. The goal of the ‘systems genetics’ approach (10, 39, 129, 171) is to achieve a mechanistic understanding of the mapping function that relates genotype to phenotype – that is, an understanding of how genetic variation is transduced through the transcriptome, proteome, and metabolome to generate phenotypes. Intermediate molecular phenotypes (‘endophenotypes’), such as transcript abundance, are also quantitative traits with genetic and environmental components of variance. The general expectation is that additive genetic variation for transcript abundance and protein activity can be related to variation in protein and metabolite abundance, which in turn can be related to organismal phenotypes (39, 116), potentially at many levels of biological organization (Fig. 1).

Using genome-wide gene expression profiles for specific tissues, it is possible to identify trait-specific transcriptional modules of co-regulated genes by measuring associations between
particular physiological traits and levels of transcript abundance. For example, studies of within-
population transcriptomic variation in *Drosophila melanogaster* have revealed that ~60% of
variable transcripts cluster into two discrete modules; levels of transcript abundance are
positively correlated within modules but negatively correlated between them (10). When flies
reared under standard growth conditions were subjected to environmental changes, the modular
organization of the transcriptome was fragmented due to co-regulated changes in sets of
environmentally responsive genes comprising ~15% of the transcriptome (10, 222). Regressions
of organismal phenotypes (resistance to starvation stress, time to recover from a chill-induced
coma, etc.) on transcript abundance implicated numerous candidate genes that formed
statistically defined transcriptional modules (10). Because transcriptional modules are typically
enriched for functionally related genes whose products interact in the same pathways, identifying
discrete modules of intercorrelated genes provides a means of constructing genetic interaction
networks – specifically, networks of causal effects that link genotype to phenotype (158).

A similar approach has been applied to an analysis of evolved population differences in
physiological performance capacities between high- and low-altitude deer mice. Relative to deer
mice that are native to low-altitude environments, high-altitude deer mice have evolved an
enhanced thermogenic capacity under hypoxia (35-37), a complex organismal performance trait
that is subject to strong directional selection in the wild (79). The enhanced thermogenic
performance of high-altitude mice is associated with upregulation of five main transcriptional
modules that influence several hierarchical steps in the O₂ transport cascade, including tissue O₂
diffusion (angiogenesis) and tissue O₂ utilization (metabolic fuel use and cellular oxidative
capacity)(35, 37). Additional transcriptional modules are associated with tissue-level traits, such
as capillary density and the relative proportion of aerobic and glycolytic fibers in skeletal
muscle, that help sustain high rates of aerobic metabolism under hypoxic conditions (122,
169)(Fig. 4). In such studies, transcripts that are significantly associated with a particular
phenotype are candidate genes for that phenotype (10, 39, 142), and can therefore be targeted for
additional experiments to test for causal effects.

Population genomics

In principle, genome-wide scans of DNA polymorphism can be used to identify candidate loci
for adaptation based on statistical signatures of positive selection. This population genomics
approach is premised on the idea that the locus-specific effects of positive selection can be detected against the genome-wide backdrop of neutral variation. Consider a genome-wide survey of DNA polymorphism in samples of individuals from a pair of populations inhabiting distinct, spatially separated environments – for example, high- and low-altitude environments. If a given trait is subject to opposing selection in the high- and low-altitude populations, then the underlying loci are expected to experience an increase in the between-population variance in allele frequency. In principle, it is possible to identify chromosomal regions that harbor such loci by exploiting theoretical predictions about the effects of spatially varying selection on patterns of DNA polymorphism at linked neutral sites. Specifically, spatially varying selection is expected to produce a peak in the between-population component of nucleotide diversity that is centered on the selected polymorphism(s), and the correlation in levels of differentiation between the selected site and linked neutral markers is expected to dissipate at a rate proportional to the genetic map distance between them (Fig. 5)(33).

