Is the Medium the Message? Biological Traits and Their Regulation

KENNETH WEISS

Much about the nature of biological processes is logically necessary but functionally arbitrary. What does this mean about biology?

We live in the information age. In the 1960s when Marshall McLuhan said the medium is the message, he meant that in our electronic times the medium itself was an important carrier of our total cultural message. The *way* we acquire information may affect us more than the information itself. What sometimes matters most is not the product, but getting the product to market, and what is done to get it there may bear little resemblance to the product itself. The purpose of advertising is to manipulate the individual, McLuhan said, making us want to become what we behold. He used automobile marketing tactics as an example, and I had my own '60s experience of that. Single, young, and frisky, I was lured into buying the latest model little red sports car, lured I'm sure by beautiful women in the ads (Figure 1, model not shown). I was sure I'd get the girls!

Science, too, has become a commodity to be sold via the media. But there is another sense in which biology and culture converge. Evidence is accumulating that a comparably large fraction of biological activity itself is used simply to bring phenotypes to the evolutionary competitive market. This fact may affect our perception of what's going on in evolution, and how highly specific traits—such as those of central interest to biological anthropology—are produced by surprisingly arbitrary mechanisms.

Thousands of proteins are used in complex ways in the 4-dimensions of space and time even within a single cell, all specified by a universal, compact DNA coding system. The code metaphor may be overstated (Kay, 2000), but information is certainly stored in DNA sequences so that while a cell or organism may die, its DNA molecules are transmitted from generation to generation. The program outlives the computer. Indeed, we often hear the genome described as the program for "computing" the organism.

Like C++ or Basic, the genetic system is a true *code* in that there is no direct functional connection between coding mechanism and protein structure. Nothing about the nucleotide triplets that code for the amino acid threonine relates to the chemical properties of threonine. Nor does the triplet correspond to the traits in which it's used. The same nucleotide triplets code for the same set of amino acids, whether the protein makes hair or bone or leaves, binds oxygen, produces color, responds to light, or gives cell walls their rigidity. Nor has it anything to do with being human, oak, or slime mold. The coding system is *logically necessary,* because we need to make (say) hemoglobin, but is *functionally arbitrary,* because it doesn't matter what codes for it. For example, four different triplet codons specify threonine (**ACT ACC**, **ACA**, **ACG**), but a protein that needs a threonine, doesn't care which codon is used to specify the threonine.

The separation of coding from function allows the coding system to work quietly in the background, so evolutionary biologists can concentrate on what has traditionally been viewed as the real aspects of life—morphology, eating, escaping, breathing, flying, and what we do with people attracted by our sports cars—without having to worry directly about how the gene products related to those traits were made. Natural selection may screen proteins for a threonine in some particular position, but it doesn't care how it got there.

How genes code for proteins is well understood, but what makes multicellular life possible is differentiation. This has to do with *when* a gene specifies a protein and which genes are expressed in a given context. The timing and context of gene expression involve an additional coding infrastructure, also embedded within the DNA sequence but working in a very different way. Gene regulation greatly increases the amount of functionally arbitrary biological activity that goes on in an organism and raises some interesting evolutionary problems.

DIFFERENTIATION AND COMPLEX ORGANISMS

It may not be appreciated how important the context-specific use of subsets of genes is to life. Making muscle and making hair are done from the same genome contained in the organism's cells. An organism pro-

Ken Weiss is Evan Pugh Professor of Anthropology and Genetics at Penn State University.

Evolutionary Anthropology 11:88–93 (2002)
DOI 10.1002/evan.10028
Published online in Wiley InterScience (www.interscience.wiley.com).

Figure 1. The medium: Ad for a 1963 Triumph TR4. (Photo courtesy Brian Sanborn. (http://www.net1plus.com/users/sanborn/).

duces multiple tissues and organs via a developmental tree of descent from a single initial cell (fertilized egg) during its lifetime. But how is selective gene expression controlled?

