

Gene Network Polymorphism Is the Raw Material of Natural Selection: The Selfish Gene Network Hypothesis

Zsolt Boldogkői

Laboratory of Neuromorphology, Department of Anatomy, Faculty of Medicine, University of Budapest, Budapest, Hungary

Received: 16 June 2003 / Accepted: 21 January 2004

Abstract. Population genetics, the mathematical theory of modern evolutionary biology, defines evolution as the alteration of the frequency of distinct gene variants (alleles) differing in fitness over the time. The major problem with this view is that in gene and protein sequences we can find little evidence concerning the molecular basis of phenotypic variance, especially those that would confer adaptive benefit to the bearers. Some novel data, however, suggest that a large amount of genetic variation exists in the regulatory region of genes within populations. In addition, comparison of homologous DNA sequences of various species shows that evolution appears to depend more strongly on gene expression than on the genes themselves. Furthermore, it has been demonstrated in several systems that genes form functional networks, whose products exhibit interrelated expression profiles. Finally, it has been found that regulatory circuits of development behave as evolutionary units. These data demonstrate that our view of evolution calls for a new synthesis. In this article I propose a novel concept, termed the selfish gene network hypothesis, which is based on an overall consideration of the above findings. The major statements of this hypothesis are as follows. (1) Instead of individual genes, gene networks (GNs) are responsible for the determination of traits and behaviors. (2) The primary source of microevolution is the intraspecific polymorphism in GNs and not the allelic variation in either the coding or the regulatory

sequences of individual genes. (3) GN polymorphism is generated by the variation in the regulatory regions of the component genes and not by the variance in their coding sequences. (4) Evolution proceeds through continuous restructuring of the composition of GNs rather than fixing of specific alleles or GN variants.

Key words: Development — Gene network — Neuroscience — Polymorphism — Population genetics — Selfish gene

Introduction

Currently, we are witnessing a shift in our view on the issue of how genes determine phenotype (Table 1). The major focus of reverse genetics*¹ is to decipher the function of a single gene. The investigation paradigm of this approach is based on a “perturbation–consequence” scheme, where first the gene expression is altered (overexpressed, knocked down,* or knocked out*), followed by analysis of the obtained phenotype. Similarly, forward genetics* seeks to identify single gene mutations, which are expected to be responsible for the altered phenotypes. The conclusions of these experiments are always invariant: the particular gene specifies the normal development of the traits or behaviors that were genetically altered. However, as yet no case has been demonstrated

Correspondence to: Zsolt Boldogkői; email: boldog@sb4.szote.u-szeged.hu

¹Terms marked with an asterisk are defined in the Glossary.

Table 1. Concepts on how genes encode phenotypes

| Era | Research paradigm | View on how genes operate |
|------------------------------|---|--|
| Classical genetics | One gene encodes one phenotype | One gene encodes one phenotype |
| Forward and reverse genetics | One gene encodes one phenotype | Dualistic or incoherent |
| Post-genome era | Functionally interrelated gene expressions encode the phenotype | Pleiotropy is the rule; genes make up functional modules |

where the lines of causality could be mapped from a single gene to a trait or behavior. In contrast, biological functions arise from interactions among many components. Complete-genome, transcriptome,* and proteome* analyses offer a new platform for understanding the genetic basis of various functions by allowing the study of gene regulation globally, rather than through the examination of individual gene effects. It has been revealed that spatiotemporal cascades of transcriptional factors control the early embryogenesis. Regulatory networks have also been shown to be included in several postnatal mechanisms, such as cell cycle control, apoptosis, cell signaling, cell metabolism, and neuronal regulation. Thus, the various functional systems appear to be coordinated by not yet precisely characterized gene expression networks, whose members are connected by both horizontal (spatial) and vertical (temporal) interactions. Comparison of whole-genome sequence data has revealed that *de novo* gene inventions cannot explain recent evolutionary events leading to differences among distinct phylogenic taxa (Waterston et al. 2002). Instead, data indicate that new combinations and modifications of preexisting molecular characters could be the means whereby phenotypic novelties are produced during the course of evolution. It has also been shown that in recent evolution the structure and function of proteins are highly conserved, while gene regulation rapidly evolves. Although, data are still sporadic, there is growing evidence for the existence of a high intraspecific variation in the cis-regulatory sequences (CRSs) of genes. Together, the elements for a new evolutionary synthesis appear to be available; however, a clear conceptual framework for this synthesis has not yet been worked out.

In an attempt to propose a theoretical account for a new synthesis, I put forward the selfish gene network hypothesis, which is based on the unification of the above findings. This hypothesis claims that (1) gene networks (GNs) coordinate all functional systems of the organisms; (2) GNs exhibit a high level of intrapopulation variation; (3) the genetic variance of GNs arises from the variation in the regulatory sequences of their components (genes); (4) the expression profiles of GN elements are mutually interdependent, which further enhances or reduces the

variation at both the gene expression and the phenotypic levels; and (5) in the modern history of evolution the source of natural selection is not the allelic diversity of individual genes varying in either the coding or the regulatory sequences, but the natural variation in GNs.

Gene Evolution—The World of “Beanbag Genetics”

Gene evolution is the fundamental principle of modern evolutionary theory. We have to make a distinction, however, between two kinds: the emergence of new genes and the adaptive changes of gene function. A new gene can arise by the duplication of a gene, which is followed by divergent evolution resulting in the co-option of one of the daughter genes for a new function. It has been proposed that the antisense DNA strand has the capability to conduct the evolution of novel genes (Zull and Smith 1990; Blalock 1990; Yomo et al. 1992; Cebrat and Dudek 1996), however, this concept is currently under debate (Boldogkői et al. 1994, 1995, 1999; Boldogkői 2000). Another method of gene evolution is based on exon shuffling (Patthy 1999), whereby proteins can gain new domains. Parasite palindromic repeat sequences that are equally capable of propagation within intergenic regions and protein-coding genes can be considered special cases of exon shuffling (Claverie and Ogata 2003). Further, horizontal transfer of gene clusters (operons) within prokaryotes has been proposed to play an important role in evolution (Lawrence and Roth, 1996).

Another type of gene evolution is the gradual improvement of gene properties. In fact, population genetics restricts its focus to the latter aspect and regards evolution as competition among various alleles differing in fitness. Ernst Mayr compared this type of modeling to sampling from a bag of colored beans and dismissed it as “beanbag genetics.” Others have also criticized these models, claiming that they are oversimplistic and tautological. Indeed, the current view of evolution, generated by population genetics, raises several unanswered questions. First, Darwin’s original concept on evolution was gradualist in the sense that he believed in a continuum of variation, on which natural selection could operate.

In contrast, Mendel and his rediscoverers assumed that traits were inherited in a discrete manner. This apparent contradiction was seemingly resolved by the neo-Darwinians, arguing that gradual quantitative differences are composed of the cumulative effects of many different loci, each behaving in a Mendelian, particulate fashion. At the time of New Synthesis, the most widespread belief among evolutionary biologists was that the majority of variations were maintained by balancing selection. That is, the polymorphism was thought to be preserved at loci because organisms carrying two different alleles are more fit than those containing two alleles of the same type. The finding that a large proportion of loci are polymorphic was problematic for this concept, because it would have imposed an extremely high genetic load for a population (Fisher 1930). Kimura (1968), in an attempt to solve this paradox, claimed that the majority of changes at the level of DNA and proteins are of little functional consequences for the organism and, therefore, are not subject to selection. However, the genetic load still remains too high, even if only a few genes were under simultaneous selection. In fact, if we examine natural populations, selection appears to be much more intense than is often assumed. In addition, there is an apparent contradiction between the rates of molecular evolution and macroevolution. The molecular clock ticks at an even rate for a specific gene. In contrast, macroevolution exhibits a highly variable tempo of evolution at various geological periods. So the major question is, How we can reconcile the constant rate of evolution at the molecular level with the sudden arise of an enormous number of new species with completely novel structural architectures? Following saltational theory of Simpson (1944) and Gould and Eldredge (1993), it is now a popular view that the speed of evolutionary changes can differ in various geological periods; however, this concern is actually not included in the mathematical formulas of population genetics. Population genetics uses terms such as “additive” or “epistatic” contributions across loci, which are intended to indicate interactions among genes. However, they only confuse the view based on a linear, interaction-free gene–phenotype relationship. This gene-centered concept appears in an extreme form in Dawkins’s selfish gene theory (1976), which claims that genes are the interest-bearers of evolution, and individuals are simply vehicles of the replicators (genes). The fundamental question that can be raised concerning the central idea of population genetics is as follows: Could the variations in protein sequences account for the adaptive differences among individuals? In fact, molecular biology has provided very little evidence so far to support this assumption.

The Four Lines of Evidence

The four groups of findings that led me to raise the selfish GN hypothesis are as follows (see also Table 2).

Functional Systems Are Coordinated by Networks of Interacting Genes

How Is the Phenotype Determined? The traditional genetic approaches are based on the assumption that a linear relationship exists between a single gene and a particular phenotype. This reductionistic view is seemingly supported by the finding that certain traits are inherited in a discrete manner, as well as by the observations that single gene mutations are frequently coupled with invariant phenotypes. However, traits including discrete traits are obviously specified by several genes. Although, in the event of monogenic inheritance, an individual carrying a certain allele exhibits a typical phenotype, this fact merely indicates that the variance of other genes playing a role in specifying this character generally does not produce variance in the particular phenotype. In addition, the “mutation–typical phenotype” relationship does not mean that the function of a given gene is to encode the normal, nonmutant trait. This fact can be highlighted by an analogue. If one removes the transistor from a radio, it will emit a howl instead of a song by Dire Straits, which state can be set back by reinstalling the transistor. From this experiment one cannot infer that the functions of the transistor are to encode the particular song and to suppress the howl. Despite the fact that the monogenic inheritance of traits is a scientifically false idea, individual genes can be the real targets of selection if their variation alone is sufficient to generate the corresponding phenotypic variance. The only problem is that those characters whose inheritance can be described by Mendelian genetics are of peripheral significance in modern biology and, hence, in evolutionary biology too. Harmful mutations often behave in a Mendelian manner and therefore are the ideal object of population genetics. However, they are simply subject to purifying selection and do not contribute to the adaptive variation of a population. Current scientific findings lead to the view that, in contrast to simple linear assembly line, complex GNs have evolved to coordinate the activities of gene expressions. These functional networks operate as (semi-)autonomous organizational units, and therefore they can be considered as the genetic bases of phenotypes.

