The frailty of adaptive hypotheses for the origins of organismal complexity

Michael Lynch*

Department of Biology, Indiana University, Bloomington, IN 47405

The vast majority of biologists engaged in evolutionary studies interpret virtually every aspect of biodiversity in adaptive terms. This narrow view of evolution has become untenable in light of recent observations from genomic sequencing and population-genetic theory. Numerous aspects of genomic architecture, gene structure, and developmental pathways are difficult to explain without invoking the nonadaptive forces of genetic drift and mutation. In addition, emergent biological features such as complexity, modularity, and evolvability, all of which are current targets of considerable speculation, may be nothing more than indirect by-products of processes operating at lower levels of organization. These issues are examined in the context of the view that the origins of many aspects of biological diversity, from gene-structural embellishments to novelties at the phenotypic level, have roots in nonadaptive processes, with the population-genetic environment imposing strong directionality on the paths that are open to evolutionary exploitation.

adaptation | genome evolution | evolvability | modularity | population genetics

Although biologists have always been concerned with complex phenotypes, the matter has recently become the subject of heightened speculation, as a broad array of academics, from nearly every branch of science other than evolutionary biology itself, claim to be in possession of novel insights into the evolution of complexity. The claims are often spectacular. For example, Kirschner and Gerhart (1) argue that evolutionary biology has been "woefully inadequate" with respect to understanding the origins of complexity and promise "an original solution to the long-standing puzzle of how small random genetic change can be converted into complex, useful innovations." However, this book and many others like it (e.g., refs. 2–5) provide few references to work done by evolutionary biologists, making it difficult to understand the perceived areas of inadequacy, and many of the ideas promoted are known to be wrong, making it difficult to appreciate the novelty. Have evolutionary biologists developed a giant blind spot; are scientists from outside of the field reinventing a lot of bad wheels; or both?

Evolutionary biology is treated unlike any science by both academics and the general public. For the average person, evolution is equivalent to natural selection, and because the concept of selection is easy to grasp, a reasonable understanding of comparative biology is often taken to be a license for evolutionary speculation. It has long been known that natural selection is just one of several mechanisms of evolutionary change, but the myth that all of evolution can be explained by adaptation continues to be perpetuated by our continued homage to Darwin's treatise (6) in the popular literature. For example, Dawkins' (7–9) agenda to spread the word on the awesome power of natural selection has been quite successful, but it has come at the expense of reference to any other mechanisms, a view that is in some ways profoundly misleading. There is, of course, a substantial difference between the popular literature and the knowledge base that has grown from a century of evolutionary research, but this distinction is often missed by nonevolutionary biologists.

The goal here is to dispel a number of myths regarding the evolution of organismal complexity (Table 1). Given that life originated from inorganic matter, it is clear that there has been an increase in phenotypic complexity over the past 3.5 billion years, although long-term stasis has been the predominant pattern in most lineages. What is in question is whether natural selection is a necessary or sufficient force to explain the emergence of the genomic and cellular features central to the building of complex organisms.

Nothing in Evolution Makes Sense Except in Light of Population Genetics

Although the basic theoretical foundation for understanding the mechanisms of evolution, the field of population genetics, has long been in place, the central significance of this framework is still occasionally questioned, as exemplified in this quote from Carroll (4), "Since the Modern Synthesis, most expositions of the evolutionary process have focused on microevolutionary mechanisms. Millions of biology students have been taught the view (from population genetics) that 'evolution is change in gene frequencies.' Isn't that an inspiring theme? This view forces the explanation toward mathematics and abstract descriptions of genes, and away from butterflies and zebras. . . The evolution of form is the main drama of life's story, both as found in the fossil record and in the diversity of living species. So, let's teach that story. Instead of 'change in gene frequencies,' let's try 'evolution of form is change in development'." Even ignoring the fact that most species are unicellular and differentiated mainly by metabolic features, this statement illustrates two fundamental misunderstandings. Evolutionary biology is not a story-telling exercise, and the goal of population genetics is not to be inspiring, but to be explanatory. The roots of this contention are fourfold.

First, evolution is a population-genetic process governed by four fundamental forces. Darwin (6) articulated one of those forces, the process of natural selection, for which an elaborate theory in terms of genotype frequencies now exists (10, 11). The remaining three evolutionary forces are nonadaptive in the sense that they are not a function of the fitness properties of individuals: mutation is the ultimate source of variation on which natural selection acts, recombination assorts variation within and among chromosomes, and genetic drift ensures that gene frequencies will deviate a bit from generation to generation independent of other forces. Given the century of work devoted to the study of evolution, it is reasonable to conclude that these four broad classes encompass all of the fundamental forces of evolution.

Second, all four major forces play a substantial role in genomic evolution. It is impossible to understand evolution purely in terms of natural selection, and many aspects of genomic, cellular, and developmental evolution can only be understood by invoking a negligible level of adaptive involvement (12, 13). Because all three nonadaptive forces of evolution are stochastic in nature, this

Table 1. A summary of some common misconceptions about evolution and complexity, and contrasting views

Myth	Reality
1. Evolution is natural selection.	Natural selection is just one of four primary evolutionary forces.
2. Characterization of interspecific differences at the molecular and/or cellular levels is tantamount to identifying the mechanisms of evolution.	The resources deployed in evolutionary change reside at the molecular level, but whereas the cataloging of such differences at the interspecific level identifies the end products of evolution, it does not reveal the population-genetic processes that promoted such change.
3. Microevolutionary theory based on gene-frequency change is incapable of explaining the evolution of complex phenotypes.	No principle of population genetics has been overturned by an observation in molecular, cellular, or developmental biology, nor has any novel mechanism of evolution been revealed by such fields.
4. Natural selection promotes the evolution of organismal complexity.	There is no evidence at any level of biological organization that natural selection is a directional force encouraging complexity. In contrast, substantial evidence exists that a reduction in the efficiency of selection drives the evolution of genomic complexity.
5. Natural selection is the only force capable of promoting directional evolution.	Both mutation and gene conversion are nonrandom processes that can drive the patterning of genomic evolution in populations with sufficiently small effective sizes (common in multicellular lineages).
6. Genetic drift is a random process that leads to noise in the evolutionary process, but otherwise leaves expected evolutionary trajectories unaltered.	By reducing the efficiency of selection, random genetic drift imposes a high degree of directionality on evolution by increasing the likelihood of fixation of deleterious mutations and decreasing that of beneficial mutations.
7. Mutation merely creates variation, whereas natural selection promotes specific mutant alleles on the basis of their phenotypic effects.	Mutation operates as a weak selective force by differentially eliminating alleles with structural features that magnify mutational target sizes.
8. Phenotypic and genetic modularity are direct products of natural selection.	There is no evidence that the modular structure of gene regulatory regions or genetic networks is directly advanced by selective mechanisms. However, the processes of duplication, degenerative mutation, and random genetic drift can lead to the passive emergence of modularity in populations of with genetic effective sizes of the magnitude found in multicellular species.
9. Natural selection promotes the ability to evolve.	There is no evidence that phylogenetic variation in the pathways open to evolutionary exploration is anything more than a by-product of physical processes that passively arise with expansions in genome size and generation length. There are no abrupt transitions in aspects of genomic architecture or gene structure between unicellular and multicellular species, nor between viruses, prokaryotes, and eukaryotes.

conclusion raises some significant technical challenges. It is tempting to think that stochastic processes have no implications for the direction of evolution. However, the effects of mutation and recombination are nonrandom, and by magnifying the role of chance, genetic drift indirectly imposes directionality on evolution by encouraging the fixation of mildly deleterious mutations and discouraging the promotion of beneficial mutations.