There are nearly countless ways in which the cells, tissues, and organisms differentiate, but the *logic* of the process is shared. As illustrated schematically in Figure 2, gene regulation involves short sequences, called *recognition sequences* (referred to as "promoters" or "enhancers"), in the DNA that flanks the protein-coding region of a gene. A gene is expressed when its enhancer sequences are physically bound by proteins called *transcription factors* (TFs). A TF is a protein coded for by its own gene somewhere in the genome, whose structure leads it to bind to a specific enhancer sequence, because the chemical "shape" of the TF molecule fits the "shape" of DNA with that sequence. Like the "Drink me!" tags on growth-potion bottles in *Alice in Wonderland,* enhancers are a gene's "Bind me!" tags for specific TFs. An enhancer is a sequence-based code in DNA, but one that carries information related to gene expression rather than protein coding.

A common way these elements are used to activate a "target" gene is shown in Figure 3. A cell produces *receptor* molecules specific to its particular function. The receptor resides in the cell membrane, with an extracellular part protruding out of the cell, that "looks" for a circulating *signaling factor* molecule that fits the extracellular binding region. The signaling factors are produced and secreted by other cells, and move through the extracellular space "looking for" a particular receptor. When the signaling factor finds a cell that expresses the receptor of its desires, the two bind together in Holy Regulamony. This event alters the receptor's intracellular part, which triggers a cascade of interaction among *second messenger* molecules that, like hand grenades, wait around inside the cell for their pins to be pulled. Eventually this cascade activates TFs that themselves are also primed and present in the cell, which then bind enhancers upstream of target genes, causing the latter to be expressed.

A dizzying hierarchy of such activity is required before a cell can become a skin or nerve cell (e.g., Figure 4). It is the *combination* of a number of specific TFs that occurs in a given cell that determines what it does. A gene must be flanked by the appropriate set of enhancer sequences for each context in which it's used. But of course, that can only happen in cells that already "know" to produce the right set of TFs, secondary messengers, or receptors. These are all coded for by other genes, which means that earlier in the cell's developmental ancestry similar mechanisms have been used to anticipate its later function. Differentiation is thus hierarchical in developmental time. Like the proverbial fleas, a regulatory cascade has another cascade upon its back to express it, and so *ad infinitum*—back to the fertilized egg.

FUNCTIONALLY ARBITRARY MECHANISMS ARE UBIQUITOUS AND EXTENSIVE

The extent of functionally arbitrary regulatory mechanisms in any organism is vast. There are many families of TF, receptor, signaling factor, and message-transduction genes in our genome. Each family can have tens of members—different genes descended from a common ancestral gene by a history of gene duplication events. And there's more. Many if not most genes are alternatively spliced, that is, use different combinations of their protein coding regions (exons), in different situations; this requires specific splicing mechanisms, coded for by other genes, for each such instance. And yet more: the genome provides code for a whole world of RNA mole-

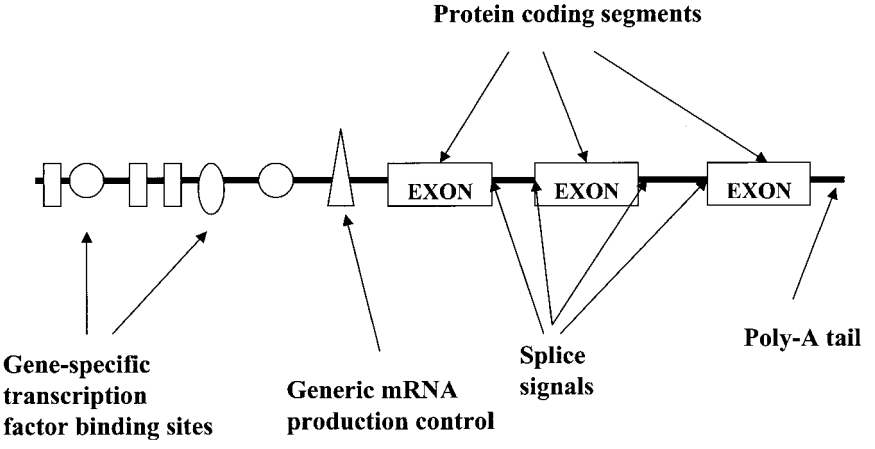

Figure 2. Schematic of information-holding regions of DNA in and near a gene. Shapes symbolize different functional sequence elements along the DNA. The intervening sequence is shown as a line.