In cases where a given allele has been rapidly driven up to high frequency by positive selection, the surrounding gene region will typically be characterized by a transient increase in LD. Newly arisen mutations will typically exhibit high LD with neighboring nucleotide polymorphisms since insufficient time has elapsed for recombination to randomize associations between alleles at linked sites. By contrast, the neutral expectation is that high-frequency alleles will typically exhibit low LD with neighboring polymorphisms, since the time needed for them to drift to high frequency will generally be sufficient to randomize associations between alleles at linked sites (100). For this reason, high-frequency alleles that exhibit high LD present a combination of features that are not expected under neutrality (162, 163). These features can provide a basis for identifying positively selected loci in genomic survey data, although selection on new mutations and selection on pre-existing variants (standing genetic variation) are expected to generate somewhat different patterns of LD and nucleotide variation (41, 81, 85, 139, 145, 146, 152, 193).

Genome scans to detect the effects of positive selection involve two main steps. The first step is to compute a given test statistic across the genome (for individual polymorphisms or chromosomal intervals that harbor multiple polymorphisms). The second step is to calculate whether locus-specific values of the test statistic exceed the critical value for rejecting a neutral model. In practice, locus-specific $P$-values are based on an empirical distribution of the test
statistic or a null distribution of the statistic under an assumed probablistic model. Regardless of
the choice of test statistic, and regardless of whether parametric or nonparametric methods are
used to identify outlier loci, it is important to appreciate that the effects of selection on patterns
of variation at or near causative loci depend strongly on the genetic architecture of the selected
trait, the intensity of selection on the trait, and the genetic structure of the population under
consideration (34, 94, 112, 114, 138, 165, 178, 185). In particular, theoretical and simulation
studies have demonstrated that the ability to detect the effects of selection at individual loci is
strongly affected by the dominance coefficient of the causative allele (194) and the initial
frequency of the causative allele at the onset of selection (81, 85, 139, 145, 146, 152, 193).

In principle, the population genomics approach can be used as a means of ‘outlier
detection’ to nominate candidate loci for adaptation, or it can be used to assess evidence for
selection on previously identified candidate loci by assessing whether such loci emerge as
outliers against a backdrop of genome-wide variation. Both approaches have been used
successfully to identify loci involved in evolutionary adaptation to particular environmental
conditions. In humans, for example, genome-wide surveys of DNA polymorphism in Tibetan
highlanders revealed evidence for strong and recent positive selection on several genes that can
be described as upstream regulators of the hypoxia inducible-factor (HIF) oxygen signaling
pathway (18, 24, 144, 172, 213, 216). The HIF family of transcription factors plays a key role in
regulating oxygen homeostasis by coordinating the transcriptional response to hypoxia. One of
the HIF genes that exhibited an especially clear signal of positive selection in the Tibetan
population was the EPAS1 gene (*endothelial PAS domain protein 1*) (Fig. 6), also known as
HIF2α, which encodes the oxygen-sensitive α subunit of the HIF-2 transcription factor. The
product of EPAS1 is known to play an especially important role in regulating the erythropoietic
response to hypoxia (153). It is especially remarkable that evidence for positive selection on
EPAS1 in Tibetan highlanders was obtained by several different research groups who used
different samples and who employed somewhat different analytical approaches (18, 24, 144, 172,
213, 216). Another HIF-regulatory gene that exhibited strong evidence for selection in both
Tibetan and Andean populations was EGLN1 (*Egl Nine homolog 1*) (24, 144, 172, 212, 213,
216), which encodes the prolyl hydroxylase isozyme (PHD2) that is responsible for hydroxylating
the α subunit of the HIF1 transcription factor. Results of these studies provide proof-of-principle
that the genome scan approach can successfully identify targets of recent positive selection. The
integration of such analyses with association studies and manipulative experiments can provide
additional insights into possible phenotypic targets of selection (120).

One reason for the popularity of the population genomic approach is that - unlike linkage
or association mapping approaches that use phenotypic variation as a starting point - genome
scans for selected loci are not biased by preconceived ideas about the specific traits that
contribute to adaptive evolution. Thus, when coupled with functional studies of the candidate
genes for adaptation that are ultimately identified, the genome scan approach holds the promise
of identifying genetic mechanisms of adaptation that were previously unanticipated (178).

However, there are many limitations of this approach. One major concern is that the identified
candidate loci for positive selection represent a highly biased subset of loci that have actually
contributed to past adaptation, so the implicated traits may have an entirely unrepresentative
genetic architecture. In particular, loci that have contributed to adaptive changes in highly
polygenic traits may be systematically under-represented in genomic maps of positive selection.