Third, the field of population genetics is now so well supported at the empirical level that the litmus test for any evolutionary hypothesis must be its consistency with fundamental population-genetic principles. Grounded in basic Mendelian processes and sampling theory, many of these principles were laid down before the elucidation of the structure of DNA. Shortly after the genetic code was cracked, a series of technological breakthroughs advanced our ability to reveal molecular variation: protein sequencing in the 1950s, surveys of protein variants in the 1960s, ribosomal RNA sequencing in the 1970s, gene sequencing in the 1980s, and whole-genome sequencing in the 1990s. Each of these episodes brought the need for new methods for analysis and interpretation, and in each case the framework was drawn largely from preexisting population-genetic theory. Thus, although we do not yet fully understand the connections between evolution at the molecular and phenotypic levels, we can be confident that the machinery to do so is in place.

Fourth, some attempts to marginalize the contributions of population genetics to our understanding of evolution have pointed to the "bean bag" genetics debate that occurred in the middle of the last century (see ref. 14). However, this is a misunderstanding, as the tensions during this period were not about the population-genetic basis of evolutionary change, but about the need to incorporate epistasis into the existing framework, something that population geneticists have now invested heavily in (15, 16). From the standpoint of its phenotypic products, evolution is more than a change in gene frequencies. Organisms are more than the sum of their parts, just as genes are more than the sum of their functional components. But if we are concerned with the process of evolutionary change, then evolution is indeed a change in genotype frequencies.

In summary, population genetics provides an essential framework for understanding how evolution occurs, grounding us in reality by clarifying the pathways that are open to evolutionary exploitation. To quote Carroll (4) again, "Simplification may indeed be necessary for news articles, but it can distort the more complex and subtle realities of evolutionary patterns and mechanisms."

Internal Versus External Evolutionary Forces

The literature is permeated with dogmatic statements that natural selection is the only guiding force of evolution, with mutation creating variation but never controlling the ultimate direction of evolutionary change (for a review, see ref. 17). This view derives from two types of arguments. First, hundreds of artificial selection experiments have generated changes in mean phenotypes well beyond the observed range in the base population in just a few dozen generations (18), inspiring the view that
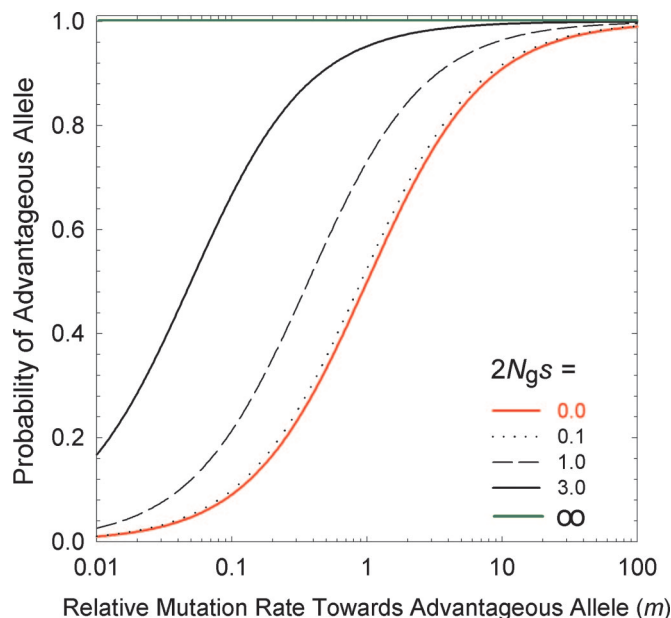
Fig. 1. The long-term probability that an allele residing at a biallelic locus will be of the selectively advantageous type, given a selective advantage s, an effective population number of gene copies of N_g, and a mutation rate to the beneficial allele m times that in the reverse direction. The red line ($2N_gs = 0.0$) denotes neutrality, whereas the green line ($2N_gs = \infty$) denotes an effectively infinite population, such that genetic drift is a negligible force.

quantitative variation is distributed over an effectively infinite number of loci with minuscule effects (19–21). Second, much of the earliest work in theoretical population genetics downplayed the ability of mutation to overcome the force of selection (22, 23). Both arguments ignore significant complications that arise in finite populations, and it is now known that genome composition is governed by biases in mutation and gene conversion, some of which (e.g., mobile-element proliferation) operate via internal drive-like mechanisms (13).

The notion that mutation pressure can be a driving force in evolution is not new (6, 24–31), and the conditions that must be fulfilled if mutation is to alter the direction of evolution relative to adaptive expectations are readily derived. Consider two alternative states at a locus, **A** and **a**, with the mutation rate of **a** → **A** being m times that of **A** → **a**, but with type **A** having a fractional selective advantage s over type **a**. Further progress requires that we specify the effective number of gene copies per locus at the population level. This quantity, N_g, which is equivalent to the effective size of a haploid population and approximately twice that for an outcrossing diploid species, is influenced by many factors, including the breeding system, temporal fluctuations in population size, and the level of recombination (which influences the sensitivity of a locus to spurious hitch-hiking effects), and is generally orders of magnitude smaller than the absolute number of reproductive adults in a population (13). With these definitions in hand, standard theory (12) shows that the fixation probability of a mutation to **A** is e^S times that for a mutation to **a**, where $S = 2N_gs$ is twice the ratio of the power of selection to the power of random genetic drift ($1/N_g$). Because the population-level rate of transition from one allelic type to another is equal to the product of the mutation rate and the fixation rate, the ratio of probabilities of being in states **A** versus **a** at selection-mutation-drift equilibrium is simply me^S (Fig. 1). This simple expression leads to two general conclusions: (*i*) regardless of the strength of selection, if $1/N_g \gg |s|$, the population will be driven to a state expected under mutation pressure alone; and (*ii*) the equilibrium composition of a population depends not on the

absolute power of mutation, but on the relative rates of forward and reverse mutations (the mutation bias).