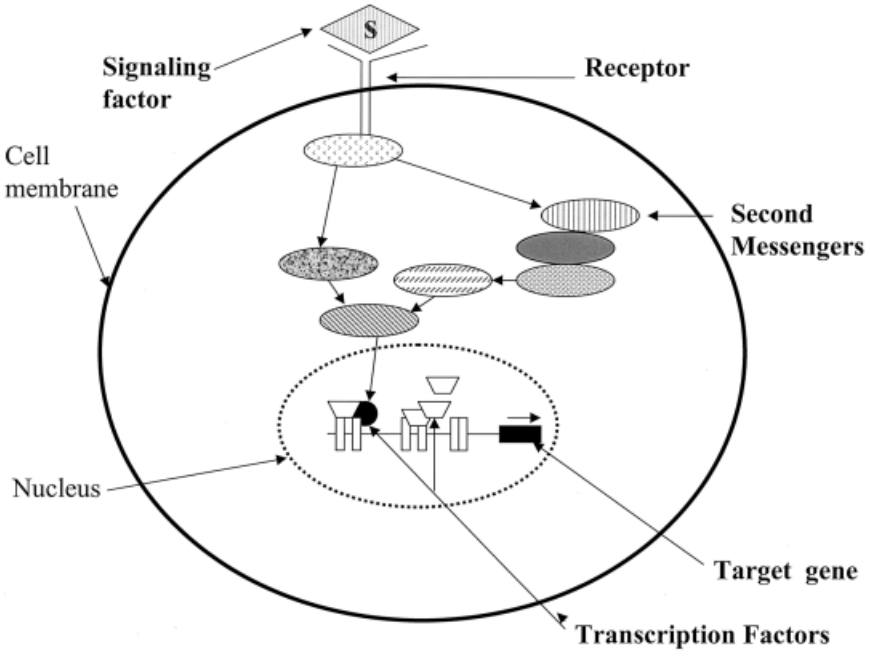

Figure 3. Schematic illustration of the cascade of receptor-mediated induction of gene expression.

cules that have biological functions of their own, based on the way they fold up and interact with other things in the cell. These include the familiar transfer RNA and ribosomal RNA. But more: part of this RNA activity has to do with other sequences scattered in many places in the genome, called imprinting centers, that regulate the expression of nearby genes by modifying their enhancer regions.

There's another strange new world of RNA molecules, also coded for in the genome, whose function appears to be to bind to specific mRNA molecules to regulate their translation into protein, or to degrade them rapidly so as to control delicate developmental timing switches. Such *antisense* genes are common in the genome, and there are also mechanisms to activate antisense-RNAs at context-specific times. Also, the 3′ end of mRNAs can include sequence signals that help localize their coded protein within a cell, using microtubule machinery as a guide, to control its translation into protein, or for other functions.

And did I mention that the packaging of DNA by histone proteins also relates to DNA sequence signals and requires histone producing and processing genes? But that's still not all! A protein that is finally coded by a gene typically has to be molded, combined, protected, transported, cleaved, folded, spindled, or mutilated (are you old enough to remember *that* McLuhan-age phrase?) before it becomes functional. Each of these processing functions is achieved by gene products, and there are many others whose job is to clean up the mess of misformed proteins that error-prone mechanisms leave behind. These genes, of course, require their own regulatory cascades.

The basic mechanisms probably were already present in the earliest cells. Many DNA binding regions for, or interaction pathways among specific regulatory factors seem to be deeply conserved phylogenetically. These "circuits" comprise a tool kit used in so many ways that it's not to be tinkered with. You may vary where you use a screw, but you use the same screwdriver. There is phylogenetic conservation in regulatory pathways associated with many basic functions, like polarity in an early embryo, photoreception, or the differentiation of neural, muscular, heart and other basic cell types. This can have important effects on how we view evolution.