This is because changes in the mean value of a polygenic trait can be produced by modest shifts
in allele frequencies at large numbers of loci that have individually small effects, so the effects of
selection at any single locus may not be readily detectable (23, 149, 150, 185). This is an
important consideration since many ecologically important (fitness-related) traits typically have a
highly polygenic basis (157).

Another important issue concerns the mapping function that relates genotype to
phenotype. It is not realistic to expect that the identification of positively selected loci will
typically lead to the clear and unambiguous identification of selected phenotypes. For example,
the noncoding *EPAS1* variants that are present at an elevated frequency among Tibetan
highlanders are associated with a reduced hemoglobin concentration (18, 172, 216), which
appears to stem from a blunted erythropoietic response to hypoxia (15-17, 19, 147, 210). There
are good reasons to think that this blunted response may be physiologically beneficial under
conditions of chronic hypoxia at high altitude (147, 148, 177, 184, 199), but that does not mean
that hemoglobin concentration was necessarily the trait under direct selection. The reduced
hemoglobin concentration may represent a second-order consequence of upstream modifications
in O₂ chemosensitivity. For example, allelic variation at *EPAS1* could affect hemoglobin
concentration indirectly through changes in the hypoxic ventilatory response, which modulates
erthropoietin production through changes in arterial O₂ tension. In such cases, a combination of
experimental and analytical methods can be used to determine whether the relationship between genotype and phenotype is causal or consequential (39, 158).

Results of genome scans are useful for generating hypotheses about candidate genes that can be followed up with functional experiments (e.g., (120)). All too often, however, results of such analyses are uncritically interpreted in purely adaptive terms without any form of experimental validation. In particular, gene-ontology enrichment analyses of outlier loci often provide a basis for concocting speculative, hand-waving stories about mechanisms of adaptation. The same practice is common in genomic surveys of sequence divergence using codon-substitution models (see Molecular evolution: analysis of interspecific divergence at specific loci). Pavlidis et al. (143) confirmed the suspicions of many skeptics by conducting neutral simulations which demonstrated, firstly, that genome scans for putative targets of positive selection can have soberingly high false-positive rates (see also (12)) and, secondly, that gene ontology annotations of false-positive outliers can be used to construct seemingly plausible narratives about genetic adaptation that are empirically unsubstantiated.

Experimental analyses of physiological traits can play an important role as an adjunct to population genomic analyses by validating selection-nominated candidate genes. However, such experiments have a far more general role to play in integrated ‘phenotype-first’ studies that reveal how variation in regulatory networks gives rise to ecologically relevant trait variation under natural conditions. As described above (Systems genetics), functional genomic analyses that are integrated with experimental analyses of physiological variation can provide the foundation and starting point for investigating the mechanistic and genetic basis of natural variation in whole-animal performance (35, 37, 42, 169, 170, 208, 209).

THE ROLE OF PHYSIOLOGY IN THE FUNCTIONAL SYNTHESIS OF EVOLUTIONARY BIOLOGY

Dean and Thornton (46) advocated an integration of molecular evolution research with mechanistic studies of genes and gene products to achieve a ‘functional synthesis of evolutionary biology.’ In addition to enriching our basic understanding of the biology of phenotypic evolution, insights into the mechanistic basis of evolutionary changes in molecular function can provide insights into features of causative mutations (dominance coefficients, epistatic interactions, and pleiotropic effects) that influence fixation probabilities and can therefore help
explain why certain types of mutation make disproportionate contributions to evolutionary change (176, 186). However, insights into mechanism are equally valuable at higher levels of biological organization, suggesting a key role for mechanistic studies of phenotypic evolution that fall squarely within the purview of organismal physiology. For example, an understanding of physiological mechanisms can be used to predict and interpret the causes of genetic correlations between traits (2, 4, 57, 59). As stated by Garland and Carter (64)(p.586): ‘…an understanding of physiological mechanisms can help in determining whether a particular pattern of phenotypic variation or covariation (e.g., an allometric relationship) represents what could possibly exist or just what selection has allowed.’