The Passive Emergence of Genome Complexity by Nonadaptive Processes

Most biologists are so convinced that all aspects of biodiversity arise from adaptive processes that virtually no attention is given to the null hypothesis of neutral evolution, despite the availability of methods to do so (32–34). Such religious adherence to the adaptationist paradigm has been criticized as being devoid of intellectual merit (35), although the field of molecular evolution has long been obsessed with potential for the "nearly neutral" accumulation of very slightly deleterious mutations (36–38). The condition for near-neutrality is fulfilled when the ratio of the powers of selection and drift is substantially <1, i.e., $|2N_gs| \ll 1$.

This simple relationship has considerable utility in attempts to evaluate the potential role of nonadaptive mechanisms in the evolution of genomic architecture (13, 39, 40). Striking phylogenetic variation exists at the level of gene and genomic architecture. The genomes of multicellular eukaryotes are invariably packed with mobile elements, and individual genes are generally subdivided by multiple introns, harbor multiple transcription-factor binding sites, and are transcribed into units containing substantial untranslated flanking sequences. In contrast, prokaryotic genomes are usually nearly completely devoid of mobile elements and introns and have genes with very simple regulatory structures, often transcribed into polycistronic units (operons) with negligible leader and trailer sequences. Most unicellular eukaryotic genomes exhibit structural features on a continuum between these two extremes.

Understanding the origins of eukaryotic genome complexity in adaptive terms is rendered difficult by the fact that each embellishment added to a gene magnifies its vulnerability to mutational inactivation, thereby encouraging its elimination from the population (13, 40). If a particular embellishment requires that n nucleotides be conserved for proper gene function, the mutational disadvantage is nu, where u is the mutation rate per nucleotide site. Each intron added to a protein-coding gene requires that $n \approx 30$ nucleotide sites within the intron and adjacent exons be reserved for proper spliceosomal recognition (41). Transcription-factor binding sites are ~10 bp in length (e.g., ref. 42), so the addition of each such site to a gene increases the degenerative mutation rate by $\approx 10u$. The average eukaryotic 5′ UTR increases the degenerative mutation rate by ~$4u$ by providing substrate for mutations to premature translation-initiation codons (43). Even completely nonfunctional intergenic DNA is a mutational hazard because it serves as a substrate for inappropriate gain-of-function mutations (44).

Numerous observations suggest that the amount of regulatory DNA associated with the average gene in a multicellular species is at least as great as the length of the coding region, and such genes typically also contain ~5–7 introns (13). Thus, the average mutational target sizes of genes in multicellular eukaryotes are more than two to three times those for prokaryotes. To become established, the modifications that led to such mutational hazards must have either had a substantial immediate selective advantage or arisen in populations with effective sizes sufficiently small to render them immune to selection. Letting $s = nu$, the latter condition requires that $2N_gnu < 1$, and because the mutational cost of individual modifications is small, generally <$30u$, it is difficult to reject the hypothesis that incremental expansions of eukaryotic gene complexity were largely driven by nonadaptive processes.

More formal justification for this claim derives from molecular population-genetic data. Assuming the silent sites of protein-coding genes to be effectively neutral, the average number of silent-site substitutions between randomly sampled nucleotide sites within a population is a function of the rates of input of new variation by mutation ($2u$) and loss by genetic drift ($1/N_g$). At mutation-drift equilibrium, the expected level of silent-site divergence within a population is equal to the ratio of these rates, $2N_gu$. Average

Table 2. Aspects of gene and genomic architectural evolution that appear to be explainable only after accounting for variation in the relative power of nonadaptive evolutionary forces

Genomic streamlining in microbial species versus genome bloating in multicellular lineages.

Nucleotide composition variation within and among genomes, genomewide A/T composition, strand asymmetry, isochores, and codon-usage bias in unicellular species.

Differential proliferation of mobile elements in unicellular versus multicellular species.

Gene number: preservation of duplicate genes by degenerative mutations (subfunctionalization).

Origin of the spliceosome by subfunctionalization and proliferation of introns in lineages of multicellular species.

Expansion of UTRs of the messenger RNAs of eukaryotes.

Origin of modular regulatory regions in eukaryotic genes.

Demise of operons in eukaryotes.

Variation in organelle genome architecture: lean in animals; bloated in land plants.

Messenger RNA editing in plant organelle genomes.

Restriction of sex chromosomes to multicellular lineages.

---

estimates of $2N_g u$ are 0.104 for prokaryotes, 0.057 for unicellular eukaryotes, 0.026 for invertebrates, 0.015 for land plants, and 0.004 for vertebrates (40). Because some weak selective forces may operate on silent sites, the average estimates of $2N_g u$ for unicellular species (with high N_g) are almost certainly downwardly biased. Thus, rearranging the criterion for effective neutrality to $2N_g u \ll 1/n$, an embellishment that increases the mutational target size of a vertebrate by $n < 250$ will be largely immune from selection, and hence free to drift to fixation, whereas the critical value of n for a prokaryote is $\ll 10$.

These observations indicate that the paths open to evolutionary exploration are fundamentally different between unicellular and multicellular species for reasons completely unassociated with organismal complexity. Because of their relatively small N_g, multicellular species are expected to accumulate gratuitous gene-structural changes without any direct selection for them and to become laden with other deleterious genomic features (13, 39). In contrast, unicellular lineages are expected to maintain streamlined genomes, not necessarily for metabolic reasons, but because of the exceptionally high efficiency of selection opposing the accumulation of mutationally hazardous DNA.

The hypothesis that expansions in the complexity of genomic architecture are largely driven by nonadaptive evolutionary forces is capable of explaining a wide range of previously disconnected observations (13, 40) (Table 2). This theory may be viewed as overly simplistic. However, simply making the counterclaim that natural selection is all powerful (without any direct evidence) is not much different from invoking an intelligent designer (without any direct evidence). If a successful adaptive counterargument is to be mounted, simpler nonadaptive models must be shown to be inadequate, and to accomplish that, something must be known about the expected pattern of evolution in the absence of selection. If nothing else, the ideas presented above provide the basis for a null model for genomic evolution. Certainly, many of the above-mentioned embellishments of eukaryotic genes have adaptive functions in today's multicellular species, but observations on current deployment may have little bearing on matters of initial origins. Most of the repatterning of the genomic real estate of eukaryotes occurred before the evolution of multicellularity (13).

Are the Origins of Organismal Complexity Also Rooted in Nonadaptive Processes?