Regulatory Circuits in Development. Developmental biology has demonstrated that specific regulatory networks control the ontogenesis. The

Table 2. The four lines of evidence supporting the selfish gene network hypothesis

| Findings | Importance in evolutionary biology |
|--|--|
| 1. Networks of interacting genes | |
| Transcription factor cascades in development | Ontogenesis is coordinated by network-like structures. |
| Pleiotropic effects of gene elimination | Interdependence of gene expressions |
| Transgenes induce rearrangements in the expression profile of endogenous genes. | Interrelated gene expressions |
| Cell signaling, cell cycle control, apoptosis, metabolic pathways | The existence of regulatory circuits |
| The importance of expression levels and patterns in cell signaling | Connected gene expressions |
| Chronic change in the neurotransmitter or receptor levels induces complex alterations. | Regulatory cascades in operation |
| Use-dependent plasticity of brain | The GN–GPN relation is highly flexible in the brain. |
| SSRIs alter the expression of several genes. | Self-organization of regulatory circuits |
| 2. Evolution of gene regulation | |
| Evolution of transcription factors | Divergence or conservation |
| Evolution of the CRSs of transcription factors | Divergence or conservation |
| CRS polymorphism in closely related species | Gene regulation could be the target of evolution |
| Cross-species ectopic expression of orthologs | Conservation of the expression of specific developmentally important genes |
| 3. The source of microevolution | |
| Intrapopulation CRS polymorphism | The potential raw material of natural selection |
| Intrapopulation variance in gene expression | CRS variance is related to expression polymorphism. |
| CRS polymorphism in strains of species | Gene regulation could be the target of evolution. |
| Developmental variants | The potential source of natural selection |
| 4. Developmental systems evolve as units | |
| Conservation of gene sequences of a circuitry | Strong negative selection for maintaining sequence constancy |
| Evolution of genes with functional conservation | Despite sequence alteration, the basic function of the circuitry is conserved. |
| Divergence of developmental pathways—retained function | The basic function is conserved, but with variance in the execution. |
| The same pathway is recruited for different functions. | Developmental pathways behave as evolutionary units. |
| Evolution of function: the co-option hypothesis | Evolving new functions of developmental modules |
| The relationships between modules are modified by evolution. | Crosstalk among gene networks is the target of selection. |
| The same regulatory networks recruited for the same function by convergent evolution | E.g. heart development in insects and mammals; difficult to explain |

molecular mechanisms underlying the early embryonic events of *Drosophila* and mouse are relatively well known. The heart of these networks consists of developmental regulatory genes, particularly those that encode transcription factors. The interaction among the network components is mediated by CRSs of regulatory genes, which generally receive multiple inputs from the upstream components of the network. The combinatorial binding pattern of transcription factors to the CRSs provides differential spatiotemporal control for the expression of target regulatory genes. Davidson et al. (2002) examined the specification of the endomesoderm of sea urchin embryos and revealed the existence of a regulatory network containing over 40 genes that participated in this process. Other studies have also reported the

importance of gene expression cascades in the regulation of kidney (Burrow, 2000), skeletal (Horton, 2003), limb (Li and Cao, 2003), and pancreatic (Wilson et al. 2003) development and in adipogenesis (Cantile et al. 2003). Von Dassow and coworkers (1999) used computer simulation to investigate whether known interactions among segment polarity genes of *Drosophila* suffice to confer the properties expected of a developmental module. They found this network robust* for varying inputs and, hence, concluded that segment polarity genes exhibit a modular design.

Neuronal Gene Expression Networks. The vertebrate brain exhibits an apparent modular design at various hierarchical levels, including anatomical,

biochemical, physiological, and functional levels. However, data are still too sporadic to have a clear picture on the modularity of gene expressions. In this section I present various examples indicating the inherent connectedness among genes. Molecular neurobiology and psychopharmacology have demonstrated that a chronic change in the level of a neurotransmitter can cause complex alterations in the expression of several other neurotransmitters and receptors in various CNS structures (Calon et al. 2002), in specific mental processes (Young and Leyton 2002; Nieoullon 2002), and can lead to complex neurological diseases, such as depression (Middlemiss et al. 2002). Analysis of the therapeutic mechanism of antidepressant selective serotonin reuptake inhibitors (SSRIs) has revealed that depression is not caused by a deficiency of serotonin. Instead, blocking serotonin reuptake induces complex reorganization in the expression of several receptors and neurotransmitters (Bymaster et al. 2002). Gene targeting experiments have shown that elimination of a single gene in mice causes pleiotropic* effects. For example, it was demonstrated that CB1 cannabinoid receptor gene (*Cnr1*) deleted mutants displayed significantly increased levels of mRNAs of several neuropeptides, including substance P, dynorphin, and enkephalin, as well as GAD67, which resulted in the functional reorganization of the basal ganglia and in behavioral alterations (Steiner et al. 1999). Similar pleiotropic effects were reported in both knockout (Gerlai 1996, 2001; Schmid et al. 1999; Kest et al. 2001; Bailey et al. 2002; Bejar et al. 2002; Cripps and Olson 2002) and transgenic animals (de la Pompa et al. 1997; MacGowan et al. 2001; Erb et al. 2001; Boldogkői and Nógrádi, 2003). These studies show that the lack of an endogenous gene or overexpression of a transgene modifies the expression of several other systems and thereby masks the desired phenotype. The major issue in each particular case is that these pleiotropic alterations represent functional compensatory adaptations or they are, at least in part, maladaptive. Whatever the case is, these data indicate that gene activities are highly connected and that traditional knockout and transgenic research has a low reliability for predicting the function of a single gene, especially at the level of behavior. It has long been known that the mammalian brain exhibits a high degree of plasticity. For example, current evidence indicates reorganization of the rat motor cortex in association with motor behavior. In the overused motor cortex up-regulations were observed, including genes coding for voltage-gated ion channels, trafficking and targeting proteins, and intracellular kinase network members (Keyvani et al. 2002).

Thus, data on various fields of neurobiology suggest the existence of an inherent interrelatedness among genes, where connected neurons serve as the

physical framework for the gene activities, which are organized into higher-order functional units.

Other Systems Controlled by Regulatory Circuits. The GN controlling the cell cycle of yeast was recently revealed (Wyrick and Young 2002). It was shown that the majority of single gene mutations in yeast affect the expression of several other genes (Featherstone et al. 2002). Currently, computational techniques have been applied for inferring the genetic regulatory networks from whole-genome expression profiles of *Escherichia coli* (Bhan et al. 2002) and *Saccharomyces cerevisiae* (Ihmels et al. 2002). A comparative mathematical analysis of the metabolic networks of 43 organisms has recently been presented (Jeong et al. 2000).

Together, there is increasing recognition that functional systems cannot be reduced to simple sums of gene activities. Instead, novel data suggest the existence of a set of complex not yet precisely elucidated multilevel interactions among genes conferring emergent system-level properties for these interactions. Accordingly, organisms appear to be composed of (semi-)autonomous developmental and genetic units termed by various names, such as regulatory networks, gene expression networks, gene nets, GNs, and modules (see, e.g., Wagner, 1996; Raff, 1996; Bonner, 1998; Cantile et al., 2003). The existence of modular design is recognized at many levels of the biological hierarchy, and it represents an important issue debated in different disciplines. One of the foremost challenges of post-genomic era investigation will be to uncover these regulatory networks coordinating embryonic development, brain function, cell signaling, and virtually every functional system of the organisms.

Evolution of Gene Regulation—A Macroevolutionary Perspective

It is evident from cross-species comparisons of orthologous sequences that the coding regions of genes are well conserved. The CRSs, on the other hand, appear to exhibit a high variability among species (Wasserman et al. 2000). However, it is difficult to ascertain the exact extent of this variability since the CRSs of multicellular organisms can be located either upstream or downstream of genes, as well as in introns, and can span up to several hundred thousand base pairs. Additionally, the precise consensus sequences for certain regulatory elements are often unknown. Further, it is practically impossible to foretell whether sequence alterations will affect transcription at all and, if so, to what extent. Besides the control of transcription by transcription factors, another possibility for transregulation is achieved by

blocking the translation of mRNAs by means of antisense (a)RNAs. The significance of this kind of regulation appears to have been considerably underestimated until recently. Computational analysis has identified 2667 human genomic loci with potential transcription from both DNA strands. Microarray analysis has revealed that as many as 60% of them could be true sense–antisense pairs (Yelin et al. 2003). Moreover, Okazaki and colleagues (2002) identified 2431 pairs of sense–antisense overlapping transcripts in the mouse genome. The extent of variability in CRSs of antisense genes remains to be determined. Ribozymes* have also been shown to play an important role in the regulation of translation. Micro (mi)RNAs* found in various phylogenic taxa direct the posttranscriptional regulation of gene expression by binding to the corresponding mRNAs. It can be presumed that a huge number of noncoding RNA genes have not yet been discovered, the reason for which is that their computational identification is difficult, because they contain few hallmarks. Thus, it appears that various RNAs play far more significant roles in the regulation of gene expression than formerly believed. Enzymes involved in protein modification (kinases, chaperones, etc.) are also transregulatory factors (TRFs), whose expression could similarly be a potential target of natural selection.

Evolution of Developmental Regulatory Genes.

Transcription factors control the expression of a number of genes; therefore, any alteration in their coding or regulatory sequences is expected to exert a substantial effect on the overall gene expression. Consequently, it might be thought that the structure and expression of transcription factors, especially those of developmental regulators, should be broadly conserved across long evolutionary distances, or their change would result in drastic phenotypic alterations; accordingly, these changes might be regarded as singular events on a macroevolutionary timescale.

Evolution of the Coding Regions of Regulatory Genes.