Mechanistic studies at different hierarchical levels of biological organization are relevant to different links in the ‘adaptive recursion’ mentioned above. Whereas experimental approaches in molecular biology and biochemistry are needed to establish causal connections between genotype and proximal (biochemical) phenotype, genetically based changes in such phenotypes are visible to selection only to the extent that they impinge on fitness-related aspects of whole-organism performance. Thus, experimental studies of whole-animal physiology are needed to establish the causal links between proximal phenotypes (subordinate traits such as enzyme activity, pathway flux, etc.) and whole-animal physiological performance, and as described above (Physiology, performance, behavior, and fitness: measuring selection in the wild), when experiments on performance are integrated with measures of survivorship and/or reproductive success of free-ranging animals, the measurement of selection gradients can be used to establish the ensuing link between performance and fitness.

The integration of molecular biology, biochemistry, and evolutionary biology has greatly enriched our understanding of mechanism and process in the evolution of biological macromolecules (46, 77). The marriage of evolutionary biology and organismal physiology has an equally important role to play in the ‘functional synthesis,’ broadly defined.

ACKNOWLEDGMENTS
We thank G. R. Scott and three anonymous reviewers for helpful suggestions. We thank C. M. Beall, Z. A. Cheviron, and B. W. M. Wone for providing figures.

GRANTS
The authors acknowledge grant support from the National Science Foundation (NSF IOS-1121273 [TG] and IOS-1354390 [JFS]) and the National Institutes of Health (GM104397 [JTB] and HL087216 [JFS]).

DISCLOSURES
No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS
All authors contributed to the writing of the manuscript.

REFERENCES


47. **Dickerson SM, and Gore AC.** Estrogenic environmental endocrine-disrupting chemical effects on reproductive neuroendocrine function and dysfunction across the life cycle. *Reviews in Endocrine and Metabolic Disorders* 8: 143-159, 2007.


FIGURE LEGENDS

Figure 1. Path diagram of hypothetical relationships among phenotypes at multiple levels of biological organization, ranging from fundamental morphological, physiological, and biochemical traits to Darwinian fitness. Following the standard conventions of path analysis, single-headed arrows indicate putatively causal relationships, whereas double-headed arrows indicate correlations between traits. Not shown are possible direct effects of subordinate traits on behavior or fitness components. For example, individual variation in circulating hormone concentrations might directly affect behavior, growth rates, sexual maturation, etc. Arnold (5) first introduced the formalities of path analysis for examining the adaptive significance of inter-individual trait variation, and it quickly became a paradigm in ecological and evolutionary physiology. The depiction here builds on a number of subsequent papers that have expanded Arnold's original presentation through both conceptual and empirical approaches (1, 28, 64, 68, 106, 140, 173, 201). Many of these have included aspects of locomotor performance because it is easy to envision how traits such as speed or stamina fit into the scheme.

Figure 2. Direct and correlated responses to selection for increased mass-independent maximal O₂ consumption (V̇O₂max) in replicate lines of house mice after eight generations. Metabolic rates represent means (± SEM) of four replicate lines per treatment. Four lines of mice were selectively bred for increased mass-independent metabolic rates (high-MMR). Another four lines were antagonistically selected for a combination of increased mass-independent metabolic rate and low mass-independent basal metabolic rate, BMR (antag-MR). The remaining four lines were maintained as non-selected controls. A) Response to directional selection for increased mass-independent maximal metabolic rate (MMR) over eight generations. Directional selection for increased MMR was imposed on high-MMR and antag-MR lines (no selection on control lines). B) Changes in mean mass-independent BMR over eight generations. Directional selection for decreased mass-independent BMR was imposed on antag-MR lines (no selection on mass-independent BMR in high-MMR or control lines).
Figure 3. Phylogeny of vertebrate steroid receptors and the evolution of hormone specificity. The most ancient steroid receptor, AncSR1, and estrogen receptors (pink) are activated by aromatized steroids, and AncSR2 and its descendants (blue) are activated by non-aromatized hormones. The black box denotes the interval in which two large-effect substitutions occurred, producing an evolutionary shift in hormone recognition.