Multicellularity is widely viewed as a unique attribute of eukaryotes, somehow made possible by the origin of a more complex cellular architecture and, without question, with the assistance of natural selection. However, it is difficult to defend this assertion in any formal way. Complex, multicellularity has only arisen twice, once in animals and once in vascular plants. One might add fungi to the list, although the number of fungal cell types is not large, and there is some question as to whether multicellularity was ancestral to the phylogenetic group that contains animals, fungi, and slime molds. In any event, the probability that two or three origins of multicellularity simply arose by chance within eukaryotes as opposed to prokaryotes is somewhere on the order of 1/4 to 1/2, well below the general standards of statistical validity. Of course, many other eukaryotes are capable of producing a few different cell types, but the same is true for prokaryotes, some of which produce radically different cell morphologies.

Nevertheless, King (45) states that "this historical predisposition of eukaryotes to the unicellular lifestyle begs the question of what selective advantages might have been conferred by the transition to multicellularity;" and Jacob (46) argues that "it is natural selection that gives direction to changes, orients chance, and slowly, progressively produces more complex structures, new organs, and new species." The vast majority of biologists almost certainly agree with such statements. But where is the direct supportive evidence for the assumption that complexity is rooted in adaptive processes? No existing observations support such a claim, and given the massive global dominance of unicellular species over multicellular eukaryotes, both in terms of species richness and numbers of individuals, if there is an advantage of organismal complexity, one can only marvel at the inability of natural selection to promote it. Multicellular species experience reduced population sizes, reduced recombination rates, and increased deleterious mutation rates, all of which diminish the efficiency of selection (13). It may be no coincidence that such species also have substantially higher extinction rates than do unicellular taxa (47, 48).

Although some aspects of the roots of the cellular interactions that constitute development are likely to reside in the resolution of adaptive conflicts between the advantages of cell–cell cooperation versus going it alone (e.g., refs. 49 and 50), it need not follow that natural selection is a sufficient force for the exit from the unicellular world. Many developmental genes previously thought to have originated in the vertebrate lineage, owing to their absence in arthropods and nematodes, are now known to be present in basal lineages of animals lacking mesoderm (the cnidarians), and by inference must have simply been lost from various invertebrate phyla (51). Numerous examples of morphological simplification exist in animals (e.g., limb loss in lizards and salamanders, coelom loss in nematodes, and mouth and anal loss in hydrothermal-vent worms), and a plausible, albeit controversial, case has even been made that prokaryotic cell architecture is a simplified derivative of that of eukaryotes (52).

Could nonadaptive processes have played a role in the evolution of something as intricate as cell architecture or developmental complexity? There are at least two ways by which such a cascade of events might be precipitated. First, intrinsically deleterious genome-level changes, such as those resulting from intron and mobile-element proliferation, must impose selection pressures for cellular defense mechanisms. It has been argued that by imposing a need to process mRNAs before their exposure to ribosomes, the establishment of spliceosomal introns provided the evolutionary pressure that led to the origin of the nuclear membrane (53, 54), and Koonin (55) has further suggested that the nonsense-mediated decay (NMD) and ubiquitin signaling pathways evolved as secondary mechanisms for minimizing the accumulation of aberrant transcripts and proteins resulting from splicing errors. This line of thinking could be taken in a number of additional directions. For example, the assembly of spliceosomal subunits occurs within intranuclear Cajal bodies (56), and aberrant transcripts flagged by the NMD pathway are degraded in cytoplasmic P bodies (57). The

nature of cause and effect in these relationships is difficult to resolve, as all hypothetical lines of defense against introns appear to have been present in the stem eukaryote (58), raising the possibility that the colonization of nuclear genes by introns followed the origin of permissive cellular features, rather than the other way around. Nevertheless, the idea that internal constraints played a role in cellular evolution is secure under either scenario.

Second, because cellular and developmental features reflect the transformation of gene-level information into gene expression, the opportunities for phenotypic evolution must ultimately be constrained by the physical resources existing at the genomic level, which as noted above are strongly influenced by nonadaptive aspects of the population-genetic environment. Reductions in N_g are expected to lead to increases in intron number and size, expansions in UTR lengths, losses of operons, the modularization of regulatory regions, and the preservation of duplicate genes by subfunctionalization (among other things; Table 2). Thus, as will be discussed more fully below, to the extent that an increase in gene-architectural complexity is a precondition for the emergence of greater complexity at the organismal level (including the hallmarks of multicellularity: multiple cell types, complex patterns of gene expression, and mechanisms of cell signaling), a long-term synergism may exist between nonadaptive evolution at the DNA level and adaptive evolution at the phenotypic level. There is no need to abandon the idea that many of the external morphological and/or behavioral manifestations of multicellularity in today's organisms are adaptive. However, if the view promoted above is correct, the relatively simple phenotypes of the Earth's smallest organisms is not an inevitable outcome of the prokaryotic body plan nor a reflection of selection for metabolic efficiency, but an indirect consequence of the barrier to the passive emergence of genomic complexity imposed by high N_g.

In summary, the near-complete absence of the concept of nonadaptive processes from the lexicon of those concerned with cellular and developmental evolution does not reflect any formal demonstration of the negligible contribution of such mechanisms, and indeed, there is no fundamental reason why development should be uniquely immune to nonadaptive evolutionary forces. One could even argue that the stringency of natural selection is reduced in complex organisms with behavioral and/or growth-form flexibilities that allow individuals to match their phenotypic capabilities to the local environment. Some of these shortcomings have recently attracted attention, and a scaffold for connecting evolutionary genetics, genomics, and developmental biology is slowly beginning to emerge (59–66).

The Passive Emergence of Modular Gene Interactions

King and Wilson (67) inspired the view that modularity and repatterning of regulatory-element utilization are the central determinants of the evolution of organismal complexity (5, 68). Although this view is not universally accepted even among developmental biologists (69), because development always involves cross-talk between gene products, one must start with a consideration of the origins of the mechanisms that allow such transactions to take place. There is no evidence that gene-regulatory modules associated with complex functions arise as *de novo* integrated units, although some biologists seem to feel otherwise (70). Rather, like all aspects of evolution, the origins of changes in genetic pathways must be a function of descent with modification. Mutant alleles arise independently at individual loci, with features being defined by prior historical contingencies. Thus, although the idea that regulatory modules with functional significance in today's organisms can only have arisen via natural selection is seductive, it remains to be determined how the stepwise alterations necessary for the construction of genetic pathways come about. The following is a preliminary outline of one approach to understanding the evolution of organismal

complexity via the development of a formal population-genetic framework for the emergence of alternative forms of gene interactions.