Evolving functions: Fast evolutionary changes of the *Odyseus* locus within *Drosophila* species was described (Ting et al. 1998). This gene caused hybrid male sterility in interspecific crosses, which led to the hypothesis that it could be involved in reproductive isolation of closely related species. Interestingly, the *Hox3* gene is conserved in chordate lineages but is highly diverged in insects (Falciani et al. 1996). It was shown that the insect *Hox* proteins evolved new activities by gaining novel functional domains (Galant and Carroll of 2002; Ronshaugen et al. 2002). Another example of homeogene evolution is the *Quox-1* gene of birds. It was shown that, in contrast to the other vertebrate class I *Hox* genes, *Quox-1* acquired a novel

domain of expression that includes fore- and mid-brain, as well as the rest of the neural anlage, and is not involved in the anterioposterior patterning of the vertebral column (at least in quails [Xue et al., 1993]).

Conserved function: Functional conservation after sequential divergence was described in several organisms. For example, cross-species ectopic expression analysis revealed that, in spite of the noteworthy divergence of the gene sequences, the homeotic gene *Deformed* of the beetle *Tribolium* rescued *Drosophila* null mutations in its ortholog* (Brown et al. 1999). It was demonstrated that the *Ubx* gene from a velvet worm species could perform similar functions in specific tissues of *Drosophila* as its endogenous ortholog (Grenier and Carroll 2000). Further, Stern et al. (1993) showed that the human *GRB2* and *Drosophila* *Drk* genes could functionally replace the *Caenorhabditis elegans* cell signaling gene *sem-5*.

Evolution of the Regulatory Region of Transcription

Divergent sequence–altered function: There is accumulating evidence for cases when the divergence of CRSs has been coupled with functional alterations. Averof and Patel (1997) described that changes in the expression pattern of the Hox genes *Ubx* and *AbdA* in different crustaceans correlated well with the modification of their anterior thoracic limbs into feeding appendages. In addition, it was found that particular CRSs of the *Hoxc8* gene were specifically changed in chicken and baleen whales. Ectopic expression experiments showed that these CRSs conducted a distinct expression pattern in a transgenic mouse embryo model (Shasikant et al. 1998; Belting et al. 1998). Evolution of regulatory elements was suggested to be the reason for the divergence of a particular aspect of larval morphology (cuticle pattern) between closely related *Drosophila* species. Accordingly, it was assumed that changes in the expression pattern of a single specific locus (*ovo/shaven baby*) could fully account for the observed morphological differences (Sucena and Stern 2000). Further, Keränen et al. (1999) found significant differences in the expression of various orthologous regulatory genes playing roles in tooth development in mouse and vole.

Divergent sequence–conserved function: In spite of the divergence of regulatory sequences, the corresponding functions were found to be remarkably stable in some instances. For example, although the minimal promoters of the *Brachyury* gene (playing a role in notochord differentiation) significantly diverged in two distantly related ascidian species, they show functional conservation (Takahashi et al. 1999). Further, it was found that, although *even-skipped stripe 2* gene expression was strongly conserved in *Drosophila*, the stripe 2 element itself underwent considerable evolutionary changes both in its binding-site sequences and in the spacing between them

(Ludwig et al. 2000). This result was explained by assuming the occurrence of counterbalancing mutations in CRSs of other genes that were fixed by selection to maintain functional constancy. Mitsialis and Kafatos (1985) showed that, in spite of significant sequence divergence, the regulatory region of chorion genes of silk moth conducted a faithful reporter gene expression pattern in *Drosophila* follicle cells. Romano and Wray (2003) demonstrated that although an extensive divergence had taken place in the *Endo16* promoter of two sea urchin species, the pattern of transcription was largely conserved during embryonic and larval development. Moreover, in reciprocal cross-species transient expression assays, the authors showed that a set of transcription factors interacting with this promoter also changed. Those investigators proposed that, despite drastic changes in the promoter sequence and the mechanism of transcriptional regulation, stabilizing selection acted to maintain a similar expression pattern of the *Endo16* gene in the two species.

Even though the differences in the amounts and timing of the expression of particular regulatory genes are minor, the effects of this on phenotypic characters can be significant. Thus, albeit stabilizing selection may act to maintain the normal function of regulatory genes, minor genetic changes can produce considerable phenotypic variance in a population. Accordingly, the alteration-versus-conservation dilemma is not necessarily a paradox in biological systems. Conservation provides stability for the operation of living systems, while alteration provides variations on which natural selection can operate. In other words, adaptive genetic variation can be produced without altering the basic function of a system. However, specific developmental structures and expression patterns are astonishingly strongly conserved across large evolutionary distances, which calls for an explanation.

Evolution of Nonregulatory Gene Expression. So far, relatively few cis-regulatory elements were studied extensively. Waterstone et al. (2002) compared 95 well-characterized CRSs of human and mouse genes and found that the extent of conservation was considerably lower than in the coding regions. Very similar results were reported by others comparing these two species (Jareborg et al. 1999). It was also demonstrated that a given Hox protein regulated different target genes in different insects, owing to the evolution of Hox protein-binding sites in the CRSs of target genes (Weatherbee et al. 1998). Comparison of gene expression patterns of closely related species can provide useful information on microevolutionary events. Dickinson (1980) examined six enzymes in 14 different tissues in 27 *Drosophila* species and found that most enzymes showed substantial level of variation in tissue-specific expression.

Intraspecific Variance in Gene Regulation—The Source of Microevolution

A key issue in evolutionary biology is the source of the biological variation on which natural selection can operate. Macroevolutionary approaches provide only limited insight into how evolution proceeds in natural populations. It is therefore essential to assess the extent of intrapopulation genetic variation, which is the raw material of natural selection. The major issue is the genetic basis underlying the genetic polymorphism. The view has long been held that phenotypic polymorphism is produced by the variation within the open reading frames (ORFs*) of genes. Novel data are now compelling us to challenge this view. A large body of information is currently available on the coding region of genes, the reason being that cDNAs can be easily generated by reverse transcription. However, we have significantly fewer data on the noncoding sequences, especially on the extent to which gene regulation varies within a population. Oleksiak and colleagues (2002), analyzing the DNA sequences of a fish (*Fundulus* genus), found that the CRSs of the examined genes were highly variable, which resulted in significant differences in gene expression level among individuals. Cowles and coworkers (2002) published data on the extent of cis-regulatory variation in several genes across four inbred mouse strains. They analyzed 69 genes and found that at least four of them showed allelic differences in expression level of 1.5-fold or greater and that some of these differences were tissue specific. It was also shown that polymorphism in the CRSs of genes associated with immune defense made up a clear majority compared to polymorphism in the coding region of these genes (Mitchison et al. 2000). By examining 313 human genes of 82 unrelated individuals, Stephens et al. (2001) demonstrated that 5' upstream regions of genes were highly variable. Delattre and Felix (2001) found that vulval cell lineages of *C. elegans* exhibited a certain intraspecific vulval cell lineage polymorphism even within a strain, but it was higher among the strains. The polymorphism was multigenic (the underlying genetics was not ascertained) and represented epistatic relationships. In addition to division patterns, polymorphisms were also observed in vulval patterning mechanisms in a Nematode species (Felix et al. 2000).

Together, although the data are still sporadic, it appears that substantial variation in gene regulation exists among individuals within populations. In addition to cis-regulatory polymorphism, variance of trans-regulatory factors can also affect gene expression. However, data on the variance of TRF expressions are virtually unavailable at the moment.

Regulatory Pathways Are the Units of Evolutionary Change

There is an emerging consensus as to the existence of developmental and evolutionary modules. However, the operation of these modules and their roles in evolution are highly debated. Surprisingly, no hypotheses have been proposed so far to explain the mechanisms of how these modules evolved on a microevolutionary time scale.

Conserved Regulatory Circuits. In recent years, it has been revealed that the networks of several signaling and transcription factors playing a role in the development are evolutionarily conserved even among distantly related phylogenetic taxa. For example, de la Pompa and colleagues investigated the role of Notch signaling in the mouse neurogenesis by analyzing the effect of *Notch1* and *RBP-Jk* gene deletions and demonstrated that the Notch pathway, its regulatory mechanisms, and its role in neurogenesis are conserved from insects to mammals (de la Pompa et al. 1997). Cripps and Olson (2002) compared the mechanisms involved in heart formation in fruit flies and mammals in the context of a network of transcriptional interactions and revealed that signaling pathways and transcription factors are evolutionarily conserved. Further, it has been shown that the same regulatory circuits were often recruited in different processes. For example, it has been demonstrated that the tyrosine kinase receptor–Ras–MAP kinase pathway plays a role in humans in controlling cell proliferation, in *C. elegans* in vulval development, and in *Drosophila* in eye development. Felix and Sternberg (1997) examined the intercellular signaling mechanism of vulval development in four nematode species and found that the same pattern of fates could be obtained by distinct networks of cell interactions. In a recent study, Podani et al. (2001) compared the system-level properties of metabolic and information networks in 43 archaeal, bacterial, and eukaryotic species and concluded that scale-free organization of these networks was more conserved throughout evolution than their content. Further, it was found that, although the eyes of insects and mammals have arisen independently, the underlying regulatory networks exhibit striking similarities (Xu et al. 1999).