Figure 4. Transcriptional modules of co-regulated genes are associated with muscle phenotype in deer mice. A) Histological analyses revealed that the gastrocnemius muscle of high-altitude deer mice has a higher capillary density (top row) and a higher areal density of oxidative fibers (bottom row) relative to that of low-altitude mice. B) Left: Transcripts that are differentially expressed between high- and low-altitude populations of deer mice ($n=68$ transcripts) cluster into two discrete modules (P1 and P2). Right: Transcripts that are differentially expressed between mice in different rearing environments (wild-caught mice sampled at their native elevations vs. F1 lab-reared mice, $n=658$ transcripts) cluster into five modules (T1-T5). C) Left: Changes in expression of the P2 module (50 transcripts) are significantly associated with evolved (non-plastic) differences in muscle capillary density between high- and low-altitude deer mice ($r^2 = 0.46, p = 0.001$). Right: Changes in expression of the T5 module (224 transcripts) are associated with environmentally induced (plastic) changes in the density of oxidative fibers in the gastrocnemius ($r^2 = 0.30, p = 0.007$). For each transcriptional module, an index of overall expression is provided by PC1 scores, where PC1 is the first axis of a principal components analysis on the correlation matrix of transcript abundance for all genes in a given module. In both panels, data points for high- and low-altitude mice are denoted as filled and open symbols, respectively. Based on data from (169).

Figure 5. Simulation results showing the effects of spatially varying selection on levels of differentiation at linked sites in a pair of populations at migration-selection equilibrium (figure modified from (33)). The Y-axis shows levels of differentiation in site-specific allele frequencies (as measured by $F_{ST}$), and the X-axis shows a measure of genetic map distance (as measured by Morgans, the distance over which the expected average number of cross-over events per generation is 0.1). Local selection was modeled as directional selection that acts in opposite directions in two populations that exchange migrants. Heterozygotes and homozygotes at the
selected locus were assigned fitnesses of $1 - (s/2)$ and $1 - s$, respectively (where $s$ is the selection coefficient); $s = 0.5$ under ‘strong’ selection and $s = 0.1$ under ‘moderate’ selection. In the absence of local selection ($s = 0$), $F_{ST}$ at linked markers is not correlated with map distance from the selected polymorphism. In the presence of local selection, levels of differentiation are especially high at sites that are closely linked to the selected site, and the rate of decay is inversely proportional to the strength of local selection. In comparisons between populations that are locally adapted to contrasting environments (e.g., high- and low-altitudes), loci that have contributed to adaptive changes in phenotype would be expected to exhibit especially high $F_{ST}$ values relative to the genome-wide average.

**Figure 6.** A genome-wide scan of allelic differentiation between population samples of Tibetans (resident at 3,200-3,500 m in Yunnan Province, China) and Han Chinese. The vertical axis of the graph shows the negative log of site-specific $P$-values for allele frequency differences between the Tibetan and Han Chinese population samples (low $P$-values denote allele frequency differences that are too large to explain by genetic drift). The horizontal axis of the graph shows the genomic positions of each assayed nucleotide site, arranged by chromosome number. The red line indicates the threshold for genome-wide statistical significance ($P = 5 \times 10^{-7}$). Values are shown after correction for background population stratification using an intragenomic control. Several non-coding variants flanking the *EPAS1* gene are highly significant outliers. Reprinted from (18).
A. Maximal metabolic rate, MMR (ml O₂ min⁻¹)
- Control
- Antagonistically selected MR
- High-selected MMR

B. Basal metabolic rate, BMR (ml O₂ min⁻¹)
- Control
- Antagonistically selected MR
- High-selected MMR
Estrogen Receptors

Estrogen Receptors

Androgen Receptors

Progesterone Receptors

Glucocorticoid Receptors

Mineralocorticoid Receptors

Estrogen-related Receptors

Aromatized A-ring

Non-Aromatized
A.)

Capillaries

Highland

Lowland

Oxidative fibers

B.)

Correlation (r)

P1

P2

T1

T2

T3

T4

T5

C.)

Residual capillary surface density (µm⁻¹)

Module expression (PC1)

Residual areal density of ox. fibers

P2

T5