Although compelling adaptive arguments are always accompanied by a formal rejection of simpler nonadaptive hypotheses, credible null hypotheses have rarely been pursued in evolutionary developmental biology. This concern is not trivial, as it has been shown that modular gene-regulatory structures (with unique transcription factors governing expression in different spatial/temporal contexts) can emerge passively, without any direct selection for modularity *per se*, starting from an initial state in which the entire expression breadth of the gene is under unified control (66). Under the scenario outlined in the top two tiers of Fig. 2, during the entire transition to a modular form of gene regulation, the spatial/temporal pattern of gene expression remains constant, with only the underlying molecular mechanisms for achieving this fixed pattern being modified. Thus, the new regulatory architecture emerges beneath a constant phenotype, without any bottleneck in fitness during the transitional phase of mixed genotypes. Such neutral transitions may help explain apparent cases of "developmental system drift," whereby closely related species achieved similar morphological structures by substantially different mechanisms (59, 63, 71–74).

The emergence of modular gene structure by the nonadaptive processes of duplication, degenerative mutation, and genetic drift is fully compatible with the known magnitudes of these forces in multicellular species (13). For example, the rate of duplication of entire genes is ~1% per gene per million years, and because small fragments of DNA are tandemly duplicated at much higher rates than entire genes (75), variation in the regulatory modules of genes must arise at least as rapidly as single-nucleotide polymorphisms. However, because of the mutational cost of allelic complexity, the likelihood of completion of semineutral modularization processes becomes negligible once $1/N_g$ becomes smaller than the excess mutational burden (66). Thus, contrary to popular belief, natural selection may not only be an insufficient mechanism for the origin of genetic modularity, but population-genetic environments that maximize the efficiency of natural selection may actually promote the opposite situation, alleles under unified transcriptional control. Under this view, the reductions in N_g that likely accompanied both the origin of eukaryotes and the emergence of the animal and land-plant lineages may have played pivotal roles in the origin of modular gene architectures on which further developmental complexity was built.

Despite the initial invariance of phenotypic expression patterns during this type of gene-architectural repatterning, the emergence of independently mutable subfunctions in modularized alleles can contribute to adaptive evolution in significant ways. For example, if the ancestral allele under unified control was subject to pleiotropic constraints associated with shared regulatory regions, modularization may open up previously inaccessible evolutionary pathways. Relief from pleiotropy can be even further facilitated following the duplication of the entire gene (bottom of Fig. 2), as complementary degenerative mutations partition cellular tasks among paralogous copies (76). This process of subfunctionalization is known to be a frequent fate of duplicate genes in multicellular species (13, 77), and theory suggests that it too is most likely to occur in populations with small N_g, again because of the mutational burden of distributing a fixed number of subfunctions over multiple genes (78). Thus, the joint operation of both processes (the emergence of gene subfunctions and their subsequent partitioning among paralogs) in the small to moderate population-size environment that exists in multicellular species provides a powerful mechanism for the passive remodeling of entire developmental genetic pathways (13).

Another peculiar aspect of developmental pathways that has defied explanation is their seemingly baroque structure (79, 80). It is common for linear pathways to consist of a series of genes whose products are essential to the activation/deactivation of the next
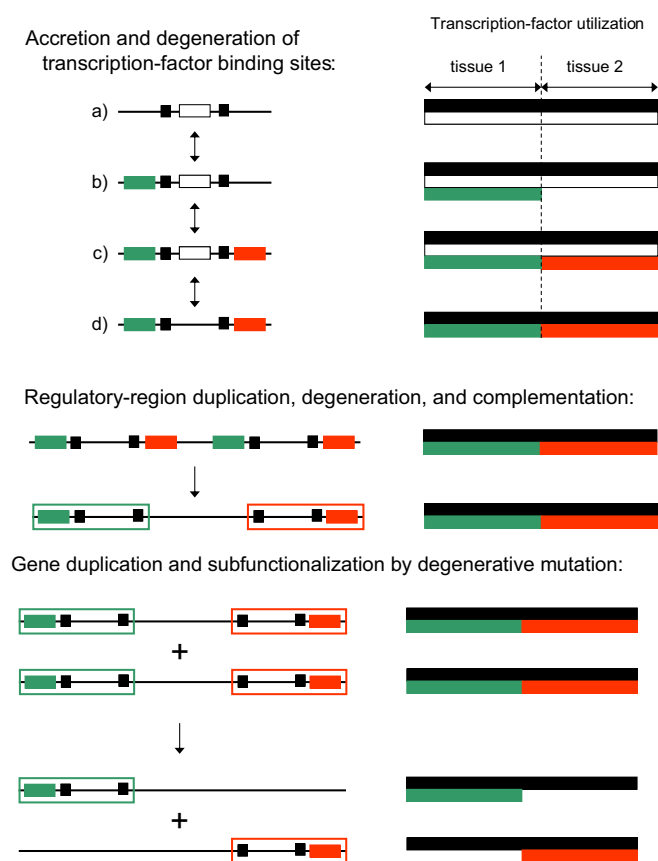
Fig. 2. The passive emergence of specialized gene functions via nonadaptive processes of duplication, degenerative mutation, and random genetic drift. (*Left*) Regulatory elements (transcription-factor binding sites) are depicted, with each regulatory element color-coded according to the transcription factor that binds to it. (*Right*) Allele-specific utilizations of transcription factors are depicted. Transcription factors denoted by black and white are ubiquitously expressed, whereas those denoted by green and red are each expressed in single, nonoverlapping tissues. For this particular gene, within their respective tissues, the green and red transcription factors are redundant with respect to the white factor, but the additional black factors are essential for complete expression. Three hypothetical phases of gene architectural modification are shown. (*Top*) Accretion and degeneration of transcription-factor binding sites. The initial allele (a) is expressed in an identical manner in both tissues, but the regulatory region sequentially acquires the green and red elements. The redundant white element is then vulnerable to loss by degenerative mutation, yielding a descendant allele with a semiindependent mode of expression, as the black element is still essential to expression in both tissues. At this stage all four alleles (a–d) are interchangeable, as each of them achieves the same pattern of phenotypic expression. (*Middle*) Regulatory-region duplication, degeneration, and complementation. The entire enhancer region is tandemly duplicated, with each component then losing a complementary (red/green) element. The resultant allele has become modularized in the sense that it harbors two independently mutable subfunctions denoted by the green and red open boxes; a mutation in either region has effects confined to a single tissue. (*Bottom*) Gene duplication and subfunctionalization by degenerative mutation. The entire gene is duplicated, with each copy becoming silenced by degenerative mutation for a complementary subfunction. The expression of each copy is now confined to a single tissue.