The mechanisms responsible for the remarkable conservation of several developmental pathways are poorly understood, but various proposals have been put forward to explain them (Table 3). (A) The gene redundancy hypothesis claims that developmental circuits were evolved to be robust against any changes, that is, to exhibit an inherent resistance to mutations and developmental noise. This hypothesis assumes that gene functions overlap; thus, the malfunction of

Table 3. Evolutionary compensation of harmful mutations

| | |
|--------------------------------------|--|
| Gene redundancy hypothesis | Overlapping gene function buffers harmful effects. |
| Within-module interaction hypothesis | Interactions within modules were designed to resist mutations. |
| Compensatory pleiotropy hypothesis | Interactions among nonmodular elements exert protection. |
| Chaperone hypothesis | Chaperones evolved to mask natural genetic variance. |
| Negative selection hypothesis | Selection constraints decrease the chance for survival and reproduction of individuals carrying the harmful mutations. |

any element is compensated by the others (see, e.g., de la Pompa et al. 1997). Basically, in the gene targeting literature, phenotype masking is generally explained in terms of an overlapping gene function (Steiner et al. 1999). Mutations in the *invected/engrailed*, *fz/dfz(2)*, and *cubitus interruptus/teashirt* genes of *Drosophila* were found to exhibit no or few phenotypic effects (Simmonds 1995; Muller et al. 1999; Gallet et al. 2000), which was explained by gene redundancy. (B) The *within-module interactivity hypothesis* (termed the robustness hypothesis by Galis et al. [2002]) proposes that the interaction of genes within a network accounts for the buffering of the harmful effects of mutations (von Dassow et al. 2000). (C) The *compensatory pleiotropy hypothesis* (termed the pleiotropy hypothesis by Galis et al. [2002]) claims that mutations do alter the expression pattern of a circuitry, but subsequent mutations in other genes exert a counterbalancing effect. That is, according to this concept the conservation of expression patterns among species reflects a high interactivity among the genes, and consequently their networks, if they exist at all, exhibit much less modularity, as proposed by von Dassow and colleagues. The robustness–pleiotropy debate originates from the observation that the gene expression patterns of some segment polarity genes and the interactions between them have been found to be highly conserved among insects (Patel et al. 1989; Brown et al. 1994; Grbic et al. 1996). Von Dassow and coworkers (2000) modeled the segment polarity network of *Drosophila* by computer simulation. They used as many as 136 differential equations to incorporate all relevant interactions. They obtained robustness to variation in the kinetic constants that govern behavior. If this is true, it must imply that evolution could rearrange inputs to modules without changing their intrinsic behavior. In contrast, Galis and colleagues (1999) evaluated the data available on this subject and found no evidence in support of the robustness hypothesis. (D) The chaperone hypothesis suggests that heat-shock protein (Hsp) 90 acts to stabilize developmental pathways by masking mutations, thereby fostering the accumulation of latent

variants that can be fixed in certain ecological circumstances (Rutherford and Lindquist 1998).

Together, the two robustness hypotheses presume that evolution created builtin mechanisms which allow the accumulation of malfunctioned mutations but block their harmful effects. In contrast, according to the compensatory pleiotropy hypothesis, buffering effects are not available when adverse mutations occur. New mutations in other genes are required for the compensatory effect to be exerted. This hypothesis must imply that the interactions between genes are designed in such a way that a novel mutation(s) in another gene(s) can easily exert a compensatory effect to stabilize the gene network. The existence of mechanisms whose function is to tolerate these mutations appears extremely unlikely. It seems much more logical that negative selection eliminates the individual(s) carrying the disadvantageous mutation before it can spread in the population if it is dominant or in a homozygous form if it is recessive. If the mutation is only mild, negative selection reduces the success of the efficiency of the transmission of the mutant gene to the subsequent generations. The above hypotheses are based on the false idea that the life of an individual is important from an evolutionary point of view. In fact, it is not. Evolution would never develop such mechanisms, which were disadvantageous for the future generations (see Interest Bearers, below, for a more detailed explanation). However, the stabilization of GNs against environmental perturbations or developmental noise would be favored by natural selection, which can appear as robustness or compensatory pleiotropy. The masking of natural variation or disadvantageous mutations can only be a side effect of this mechanism, not a direct target of natural selection.

Evolving Regulatory Circuits—The Co-option Hypothesis. The divergence of developmental pathways among various phylogenic groups has been described in several cases. The basic functions of these pathways, however, were conserved. On the other hand, some evidence has been published of the function of developmental circuits undergoing change. It has been demonstrated that genes connected by regulatory linkages can be co-opted as units to serve a novel function throughout evolution. Keys and colleagues (1999) hypothesized that the reorganization of a whole pathway of molecular interactions might be involved in the determination of eyespot patterns in butterfly wings. They speculated that this process did not include each gene of the pathway, but only one or two key genes, which originally had a different role in the development. The genetic alteration of these key genes was assumed to induce changes in the expressions of other members of the circuit through existing regulatory linkages. Spitz et al. (2001) demonstrated that

novel limb-specific CRSs played a role in the co-option of Hox genes during the evolution of tetrapod limbs. Thus, the co-option hypothesis claims that natural selection can find new uses for existing gene networks by changing either the function of the component genes or their regulation (True and Carroll 2002).

Collectively, regulatory circuits appear to behave as evolutionary units. Some developmentally significant pathways display a high conservation of expression patterns across large evolutionary distances, while other pathways diverge. Naturally, strong selection constraints must exist for the conservation of the basic function of a circuitry, and especially of those which are developmentally important. However, I assume that a great number of data on sequential divergence will be produced in the future, which will reveal substantial phenotypic variation with a retained native function of these pathways. In fact, change in the function of a regulatory circuit appears to be an extremely rare, but very important macroevolutionary event. The proposed models rather consider the consequences of earlier macroevolutionary events on the present-day structure and operation of these modules, but do not attempt to give an explanation of the mechanism or of how the evolution proceeds on a microevolutionary time scale. Specifically, the major problem is that there are no considerations on the origin or nature of genetic variations which are the prerequisites of any evolutionary events.

The Selfish Gene Network Hypothesis

Definitions

The Gene Network. A *gene network* (GN) is a cluster of genes with a high density of internal interactions and sparse connection to the rest. GNs are the functional organization of genes into higher-order units below the scale of the entire genome but above the regulatory circuits (pathways, networks), which terms are often used for the cascades of transcription factors or of signaling pathways. The communication between GNs occurs at the system level rather than between the individual components of the systems, that is, the input signals are processed by the network as a whole and their output signals are produced as a joint effect of their elements. In consequence of our incomplete knowledge on the exact regulatory connections, it is difficult at present to ascertain whether specific GNs form nonoverlapping clusters or distinguishable, but to a certain degree intermingled, groups, or more or less diffuse networks. At this moment, therefore, it is not possible to define exactly the physical underpinnings of GNs. As an example, GNs controlling the development of various organs all include the early embryonic cascades of tran-

scription factors and other substances; that is, they all overlap in the early ontogenic stages. Hence, particular developmental regulatory circuits form only submodules of a specific GN.

I neglect the polymorphism in the coding region of genes, which, I assume, plays a marginal role in the modern history of evolution. To be more precise, evolution of ORFs, especially those of transcription factors and other TRFs, may have significant macro-evolutionary importance; however, I do not think that they produce adaptive variance in a population.

A particular gene has several gene expression variants (*gevs*) differing in their CRSs (Fig. 1a). A GN allele is composed of a specific set of *gevs* of functionally related genes (Fig. 1b). Manifested GN alleles are the subset of GN alleles which are actually present in a given generation of a population, while latent GN alleles are only theoretical possibilities of specific *gev* combinations. Certain *gevs* interact at a given time (horizontal interaction), but others are downward or upward components of an expression cascade (vertical interaction), which relation is not necessarily unidirectional. I refer to the specific set of *gevs* which interacts horizontally as a snapshot GN (allele). Thus, a GN is composed of a consecutive series of snapshot GNs.

The Gene Product Network. A *gene product network* (GPN) is a set of gene products (proteins, RNAs) encoded by a GN and is comprised of specific gene expression profiles (*geps*). A *gep* is defined as the amount and spatiotemporal distribution of a gene product encoded by a *gev*. A functionally related cluster of *geps* makes up a GPN allele (Fig. 1b). A *gep* is determined by both the CRSs of its *gev* and the interaction with other *gev* products belonging to the same GPN. Thus, *geps* of a GPN exhibit mutually interdependent expression profiles. An alteration in the expression of any components of a GPN can have an effect on the overall expression of the network (Fig. 1c). The elements of a GN have various importance in determining the overall expression profile of a GPN. For example, certain developmental transcription factors or specific neurotransmitters or receptors in the brain have eminent roles in the operation of GPNs. Thus, the connectivity between genes is unevenly distributed, and this relation is hierarchical. Further, in early embryogenesis, epigenetic factors derived from maternal cells play roles in the determination of GPNs.

Coding of the Phenotype. Organisms are dynamic, continuously altering systems, and therefore, we can only see a glimpse of changes at a given time. I make a distinction between transitory and final phenotypes. Transitory phenotypes appear during the course of embryogenesis, while final phenotypes ap-

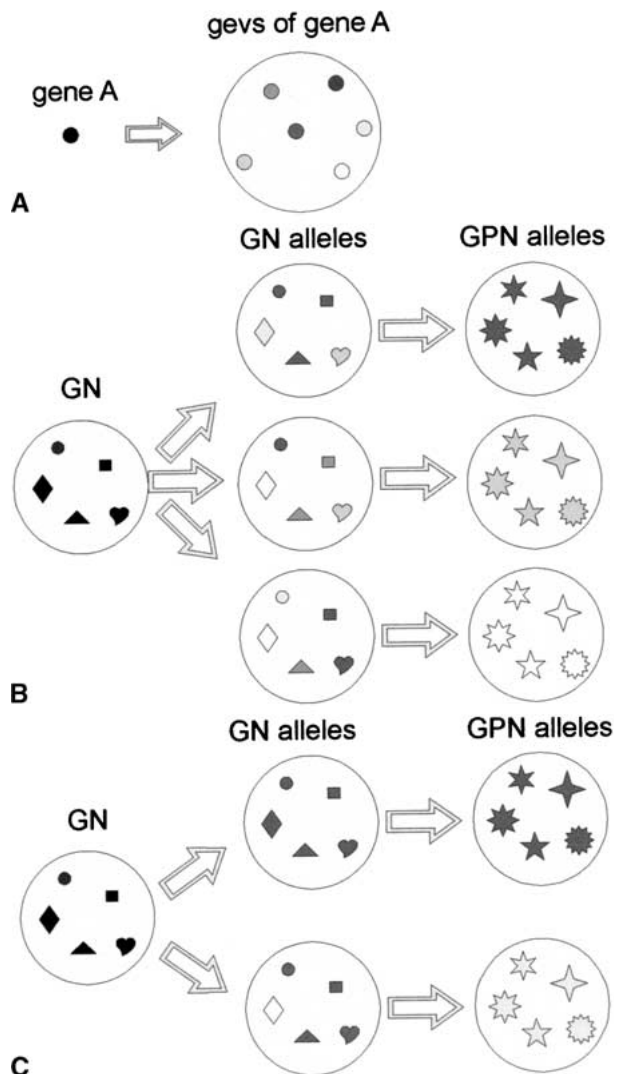


Fig. 1. What is a gene network? **A** Gene expression variants (*gevs*). A specific gene can have several *gevs* differing in their regulatory sequences. Allelic variants differing in their coding regions are neglected. **B** Gene network (GN) and gene product network (GPN) alleles. A GN comprises functionally related genes. A GN allele comprises a specific combination of *gevs* of functionally related genes. A GPN allele is composed of a specific set of gene expression profiles (*gep*). Each individual contains two copies of GN alleles, but for the sake of simplicity, I assume a haploid case. **C** A GPN is a dynamic system, therefore, the replacement of a single *gev* with another one (which can occur in sexual reproduction or by mutation) usually results in the alteration of the expression of several other members (*geps*) of a GPN allele.

pear at a certain postnatal stage and remain relatively stable. A particular phenotype is determined by both the genetic program and the effects of the external environment and the internal milieu (IM). The line of causality between GN and GPN is bidirectional. There are two hypothetical cases for their relation.