downstream member, with only the expression of the final component in the series having an immediate phenotypic effect. For example, the product of gene D may be necessary to turn on gene C, whose product turns on gene B, whose product finally turns on gene A, which carries out an essential organismal function. Pathways involving only inhibitory steps also exist, and these lead to an alternating series of high and low expression, depending on the state of the first gene in the pathway. It is often unclear whether such
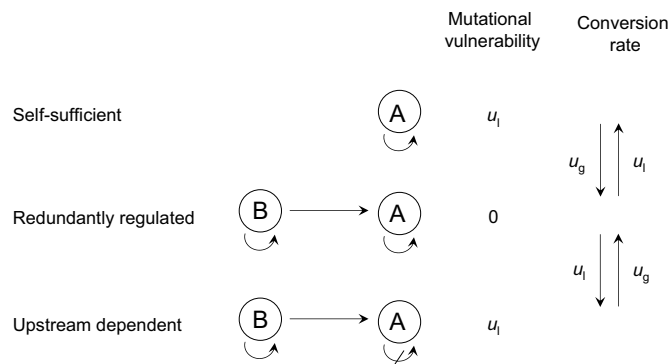
Fig. 3. A series of allelic states for locus A, defined by the ability to self-express and/or be activated by an upstream transcription factor B. Mutational rates of gain and loss of regulatory abilities are denoted by u_g and u_l, here for simplicity assumed to be the same for both self-activation and upstream activation. The redundantly regulated allele is invulnerable to single-loss mutations.

complexity has any advantages over the simple constitutive expression or self-regulation of the final member of the pathway.

In principle, pathway augmentation may be driven entirely by the nonadaptive processes of duplication, degeneration, and genetic drift. Consider the series of regulatory states for gene A in Fig. 3. In the simplest case, A carries out some function in a self-sufficient fashion, but in a series of steps, it can become completely reliant on upstream activation by transcription factor B. A scenario like this could unfold in the following way. Initially, A becomes sensitive to activation by B, either because gene A has acquired a binding site for factor B, or because factor B acquires a fortuitous mutation that causes it to serve as an activator of A. At this point, gene A has redundant activation pathways, and is therefore insensitive to loss of one of them. Should a degenerative mutation cause a redundantly regulated allele of A to lose the ability to self-regulate, B will have been established as an essential activator, i.e., the pathway will have been augmented by a step. In principle, this process could be repeated anew as B acquires sensitivity to a further upstream gene C and loses the ability to constitutively express.

As in the case of subfunction fission and duplicate-gene subfunctionalization, the probability of establishment of these types of changes will depend on N_g. This is because a redundantly regulated allele has a weak mutational advantage equal to the rate of loss of a regulatory site (u_l); one such mutation will result in the nonfunctionalization of either a self-regulated or an upstream-dependent allele, but will leave the function of a redundantly regulated allele unaltered. If $1/N_g > u_l$, such an advantage will be too small to be influenced by selection, and the population will evolve to an allelic state that simply depends on the relative rates of gain and loss of regulatory sites (u_g and u_l in Fig. 3). In contrast, if $1/N_g < u_l$, the accumulation of upstream-dependent alleles will be inhibited by their weak mutational burden and their lack of function in genetic backgrounds that fail to support A–B crosstalk. Thus, whereas small N_g may promote the passive elongation of genetic pathways, large N_g has the opposite effect. This does not mean that the augmentation of obligatory pathways cannot occur in very large populations, but if such changes are to occur, they must be of immediate selective advantage.

Evolvability

All replicating populations are capable of evolution, but it has recently been argued that some species are better at it than others, with natural selection directly advancing features of genomic architecture, genetic networks, and developmental pathways to promote the future ability of a species to adaptively evolve. Such speculation, which is almost entirely restricted to molecular and cell biologists and those who study digital organ-

isms (e.g., refs. 1, 2, and 81–88), has been subject to considerable criticism by evolutionary biologists (e.g., refs. 89–93). The term evolvability has long been in use in quantitative genetics, where is it has a precise definition closely related to the concept of heritability, i.e., the relative amount of standing variation that is subject to a response to natural selection (94, 95). However, the above-mentioned authors use the word in a rather different way, loosely defining evolvability to be the ability of a lineage to generate useful adaptive variation via mutational flexibility. Regardless of the definition, the idea that variation in evolvability exists among species is secure, as it has long been known that organisms and classes of traits vary in their propensities to respond to natural selection (18). Less secure is the idea that the ability to evolve itself is actively promoted by directional selection. Four reasons for skepticism follow.

First, evolution is a population-level feature. Thus, if an organismal feature that modifies the ability to evolve is to be advanced directly by adaptive mechanisms, selection must operate efficiently at a higher level of organization than the individual. This requires a significantly subdivided population structure, with levels of evolvability being positively correlated with population longevity and/or productivity. Because populations survive longer than individuals, such group selection is expected to be a much weaker force than individual selection, and necessarily operates on much longer time scales. If evolvability is to be subject to selective advancement, at least three stringent conditions must be met: (*i*) the group advantages of the genomic attribute that enhances evolvability must exceed any conflicting pressures operating at the individual level; (*ii*) the enhanced capacity for rapid evolutionary change must persist over the time scale of group selection; and (*iii*) while en route to fixation at the population level, the alleles that promote evolvability must remain tightly linked to the loci whose evolution is advantageous. Do such conditions ever exist in nature? The evidence for individual-level selection is overwhelming (96, 97), but aside from the matter of kin selection in behavioral evolution (98–100), the evidence for the operation of group selection is weak, although some investigators remain more optimistic than others (101, 102).

Second, it is by no means clear that an enhanced ability to evolve is generally advantageous. The dynamics of genetic variance for quantitative traits is complex, with selection modifying allele frequencies at epistatically interacting loci in ways that can either increase or decrease heritabilities, regardless of the advantage of the traits under selection (16). In addition, one can just as easily point to a long list of pathologies that can arise from an overly rapid proliferation of a new phenotype, and such scenarios have motivated a completely alternative, and equally speculative, view, that selection can favor mechanisms that suppress evolvability (103). Furthermore, theoretical studies have shown that the kinds of complexities that are often focused on by those enamored with evolvability (e.g., increased dimensionality and modularity) can actually inhibit the rate of adaptive evolution (104–106). Although the arguments are technical, they are no more abstract that the verbal reasoning of the evolvability school.

Third, there is no evidence that phylogenetic variation in evolutionary features reflects anything more than diversity in variation-generating factors that exist for purely physical reasons. For example, the biological features most likely to influence the ability to evolve, recombination and mutation rates, vary by orders of magnitude among species, with no sudden discontinuities in the lineages imagined to be most evolvable (animals and land plants) (13, 40). Such variation appears to be a simple by-product of alterations in chromosome lengths and numbers of germ-line cell divisions: because chromosome number is independent of genome size, and there is typically one meiotic cross-over event per chromosome, a doubling in genome size generally results in a 50% reduction in the recombination rate per physical distance; and because a large fraction of mutations are generated during replication, a doubling in the number of germ-line cell divisions doubles the per-generation mutation rate.

Fourth, comparative genomics provides no support for the idea that genome architectural changes have been promoted in multicellular lineages so as to enhance their ability to evolve (13). Indeed, other than the appearance of spliceosomal introns, some forms of mobile elements, and organelles in the stem eukaryote, there are no discontinuities in the basic features of genomes across the entire domain of cellular life. Moreover, as noted above, the additional genomic complexities of multicellular eukaryotes appear not to have arisen by positive selection but instead to have emerged passively in population-genetic environments where the efficiency of selection is relaxed, quite contrary to the view espoused by the evolvability school. Many unicellular species are excluded from certain evolutionary pathways that are open to multicellular species, and vice versa, but this is simply an indirect consequence of the altered power of nonadaptive evolutionary forces in these different contexts, not a direct outcome of natural selection for the ability to engage in particular evolutionary pursuits.

Closing Comments

Because it deals with observations on historical outcomes, frequently in the face of incomplete information, the field of evolution attracts significantly more speculation than the average area of science. Nevertheless, a substantial body of well tested theory provides the basis for understanding the pathways that are open to evolutionary exploration in various population-genetic contexts. Four of the major buzzwords in biology today are complexity, modularity, evolvability, and robustness, and it is often claimed that ill-defined mechanisms not previously appreciated by evolutionary biologists must be invoked to explain the existence of emergent properties that putatively enhance the long-term success of extant taxa. This stance is not very different from the intelligent-design philosophy of invoking unknown mechanisms to explain biodiversity. Although those who promote the concept of the adaptive evolution of the above features are by no means intelligent-design advocates, the burden of evidence for invoking an all-powerful guiding hand of natural selection should be no less stringent than one would demand of a creationist. If evolutionary science is to move forward, the standards of the field should be set no lower than in any other area of inquiry.

The field of population genetics is technically demanding, and it is well known that most biologists abhor all things mathematical. However, the details do matter in the field of evolutionary biology. As discussed above, many aspects of biology that superficially appear to have adaptive roots almost certainly owe their existence in part to nonadaptive processes. Such conclusions would be difficult to reach without a formal population-genetic framework, but they equally rely on observations from molecular, genomic, and cell biology. Such conclusions also raise significant challenges. If complexity, modularity, evolvability, and/or robustness are entirely products of adaptive processes, then where is the evidence? What are the expected patterns of evolution of such properties in the absence of selection, and what types of observations would be acceptable as a falsification of a null, nonadaptive hypothesis?

This tone of dissent is not meant to be disrespectful. The development of a mature field of evolutionary biology requires the participation of not just population geneticists, but molecular, cell, and developmental biologists. However, the integration of these fields needs to be a two-way street. Because the forces of mutation, recombination, and genetic drift are now readily quantifiable in multiple species, there is no longer any justification for blindly launching suppositions about adaptive scenarios without an evaluation of the likelihood of nonadaptive alternatives. Moreover, if the conclusion that nonadaptive processes have played a central

role in driving evolutionary patterns is correct, the origins of biological complexity should no longer be viewed as extraordinarily low-probability outcomes of unobservable adaptive challenges, but expected derivatives of the special population-genetic features of DNA-based genomes. A similar point has been made previously by Kauffman (3), although his conclusions were derived from models far removed from mainstream population genetics.