(a) The reciprocal interaction between GNs and GPNs can induce a series of consecutive transformations of both snapshot GNs and snapshot GPNs (Fig. 2a). In other words, the composition of the gene cluster that is actually interacting changes during the

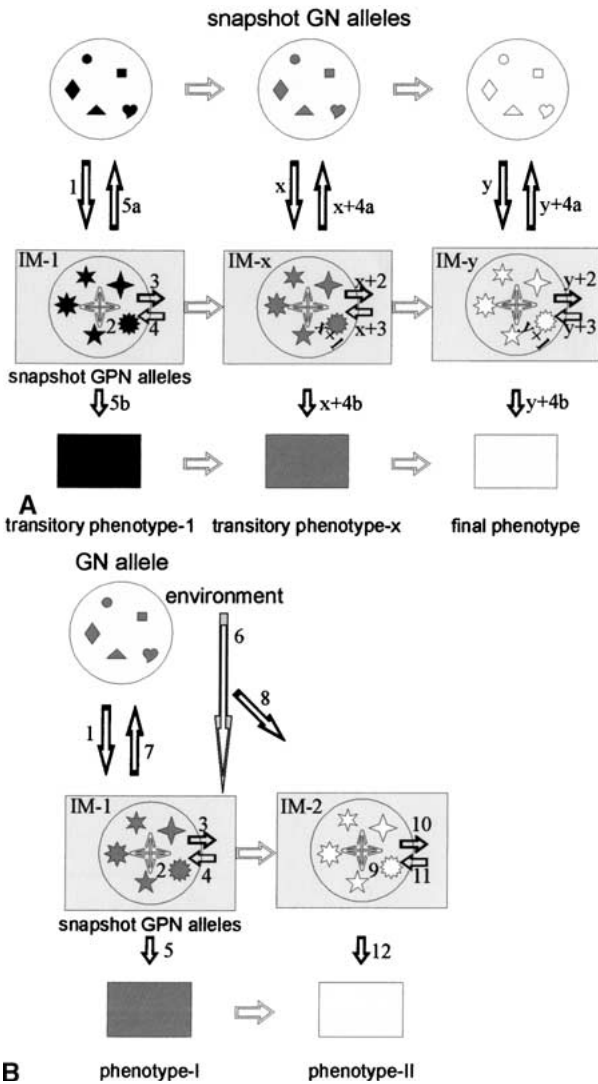


Fig. 2. Determination of phenotype. **A** The succession of snapshot GN and GPN alleles are controlled by their predetermined interactions. I term a trait or a behavior a transitory phenotype until it has reached its final state during ontogenesis. Although embryogenesis is directed by dynamic interacting processes, it shows a very conservative flow of progression. The reason for this is that all of the forthcoming interactions are “precalculated” in the genetic program, which is highly protected from perturbations. A given form with different shades represents different genes or *geps* (stars). **B** Mammalian brain states adaptively change according to alteration of the environment but depending on its former state.

course of time; some genes join, while others skip out from a particular functional network. An important question is the role of the environment and the IM in this process. The example of identical twins indicates that the genetic program very conservatively encodes the body plan, implying that the environment does not play a significant role in the normal morphogenesis, although some examples for the developmental plasticity exist: polyphenism, e.g., cast or sex determination in bees and certain fish species, respectively, or the effect of feeding on grasshoppers development. That is, a very rigid relationship appears to exist between GNs and GPNs during the

course of ontogenesis. However, development is a continuous exquisite interplay among GNs, GPNs, and the IM. That is, this hard-wired determination is programmed by a series of dynamically regulated steps. Why, then, are identical twins identical? They are so because the interplay between genes and the IM also occurs according to a preprogrammed cascade of events which exhibit a high resistance against a high range of environmental effects. However, IM can be experimentally perturbed, which can result in abnormal morphogenesis (Liu et al. 1998).

(b) In contrast to body plan, mental processes of higher-order organisms exhibit high flexibility in their response to external effects (Fig. 2b). For example, in humans, mental events are determined both by hard-wired ontogenesis and by the historical personal development of the individual and the society and culture in which that individual is embedded. A certain composition of a GN usually determines a specific GPN and phenotype (Fig. 3, I). However, in the brain, the GN–GPN relation is affected by the environment, resulting in the redundant coding of GPNs by GNs (Fig. 3, II). Furthermore, it is proposed by the interactionist philosophy that the mind (phenotype) can causally interact with the body (firing pattern of specific neurons governed by certain GPNs) (Popper and Eccles 1986). In addition, in some reported cases GPNs had the ability to stabilize their states (Fig. 3, III). Another example is the compensatory effect observed in mutant and knock-out animals. In this case both the GNs and the GPNs undergo changes due to mutation, however, the phenotype remains unchanged (Fig. 3, IV). Thus, GNs and GPNs are dynamic and self-organizing systems, which can respond to the various effects derived from both the genetic composition and the internal or external environment in a variety of ways.

The Unit of Selection

The paradox in the problem of *the unit of selection* is that natural selection favors individuals with adaptive phenotypes, while the genetic material, and not the phenotype, is transmitted to the next generation. Therefore, the question of how DNA encodes the phenotype is critical. Traits and behaviors are specified by the concerted action of a particular set of gene products interacting at various organizational levels, including molecular, cellular, and higher-level modules, in both direct and indirect manners, and through mediators such as hormones, neurotransmitters, electric impulses, signaling ligands etc. The contribution of a given gene to the specification of a certain trait depends on the composition of the specific GN. In other words, the expression profile of a particular gene is dependent on the overall expression properties of the

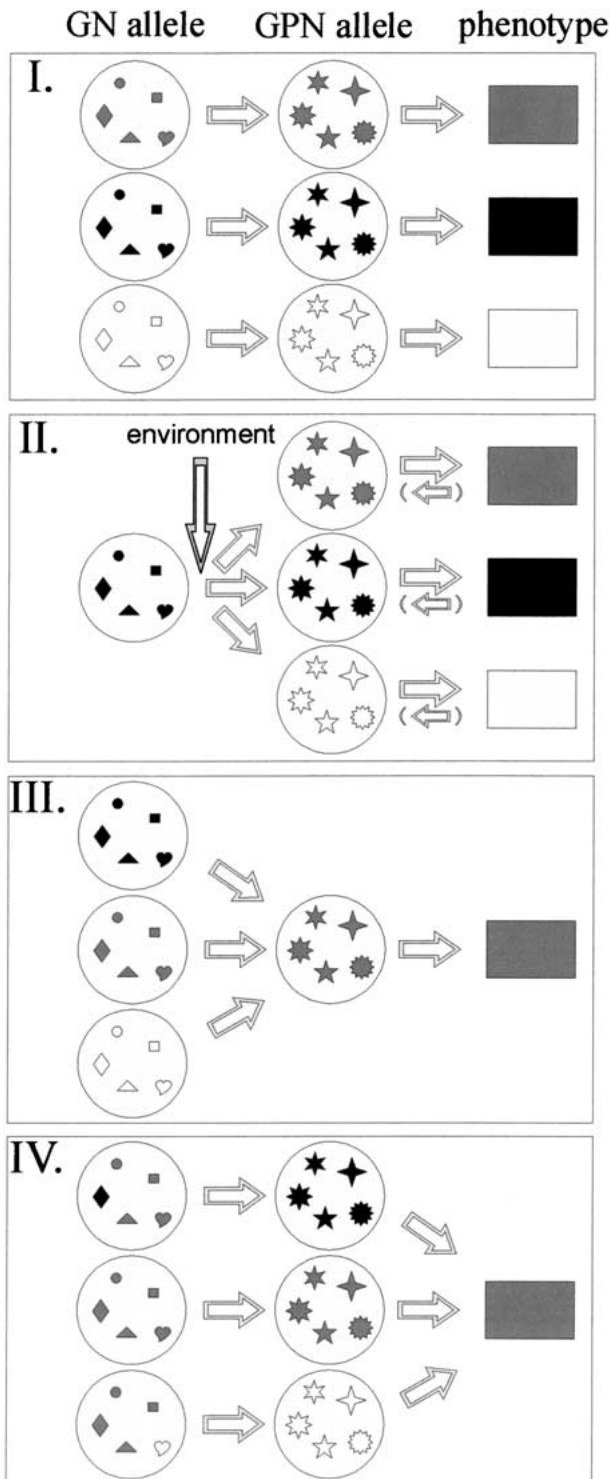


Fig. 3. Gene network–gene product network–phenotype. **I** A given GN allele determines a certain GPN allele, which in turn specifies a certain phenotype. There are no stabilizing effects to maintain uniform GPNs or phenotypes. **II** rain mechanisms were evolved to have high flexibility for responding to environmental stimuli. Therefore, a certain GN allele can specify multiple GPNs and, hence, phenotypes. A potential two-way interaction between the body (GPN, firing patterns of neurons) and mind (phenotype) (downward causation) has been suggested by the “interactionalist” philosophy. GPNs can act **(III)** to preserve their identity or **(IV)** to produce the same phenotype from different GPN compositions. **IV** represents the compensatory effects observed, for example, in knockout animals. Since they were produced using inbred mouse strains, GN alleles differ in only one element in this figure. The system properties presented in **III** and **IV** are consequences of evolution resulting from resistance to environmental and developmental perturbations.

state, including the earliest events of life such as maternal effects and imprinting (epigenetics).