1. Kirschner M, Gerhart J (2005) *The Plausibility of Life* (Yale Univ Press, New Haven, CT).
2. Depew DJ, Weber DH, eds (1985) *Evolution at a Crossroads: The New Biology and the New Philosophy Science* (MIT Press, Cambridge, MA).
3. Kauffman SA (1993) *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford Univ Press, Oxford).
4. Carroll SB (2005) *Endless Forms Most Beautiful* (Norton, New York).
5. Davidson EH (2006) *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution* (Academic, New York).
6. Darwin C (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, London).
7. Dawkins R (1976) *The Selfish Gene* (Oxford Univ Press, New York).
8. Dawkins R (1986) *The Blind Watchmaker* (Norton, New York).
9. Dawkins R (1996) *Climbing Mount Improbable* (Norton, New York).
10. Crow JF, Kimura M (1970) *An Introduction to Population Genetics Theory* (Harper & Row, New York).
11. Bürger R (2000) *The Mathematical Theory of Selection, Recombination, and Mutation* (Wiley, New York).
12. Kimura M (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ Press, Cambridge).
13. Lynch M (2007) *The Origins of Genome Architecture* (Sinauer, Sunderland, MA).
14. Felsenstein J (1975) *Evolution (Lawrence, Kans)* 29:587–590.
15. Wolf JB, Brodie ED, III, Wade MJ, eds (2000) *Epistasis and the Evolutionary Process* (Oxford Univ Press, New York).
16. Carter AJ, Hermisson J, Hansen TF (2005) *Theor Popul Biol* 68:179–196.
17. Stoltzfus A (2006) *Evol Dev* 8:304–317.
18. Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics* (Longman, New York).
19. Kimura M (1965) *Proc Natl Acad Sci USA* 54:731–736.
20. Lande R (1975) *Genet Res* 26:221–235.
21. Bulmer MG (1980) *The Mathematical Theory of Quantitative Genetics* (Oxford Univ Press, Oxford).
22. Fisher RA (1930) *The Genetical Theory of Natural Selection* (Clarendon, Oxford).
23. Haldane JBS (1932) *The Causes of Evolution* (Longmans Green, New York).
24. Darwin C (1866) *The Variation of Animals and Plants Under Domestication* (John Murray, London).
25. Morgan TH (1925) *Evolution and Genetics* (Princeton Univ Press, Princeton).
26. Dover G (1982) *Nature* 299:111–117.
27. Nei M (1987) *Molecular Evolutionary Genetics* (Columbia Univ Press, New York).
28. Nei M (2005) *Mol Biol Evol* 22:2318–2342.
29. Cavalier-Smith T (1997) *Trends Genet* 13:6–9.
30. Yampolsky LY, Stoltzfus A (2001) *Evol Dev* 3:73–83.
31. Stoltzfus A (2006) *Mol Biol Evol* 23:1852–1862.
32. Lande R (1976) *Evolution (Lawrence, Kans)* 30:314–334.
33. Lynch M, Hill WG (1986) *Evolution (Lawrence, Kans)* 40:915–935.
34. Lynch M (1994) in *Ecological Genetics*, ed Real L (Princeton Univ Press, Princeton), pp 86–108.
35. Gould SJ, Lewontin RC (1979) *Proc R Soc London Ser B* 205:581–598.
36. Ohta T (1972) *J Mol Evol* 1:150–157.
37. Ohta T (1973) *Nature* 246:96–98.
38. Ohta T (1974) *Nature* 252:351–354.
39. Lynch M, Conery JS (2003) *Science* 302:1401–1404.
40. Lynch M (2006) *Mol Biol Evol* 23:450–468.
41. Lynch M (2002) *Proc Natl Acad Sci USA* 99:6118–6123.
42. Gasch AP, Moses AM, Chiang DY, Fraser HB, Berardini M, Eisen MB (2004) *PLoS Biol* 2:2202–2219.
43. Lynch M, Scofield DG, Hong X (2005) *Mol Biol Evol* 22:1137–1146.
44. Hahn MW, Stajich JE, Wray GA (2003) *Mol Biol Evol* 20:901–906.
45. King N (2004) *Dev Cell* 7:313–325.
46. Jacob F (1977) *Science* 196:1161–1166.
47. Raup DM (1978) *Paleobiology* 4:1–15.
48. Stanley SM (1985) *Paleobiology* 11:13–26.
49. Maynard Smith J, Szathmáry E (1995) *The Major Transitions in Evolution* (Freeman, San Francisco).
50. Michod RE (1999) *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality* (Princeton Univ Press, Princeton).
51. Technau U, Rudd S, Maxwell P, Gordon PM, Saina M, Grasso LC, Hayward DC, Sensen CW, Saint R, Holstein TW, *et al.* (2005) *Trends Genet* 21:633–639.
52. Kurland CG, Collins LJ, Penny D (2006) *Science* 312:1011–1014.
53. Martin W, Koonin EV (2006) *Nature* 440:41–45.
54. Lopez-Garcia P, Moreira D (2006) *BioEssays* 28:525–533.
55. Koonin EV (2006) *Biol Direct* 1:22.
56. Stanek D, Neugebauer KM (2006) *Chromosoma* 115:343–354.
57. Sheth U, Parker R (2006) *Cell* 125:1095–1109.
58. Lynch M, Hong X, Scofield DG (2006) in *Nonsense-Mediated mRNA Decay*, ed Maquat LE (Landes Bioscience, Georgetown, TX), pp 197–211.
59. Johnson NA, Porter AH (2000) *J Theor Biol* 205:527–542.
60. Johnson NA, Porter AH (2001) *Genetica* 112/113:45:58.
61. Stern DL (2000) *Evolution (Lawrence, Kans)* 54:1079–1091.
62. Delattre M, Felix MA (2002) *BioEssays* 23:807–819.
63. True JR, Haag ES (2001) *Evol Dev* 3:109–119.
64. Rockman MV, Wray GA (2002) *Mol Biol Evol* 19:1991–2004.
65. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA (2003) *Mol Biol Evol* 20:1377–1419.
66. Force A, Cresko WA, Pickett FB, Proulx S, Amemiya C, Lynch M (2005) *Genetics* 170:433–446.
67. King MC, Wilson AC (1975) *Science* 188:107–116.
68. Carroll SB (2005) *PLoS Biol* 3:1160–1166.
69. Alonso CR, Wilkins AS (2005) *Nat Rev Genet* 6:709–715.
70. Davidson EH, Erwin DH (2006) *Science* 311:796–800.
71. Sommer RJ (1997) *BioEssays* 19:225–231.
72. Ludwig MZ, Bergman C, Patel NH, Kreitman M (2000) *Nature* 403:564–567.
73. Ruvinsky I, Ruvkun G (2003) *Development (Cambridge, UK)* 130:5133–5142.
74. Tsong AE, Tuch BB, Li H, Johnson AD (2006) *Nature* 443:415–420.
75. Katju V, Lynch M (2003) *Genetics* 165:1793–1803.
76. Force A, Lynch M, Pickett B, Amores A, Yan Y-L, Postlethwait J (1999) *Genetics* 151:1531–1545.
77. Prince VE, Pickett FB (2002) *Nat Rev Genet* 3:827–837.
78. Lynch M, O'Hely M, Walsh B, Force A (2001) *Genetics* 159:1789–1804.
79. Wilkins AS (2002) *The Evolution of Developmental Pathways* (Sinauer, Sunderland, MA).
80. Wilkins AS (2005) *Brain Res Bull* 66:495–509.
81. Kirschner M, Gerhart J (1998) *Proc Natl Acad Sci USA* 95:8420–8427.
82. Gerhart J, Kirschner M (1997) *Cells, Embryos, and Evolution* (Blackwell, Malden, MA).
83. Rutherford SL, Lindquist S (1998) *Nature* 396:336–342.
84. True HL, Lindquist SL (2000) *Nature* 407:477–483.
85. Caporale LH (2003) *Am Sci* 91:234–241.
86. Earl DJ, Deem MW (2004) *Proc Natl Acad Sci USA* 101:11531–11536.
87. Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) *Proc Natl Acad Sci USA* 103:5869–5874.
88. Federici D, Downing K (2006) *Artif Life* 12:381–409.
89. Williams GC (1966) *Adaptation and Natural Selection* (Princeton Univ Press, Princeton).
90. Dickinson WJ, Seger J (1999) *Nature* 399:30.
91. Partridge L, Barton NH (2000) *Nature* 407:457–458.
92. Brookfield JF (2001) *Curr Biol* 11:R106–R108.
93. Sniegowski PD, Murphy HA (2006) *Curr Biol* 16:R831–R834.
94. Houle D (1992) *Genetics* 130:195–204.
95. Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits* (Sinauer, Sunderland, MA).
96. Endler JA (1986) *Natural Selection in the Wild* (Princeton Univ Press, Princeton).
97. Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CE, Hoang A, Gibert P, Beerli P (2001) *Am Nat* 157:245–261.
98. Hamilton WD (1964) *J Theor Biol* 7:1–16.
99. Hamilton WD (1964) *J Theor Biol* 7:17–52.
100. Wilson EO (1975) *Sociobiology: The New Synthesis* (Belknap, Cambridge, MA).
101. Coyne JA, Barton NH, Turelli M (2000) *Evolution (Lawrence, Kans)* 54:306–317.
102. Goodnight CJ, Wade MJ (2000) *Evolution (Lawrence, Kans)* 54:317–324.
103. Altenberg L (2005) *Artif Life* 11:427–443.
104. Orr HA (2000) *Evolution (Lawrence, Kans)* 54:13–20.
105. Welch JJ, Waxman D (2003) *Evolution (Lawrence, Kans)* 57:1723–1734.
106. Haygood R (2006) *Mol Biol Evol* 23:957–963.