The Origin and Nature of Genetic Variation

Population genetics regards mutations as the sole sources of variation and suggests that the role of the sex is restricted to an intermingling of the genomes in order to bring together favorable mutations in the various genes of different individuals. According to the selfish GN hypothesis, the sex has an additional function: it generates novel phenotypic variants by continuously mixing the compositions of the GNs, thereby allowing chromosomal rearrangement and recombination. The sex can also decrease the variance by converting several manifested GN alleles to latent GN alleles in the next generation. Hence, if we compare two consecutive generations of a population, we may not observe a net increase in the variance. However, in contrast with gene variants, newer and newer GN alleles appear from generation to generation, even in the absence of new mutations (Fig. 4). As a consequence, at a given period spanning several generations, the total variance of manifested GN alleles will be far higher than the allelic diversity of the genes. Additionally, *gevs* form clusters (GNs), and the variation of the clusters is much higher than the variability of their components alone. Selection and genetic drift both act to decrease genetic diversity. Linkage disequilibrium and assortative mating cause only nonrandom distributions in the population genetic models. However, the selfish GN hypothesis claims that they both decrease the manifested genetic variance, since they restrict free combinations of *gevs*. The nonrandom association of genes belonging to a functional gene cluster would result in the co-inheritance of particular *gevs*. Colocalization of functionally related genes in the genome is evolutionarily advantageous for keeping favorable *gev* variants together; however, it is unfavorable for the efficient incorporation of newly appeared beneficial variants, because the close proximity reduces the frequency of

other components of the GN to which they belong. In addition, biological systems have a “history” component, due to the fact that they undergo ontogenesis. This means that, for predicting future events, it is not enough to know the interacting elements and the rules of communication, but in most cases, we also have to be aware of the earlier events that created the present

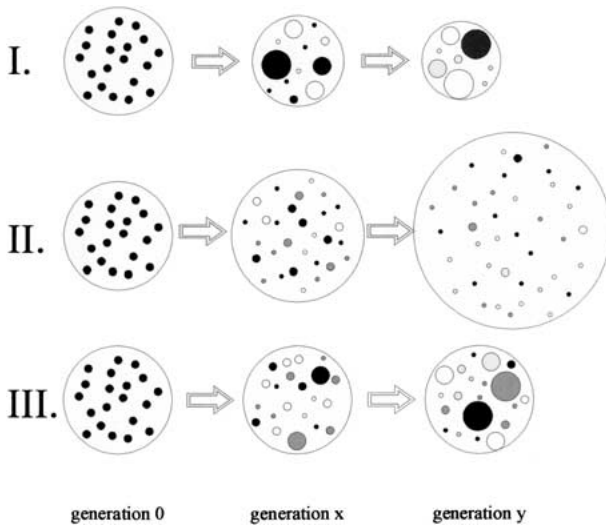


Fig. 4. Gene network evolution. For the sake of simplicity, in the first stage (generation 0), GNs comprise an equal frequency of manifested GN alleles. However, the large portion of GN alleles is latent. **I** Evolution without new mutations. In the absence of new mutations, novel GN alleles can appear as a result of sexual reproduction. However, the number of manifested GN variants will be decreased, and their frequency will be unevenly distributed due to selection, genetic drift, assortative mating, and linkage disequilibrium. In generation x , several latent GN alleles become manifested due to chromosomal rearrangement and recombination, but others become or remain latent. There are two possibilities regarding the change of the frequencies of GN alleles. (1) If selection and random drift do not result in the elimination of any *gevs*, then the total number of GN alleles will remain the same, but the ratio of manifested to latent variation will be decreased. (2) If selection and random drift eliminate some *gevs*, then the number of both manifested and latent variants will be decreased and some GN alleles will irreversibly disappear. **II** Evolution without selection and genetic drift. If only mutations occur, the number of both GN alleles and GPN variants will be increased, but their frequency will be decreased. **III** Evolution with selection, genetic drift, and mutation. Novel GN alleles are produced by both mutation and sexual reproduction. The *gevs* can be eliminated by selection and random drift. The size of the circles corresponds to the frequency of specific GN alleles. Black, GN alleles at generation 0; white, new GN alleles at generation x ; light gray, new GN alleles at generation y ; dark gray, new mutations.

recombination. Functional gene clustering is common in bacteria, whereas in eukaryotes it appears to be the exception rather than the rule (Lawrence 1999). Thus, these two groups of organisms appear to pursue two different evolutionary strategies with respect to the preservation of the integrity of GN variants. Geographical isolation and other ecological situations evoking “bottleneck” effects very probably play eminent roles in the restriction of the genetic variance of GNs and thereby in the creation of characteristic morphotypes of particular phylogenetic groups. The variance of GPNs and phenotypes is also decreased by the self-regulatory nature of specific GNs.

Together, the selfish GN hypothesis claims that, instead of a single gene, a GN encodes a particular phenotype. Further, gene products interact, and the outcome of this interaction is determined by the genetic composition of the CRSs of the genes making up GNs.

Microevolution of Gene Networks

The microevolution of gene networks involves the selective survival and reproduction of individuals possessing various GN alleles. Due to the effect of selection and genetic drift, the particular GNs tend to become uniform. However, this state will never be reached, because GNs are continuously intermingled by the sex, traits are predominantly neutral, selection is usually too soft, environmental changes continuously require novel compositions of GPNs, and new CRS mutations frequently arise (Fig. 4). Accordingly, natural selection acts constantly to improve the composition of specific GNs in the sense of the Red Queen hypothesis (Van Valen 1973), which claims that adaptivity is not increased in an absolute sense over time but simply follows the ever-changing environment. Compared to the gene-centered evolution of population genetics, GN evolution has basically different characteristics, which are summarized in Table 4. Evolution is not solely a reflection of the competition between alleles differing in fitness, but the continuous reorganization of the GN structure, driven by mutation, selection, genetic drift, and sex. In most cases, however, these forces will result in the generation of GNs with novel compositions (not necessarily with lower polymorphism), not in the fixation of a certain GN allele. Two seemingly opposing phenomena can be observed in nature: (a) in natural populations, especially in higher-order species, practically every individual is unique, indicating a huge genetic variation; (b) different species and higher taxa exhibit typical morphological characters, indicating that the genetic variance is somehow canalized. Dog breeds, representing a great number of morphotypes, might well be ideal experimental models for analysis of the role of genetic variance in developmental and, for instance, neural GNs and their role in evolution. The existence of a high number of latent variations means that significant evolutionary changes may occur without the involvement of new mutations, simply through a shuffling of the genomes by the sex, followed by a selection for favorable GNs and/or the elimination of a substantial degree of variance by pure chance (in bottleneck situations). However, the significance of this type of evolution cannot yet be estimated.

Together, evolution proceeds in a fluid-like manner, by continuously restructuring the composition of the GNs. Natural selection does not act purely to increase the frequency of certain GN alleles, but also continuously to improve its composition and, in turn, certain strategies (not only behavior). This is achieved by pushing the GN compositions in certain directions without a preset “winning post” and without a homogeneous state ever being reached. This point is a

Table 4. Comparison of selfish gene network hypothesis with population genetics

| Terms | Population genetics | Selfish GN hypothesis |
|---|--|---|
| The genetic source of natural selection | Natural variance in the coding region of genes | Natural variance in CRSs of genes forming gene networks |
| The target of natural selection | Traits encoded by single genes or, by the additive effect of several genes | Traits encoded by interacting genes pertaining to gene networks |
| The unit of evolutionary change | Genes | Gene networks |
| Mechanism of natural selection | Differential transmission of gene variants to the next generation | Changing of the genetic composition of a gene network across generations |
| The result of natural selection | Enrichment or fixation of an allele | Continuous reorganization of GN structure |
| The concept of fitness | Central | Difficult to operationally use |
| Generation of varieties | Mutation | Mutation and sex |
| Interaction among genes | Confusing | Essential |
| Coevolution of genes | Not included | Essential |
| Significance of gene evolution | Fundamental | Marginal in recent evolution |
| Extent of genetic variation | Low | Very high |
| Extent of adaptive variation | Only neutral or harmful mutations | Presumably high |
| Hidden phenotypic variation | Recessive traits | Latent GN alleles |
| The nature of quantitative traits | Sum or average of discrete traits | Continuous |
| The role of the sex | Assembles favorable mutations occurring in different individuals | Assembles favorable mutations and generates novel phenotypes |
| Fitness optimum | Exists | Does not exist |
| Linkage disequilibrium | Nonrandom distribution of alleles | Deplete variance |
| Assortative mating | Nonrandom distribution of alleles | Deplete variance |
| Genetic load | Problem | Does not exist |
| Evolution without mutation | Minor steps | Potentially significant steps |
| “Hopeful monsters” | Irrelevant | Singular developmental events can create them. They might be the progenitors of macroevolutionary jumps. |
| Interest bearers (in a metaphoric sense) | Genes (more precisely alleles) | Specific GNs directly related to reproduction success |

vital difference between the allele competition-centered approaches (population genetics and selfish gene theory) and the selfish GN hypothesis.

Interest Bearers

Dawkins' selfish gene theory considers genes as *real interest bearers*. In this concept, the identical alleles have a common interest: to win over other alleles. Clearly, this struggle, if it exists, is not performed by each gene of an organism, nor is it the joint effect of all of the genes of the genome. The phrase “interest bearer” is only a metaphor, which should be applied to those strategies which are directly related to reproduction, self-defense, or the “altruistic” defense of progenies and other relatives. These strategies are autonomous and do not require relatedness with the phenotypes encoded by the rest of the genome. In order to distinguish the GNs underlying such strategies from other GNs, I call them self-serving GNs. In a metaphoric sense, interest bearers (self-serving GNs) force other GNs to be beneficial in their struggle. Individuals appear to strive to increase the chance of the transmission of their own genetic material in the next generation of a population.

However, this struggle is performed by only a small number of GNs. Hence, while various self-serving GN variants tend to increase their frequency in a population, the rest of the genome (other GNs) “benefits” from this struggle simply because of a certain strength of physical linkage, but their “interests” do not matter at all. According to the selfish GN hypothesis, if such kinds of interest bearers really exist, they are “interested” in the evolutionary development of their strategy and not in the spreading of their identical copies in the population.

Macroevolution of Gene Networks

“Within-System” Evolution. (a) *Coevolution for interaction*. The members of GNs evolve to interact with each other, thereby forming a dynamic network. Any alteration in the expression of one of their components will induce changes in the expression profiles in several other components, (b) *Coevolution for formation of functional units*. The interactions among genes are canalized in order for the whole network to behave as a system, (c) *Coevolution for robustness*. The GN component evolved to provide functional stability of developmental systems against perturbations derived from the environment or the organisms themselves

(but not to buffer harmful mutation effects or mask natural variance). (d) *Coevolution for adaptive flexibility*. For example, the mammalian brain evolved to be extraordinarily flexible in order to adaptively respond to environmental challenges.

“Between-System” Evolution. (a) *Evolution of combinatorial gene regulatory strategies*. The target of natural selection can be an existing allocation mechanism, which is responsible for the determination of *geps* of a network. This allocation mechanism is primarily based on the combination of transcription factors, which are represented in a limited number in an organisms regulatory repertoire, (b) *The communication among GNs* can also be the target of natural selection.

Priority Shift in Evolution

Proteins are brilliantly designed devices, which had to be created by adaptive evolution. I have no doubts that this was the case at the time of the emergence of new genes. In the early history of life creation of new genes, improvement of their functions and establishment of the rule of interactions among gene products had primary significance. Further on, new priorities have emerged: the regulation of gene expression and the formation of functionally interrelated gene expression networks.

Some Conclusions

Some conclusions can be drawn on the basis of the selfish GN hypothesis. The paradox of too high a genetic load of the multiple gene fixations in a population does not arise, because in most cases there is no optimal fitness toward which the allele frequencies should be pushed by selection. The seeming contradiction that genes are discrete entities, while phenotypes are predominantly continuously distributed in a population, is no longer a contradiction, since gene expression polymorphism itself is continuous. The selfish GN hypothesis not only accepts the appearance of “hopeful monsters” in singular events, but suggests that such events could play a major role in the emergence of key innovations and in major evolutionary transitions.

Conclusion

We are currently witnessing the emergence of a new epistemology of biology based on the recognition that understanding the operation of biological systems requires system-level analysis. Currently, our conception on the mechanism of evolution is in the paradigm trap created by the gene-centric view of population genetics. However, novel results in various fields of molecular biology support the modular

organization paradigm, which calls for the reevaluation of contemporary evolutionary biology. The fundamental assertions of the selfish GN hypothesis are as follows. (1) GNs, and not individual genes, are the units of natural selection. (2) The genetic polymorphism of GNs is produced by the genetic variance in the regulatory regions of the network components. (3) Natural phenotypic diversity is generated by the genetic polymorphism and by the interaction among the components. (4) Evolution proceeds by continuously restructuring the GNs, and not by fixing certain allelic variants. Hence, the concept of “fitness” in the sense used by population genetics is useless for describing evolution. The proposed conceptual framework will hopefully provide an operational basis for novel mathematical approaches for modeling microevolution and perhaps macroevolutionary events.

Glossary

Antisense (a) RNA: Transcribed from the complementary DNA strand and generally playing a role in the block of translation from the mRNA transcribed from the same DNA stretch.

Epistasis: The condition where one gene has an effect on the expression of another gene.

Forward genetics: An experimental approach where the aim of the investigation is to find the genetic background of a phenotype obtained by spontaneous or induced mutations.

Hypomorphic mutation: A mutation in the regulatory sequence of a gene that does not completely abolish gene expression from the affected gene.

Knockout technology: Targeted gene ablation by means of homologous or site-specific recombination.

Knockdown technology: Posttranscriptional gene silencing carried out by small interfering (si)RNAs.

Loss-of-function mutation: The gene function is completely abrogated by mutation.

Micro (mi)RNA: An endogenous posttranscriptional down-regulation mechanism achieved by small RNAs with a hairpin structure.

Ortholog: A gene from one species which corresponds to a gene in another species.

Pleiotropy: A mutation which affects the operation of other genes.

Proteome: The full sets of proteins encoded by the genome.

Reverse genetics: An experimental approach where the first step is a modification of gene expression (mutation, overexpression), followed by analysis of the phenotype obtained.

Ribozyme: An RNA with enzyme function.

Robustness: The ability of a system to retain its original state in spite of receiving variable inputs.

Elements of the selfish gene network hypothesis

| | |
|--|--|
| CRSs (cis-regulatory sequences) | Promoters, enhancers, and other DNA sequences, such as MAR, LAR, SAR |
| TRFs (trans-regulatory factors) | Gene products (transcription factors and other regulatory proteins, aRNAs, miRNAs, ribozymes), which directly regulate gene expression or process function |
| <i>gevs</i> (gene expression variants) | Variants of a specific gene differing in their CRSs |
| <i>gep</i> (gene expression profile) | The expression level and spatiotemporal distribution of a protein or RNA encoded by a particular <i>gev</i> |
| Gene | Directs the production of a protein |
| Gene network (GN) | A set of genes connected by regulatory linkages. Functional modules behave as systems in their operation. |
| GPN (gene product network) | The expression level and spatiotemporal distribution of proteins or RNAs encoded by <i>gevs</i> belonging to a particular GN |
| Allele | Gene variants differing in their coding or regulatory sequences |
| Gene Network (GN) allele | A particular composition of GNs by <i>gevs</i> |
| GPN alleles | A particular composition of GPNs by <i>geps</i> |
| Manifested GN alleles | The pool of GN alleles existing in a certain generation of a population |
| Latent ON slides | Nonmanifested <i>gev</i> combinations in a certain generation of a population |
| Snapshot GN (GPN) | A subset of a GN (GPN), comprising only those genes (<i>gevs</i> , <i>geps</i>) which interact at a given time. |
| Transitory phenotypes | Phenotypes appear transiently during ontogenesis. |
| Final phenotypes | Phenotypes have reached their ultimate structure and function. |

Open reading frame (ORF): A part of the mRNA (or cDNA) that spans from ATG to the first stop codon. It encodes the exons of a gene.

Transcriptome: The full sets of RNAs encoded by the genome.

References

- Averof M, Patel NH (1997) Crustacean appendage evolution associated with changes in Hox gene expression. *Nature* 388:682–686
- Bailey A, Ledet C, Kelly M et al. (2002) Changes in spinal delta and kappa opioid systems in mice deficient in the A2A receptor gene. *J Neurosci* 22:9210–9220
- Bejar R, Yasuda R, Krugers H et al. (2002) Transgenic calmodulin-dependent protein kinase II activation: Dose-dependent effects on synaptic plasticity, learning, and memory. *J Neurosci* 22:5719–5726
- Belting HG, Shashikant CS, Ruddle FH (1998) Modification of expression and cis-regulation of *Hoxc8* in the evolution of diverged axial morphology. *Proc Natl Acad Sci USA* 95:2355–2360
- Bhan A, Galas DJ, Dewey TG (2002) A duplication growth model of gene expression networks. *Bioinformatics* 18:1486–1493
- Blalock JE (1990) Complementarity of peptides specified by sense and antisense strand of DNA. *Trends Biotechnol* 8:40–144
- Boldogkői Z (2000) Coding in the noncoding DNA strand: A novel mechanism of gene evolution?. *J Mol Evol* 51:600–606
- Boldogkői Z, Barta E (1999) Specific amino acid content and codon usage account for the existence of overlapping ORFs. *Biosystems* 51:95–100
- Boldogkői Z, Murvai J (1994) A novel explanation for the existence of open reading frames on latency-associated transcripts of alphaherpesviruses. *Virus Genes* 9:47–51
- Boldogkői Z, Nógrádi A (2003) Gene and cancer therapy—Pseudorabies virus: A novel research and therapeutic tool?. *Curr Gene Ther* 3:155–182
- Boldogkői Z, Murvai J, Fodor I (1995) G and C accumulation at silent positions of codons produces additional ORFs. *Trends Genet* 11:125–126
- Boldogkői Z, Schütz B, Sallach J et al. (2002) P2X(3) receptor expression at early stage of mouse embryogenesis. *Mech Dev* 118:255–260
- Bonner JT (1988) The evolution of complexity. Princeton University Press, Princeton, NJ
- Brown SJ, Patel NH, Denell RE (1994) Embryonic expression of the single *Tribolium* engrailed homolog. *Dev Genet* 15:7–18
- Brown S, Holtzman S, Kaufman T et al. (1999) Characterization of the *Tribolium* Deformed ortholog and its ability to directly regulate Deformed target genes in the rescue of a *Drosophila* Deformed null mutant. *Dev Genes Evol* 209:389–398
- Burrow CR (2000) Regulatory molecules in kidney development. *Pediatr Nephrol* 14:240–253
- Bymaster FP, Zhang W, Carter PA et al. (2002) Fluoxetine, but not other selective serotonin uptake inhibitors, increases norepinephrine and dopamine extracellular levels in prefrontal cortex. *Psychopharmacology* 160:353–361
- Calon F, Morissette M, Ghribi O et al. (2002) Alteration of glutamate receptors in the striatum of dyskinetic 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine-treated monkeys following dopamine agonist treatment. *Prog Neuropsychopharmacol Biol Psychiatry* 26:127–138
- Cantile M, Prociò A, D'armentino M et al. (2003) HOX gene network is involved in the transcriptional regulation of in vivo human adipogenesis. *J Cell Physiol* 194:225–236
- Cebat S, Dudek MR (1996) Generation of overlapping open reading frames. *Trends Genet* 12:12
- Claverie J, Ogata H (2003) The insertion of palindromic repeats in the evolution of proteins. *Trends Biochem Sci* 28:75–80
- Cowles CR, Hirschhorn JN, Altshuler D, Lander ES (2002) Detection of regulatory variation in mouse genes. *Nat Genet* 32:432–437
- Cripps RM, Olson EN (2002) Control of cardiac development by an evolutionarily conserved transcriptional network. *Dev Biol* 246:14–28
- Davidson EH, Rast JP, Olivery P et al. (2002) A genomic regulatory network for development. *Science* 295:1669–1678
- Dawkins R (1976) The selfish gene Oxford University Press, Oxford
- de la Pompa JL, Wakeham A, Correia KM et al. (1997) Conservation of the Notch signaling pathway in mammalian neurogenesis. *Development* 124:1139–1148
- Delattre M, Felix MA (2001) Polymorphism and evolution of vulval precursor cell lineages within two nematode genera, *Caenorhabditis* and *Oscheius*. *Curr Biol* 11:631–643
- Dickinson WJ (1980) Evolution of patterns of gene expression in hawaiian picture-winged *Drosophila*. *J Mol Evol* 16:73–94

- Eldredge N, Gould SJ (1972) Punctuated equilibria: An alternative to phyletic gradualism In: Models in paleobiology, Freeman, Copper and Co., San Francisco
- Erb C, Troost J, Kopf S et al. (2001) Compensatory mechanisms enhance hippocampal acetylcholine release in transgenic mice expressing human acetylcholinesterase. *J Neurochem* 77:638–646
- Falciani F, Hausdorf B, Schroder R, Akam M, Tautz D, Denell R, Brown S (1996) Class 3 Hox genes in insects and the origin of zen. *Proc Natl Acad Sci USA* 93:8479–8484
- Featherstone DE, Broadie K (2002) Wrestling with pleiotropy: Genomic and topological analysis of the yeast gene expression network. *BioEssays* 24:267–274
- Félix MA, Sternberg PW (1997) Two nested gonadal inductions of the vulva in nematodes. *Development* 124:253–259
- Félix MA, De Ley P, Sommer RJ, Frisse L, Nadler SA, Thomas WK, Vanfleteren J, Sterberg PW (2002) Evolution of vulva development in the Cephalobina (Nematoda). *Dev Biol* 221:68–86
- Fisher RA (1930) The genetical theory of natural selection Oxford University Press, Oxford,
- Galant R, Carroll SB (2002) Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* 415:910–913
- Galis F, van Dooren TJ, Metz JA (2002) Conservation of the segmented germband stage: robustness or pleiotropy?. *Trends Genet* 18:504–509
- Gallet A, Angelats C, Kerridge S, Therond PP (2000) Cubitus interruptus-independent transduction of the Hedgehog signal in *Drosophila*. *Development* 127:5509–5522
- Grbic M, Nagy LM, Carroll S B, Strand M (1996) Polyembryonic development: Insect pattern formation in a cellularized environment. *Development* 122:795–804
- Gerlai R (1996) Gene-targeting studies of mammalian behavior: Is it the mutation or the background genotype?. *Trends Neurosci* 19:177–181
- Grenier JK, Carroll SB (2000) Functional evolution of the Ultrabithorax protein. *Proc Natl Acad Sci USA* 97:704–709
- Horton WA (2003) Skeletal development: insights from targeting the mouse genome. *Lancet* 362:560–569
- Ihmels J, Friedlander G, Bergmann S et al. (2002) Revealing modular organization in the yeast transcriptional network. *Nat Genet* 31:370–377
- Jareborg N, Birney E, Durbin R (1999) Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. *Genome Res* 9:815–824
- Jeong H, Tombor B, Albert R et al. (2000) The large-scale organization of metabolic networks. *Nature* 407:651–654
- Keranen SV, Kettunen P, Aberg T et al. (1999) Gene expression patterns associated with suppression of odontogenesis in mouse and vole diastema regions. *Dev Genes Evol* 209:495–506
- Kest B, Hopkins E, Palmese CA et al. (2001) Morphine tolerance and dependence in nociceptin/orphanin FQ transgenic knockout mice. *Neuroscience* 104:217–222
- Keys DN, Lewis DL, Selegue JE et al. (1999) Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science* 283:532–534
- Keyvani K, Witte OW, Reinecke S et al. (2002) Use-related gene expression patterns of rat motor cortex. *Neuroreport* 13:953–956
- Kimura M (1968) Evolutionary rate at the molecular level. *Nature* 217:624–626
- Lawrence JG, Roth JR (1996) Selfish operons: Horizontal transfer may drive the evolution of gene clusters. *Genetics* 143:1843–1860
- Liu A, Joyner AL, Turnbull DH (1998) Alteration of limb and brain patterning in early mouse embryos by ultrasound-guided injection of Shh-expressing cells. *Mech Dev* 75:107–115
- Ludwig MZ, Bergman C, Patel NH et al. (2000) Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403:564–567
- MacGowan GA, Du C, Wieczorek DF et al. (2001) Compensatory changes in Ca(2+) and myocardial O(2) consumption in beta-tropomyosin transgenic hearts. *Am J Physiol Heart Circ Physiol* 281:H2539–H2548
- Middlemiss DN, Price GW, Watson JM (2002) Serotonergic targets in depression. *Curr Opin Pharmacol* 2:18–22
- Mitchison NA, Muller B, Segal RM (2000) Natural variation in immune responsiveness, with special reference to immunodeficiency and promoter polymorphism in class II MHC genes. *Hum Immunol* 61:177–181
- Mitsialis SA, Kafatos FC (1985) Regulatory elements controlling chorion gene expression are conserved between flies and moths. *Nature* 317:453–456
- Müller H, Samanta R, Wieschaus E (1999) Wingless signaling in the *Drosophila* embryo: Zygotic requirements and the role of the frizzled genes. *Development* 126:577–586
- Nieoullon A (2002) Dopamine and the regulation of cognition and attention. *Prog Neurobiol* 67:53–83
- Okazaki Y et al. (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420:563–573
- Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nat Genet* 32:261–266
- Patel NH, Kornberg TB, Goodman CS (1989) Expression of engrailed during segmentation in grasshopper and crayfish. *Development* 107:201–212
- Patthy L (1999) Genome evolution of exon-shuffling—A review. *Gene* 238:103–114
- Podani J, Oltvai ZN, Jeong H et al. (2001) Comparable system-level organization of Archaea and Eukaryotes. *Nat Genet* 29:54–56
- Popper KR, Eccles JC (1986) The self and its brain Routledge, New York,
- Raff R (1996) The shape of life development and the evolution of animal form University of Chicago Press, Chicago
- Romano LA, Wray GA (2003) Conservation of Endo16 expression in sea urchins despite evolutionary divergence in both cis- and trans-acting components of transcriptional regulation. *Development* 130:4187–4199
- Ronshaugen M, McGinnis N, McGinnis W (2002) Hox protein mutation and macroevolution of the insect body plan. *Nature* 415:914–917
- Rutherford SL, Lindquist S (1998) Hsp90 as a capacitor for morphological evolution. *Nature* 396:336–342
- Schmid G, Wang ZQ, Wesierska-Gadek J (1999) Compensatory expression of p73 in PARP-deficient mouse fibroblasts as response to a reduced level of regularly spliced wild-type p53 protein. *Biochem Biophys Res Commun* 255:399–405
- Schütz B, Weihe E, Eiden LE (2001) Independent patterns of transcription for the products of the rat cholinergic gene locus. *Neuroscience* 104:633–642
- Shashikant CS, Kim CB, Borbely MA et al. (1998) Comparative studies on mammalian Hoxc8 early enhancer sequence reveal a baleen whale-specific deletion of a cis-acting element. *Proc Natl Acad Sci USA* 95:15446–15451
- Simmonds AJ, Brook WJ, Cohen SM, Bell JB (1995) Distinguishable functions for engrailed and invected in anterior-posterior patterning in the *Drosophila* wing. *Nature* 376:424–427
- Simpson GG (1944) Tempo and mode in evolution Columbia University Press, New York
- Spitz F, Gonzalez F, Peichel C et al. (2001) Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. *Genes Dev* 15:2209–2214
- Steiner H, Bonner TI, Zimmer AM et al. (1999) Altered gene expression in striatal projection, neurons in CBI cannabinoid receptor knockout mice. *Proc Natl Acad Sci USA* 96:5786–5790

- Stephens JC, Schneider JA, Tanguay DA et al. (2001) Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293:489–493
- Stern MJ, Marengere LE, Daly RJ et al. (1993) The human GRB2 and *Drosophila* Drk genes can functionally replace the *Caenorhabditis elegans* cell signaling gene sem-5. *Mol Biol Cell* 4:1175–1188
- Sucena E, Stern DL (2000) Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of *ovo/shaven-baby*. *Proc Natl Acad Sci USA* 97:4530–4534
- Takahashi H, Mitani Y, Satoh G et al. (1999) Evolutionary alterations of the minimal promoter for notochord-specific Brachyury expression in ascidian embryos. *Development* 126:3725–3734
- Ting CT, Tsauro SC, Wu ML et al. (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282:1501–1504
- True JR, Carroll SB (2002) Gene co-option in physiological and morphological evolution. *Annu Rev Cell Dev Biol* 18:53–80
- Van Valen L (1973) A new evolutionary law. *Evol Theory* 1:1–30
- von Dassow G, Munro E (1999) Modularity in animal development and evolution: elements of a conceptual framework for EvoDevo. *J Exp Zool* 285:307–325
- von Dassow G, Meir E, Munro EM et al. (2000) The segment polarity network is a robust developmental module. *Nature* 406:188–192
- Wagner GP (1996) Homologues, natural kinds and the evolution of modularity. *American Zoologist* 36:36–43
- Wasserman WW, Palumbo M, Thompson W et al. (2000) Human-mouse genome comparisons to locate regulatory sites. *Nat Genet* 26:225–228
- Waterston RH, Lindblad-Toh K, Birney E et al. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562
- Weatherbee SD, Halder G, Kim J et al. (1998) Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the *Drosophila* haltere. *Genes Dev* 12:1474–1482
- Wilson ME, Scheel D, German MS (2003) Gene expression cascades in pancreatic development. *Mech Dev* 120:65–80
- Wyrick JJ, Young RA (2002) Deciphering gene expression regulatory networks. *Curr Opin Genet Dev* 12:130–136
- Xu PX, Zhang X, Heaney S et al. (1999) Regulation of Pax6 expression is conserved between mice and flies. *Development* 126:383–395
- Xue ZG, Xue XJ, Quox-1, LeDourin NM (1993) Quox-1, an Antp-like homeobox gene of the avian embryo: A developmental study using a Quox-1-specific antiserum. *Mech Dev* 43:149–158
- Yelin R, Dahary D, Sorek R et al. (2003) Widespread occurrence of antisense transcription in the human genome. *Nat Biotechnol* 21:379–386
- Young SN, Leyton M (2002) The role of serotonin in human mood and social interaction. Insight from altered tryptophan levels. *Pharmacol Biochem Behav* 71:857–865
- Yomo T, Urabe I, Okada H (1992) No stop codons in the antisense strands of the genes for nylon oligomer degradation. *Proc Natl Acad Sci USA* 89:3780–3784
- Zull JE, Smith SK (1990) Is genetic redundancy related to retention of structural information in both DNA strands?. *Trends Biochem Sci* 15:257–261