RECONSTRUCTION: THE ANCIENT WORLD

Since the 1800s, biologists have reconstructed ancestral body plans for groups of animals (e.g., the first chordate) by building an hypothetical Cootie with all the shared elements, like four limbs, two eyes, kidneys, muscles, heart, etc. But this may be wrong. Rather than an assembly of

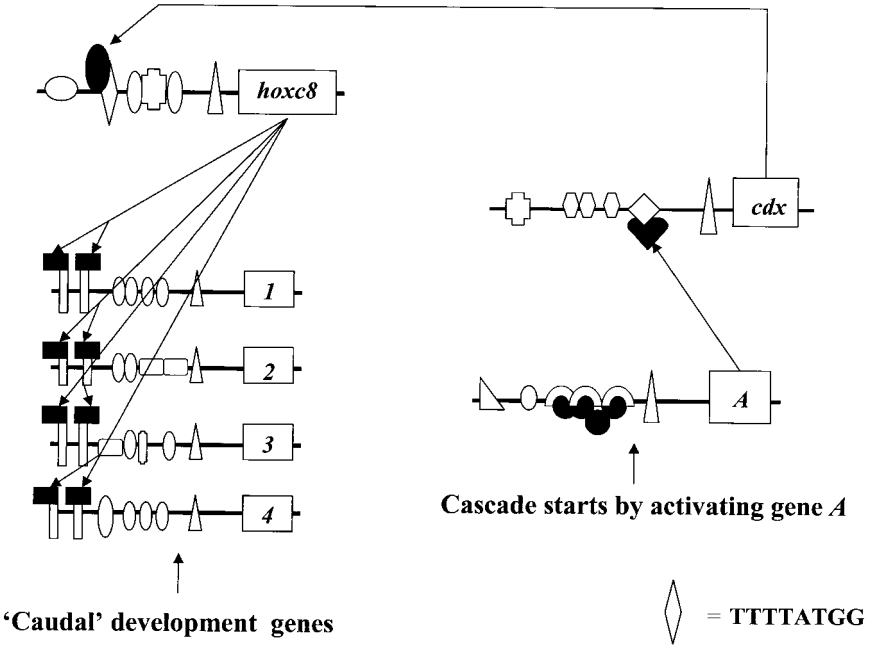

Figure 4. A hierarchy of circuits contributing to activate a set of functional genes.

ready-made proto-organs, such ancestors may be better reconstructed as having a set of primitive regulatory circuits for the basic functions represented in those organs (see Davidson, 2001). Fly, octopus, and human eye development share the use of a homologous *pax6* regulatory gene circuit for photoreceptors, but their actual eyes are so different that for centuries they were used as exemplars of analogy—independent convergent evolution. But the regulatory circuits reveal an underlying element of what may be true homology. We have viewed the world in terms of macroscopic phenotypes, when it may be more useful to think of the regulatory types that make them possible. Who knows what such an *Ur*Cootie may have actually looked like!

These notions have reassuring implications for the evolution of complex organisms. Once evolved, the regulatory toolkit remained available for use. Because enhancer sequences are only a few base pairs long (e.g., **TTTATGG**), they can arise even in random sequences flanking a gene by just a few mutational changes. This would "attract" TFs to that gene, perhaps to activate it in some new context. Since a gene is only expressed when a particular *combination* of TFs binds its nearby DNA, there may be little cost to littering the genome with enhancer sequences. It may be easier to evolve new function by changing expression patterns of existing genes than to require an entirely new gene. A cost of doing business this way, however, is the high fraction of an organism's energy, and the information in its genome that is required. Yet it is the final traits rather than the way they get delivered that have traditionally attracted the attention of evolutionary biologists, are used in comparative and cladistic studies, are found in the fossil record, and appeared to show us how organisms evolved adaptively. Obviously, the cost was worth it.

Where is the Reality? In the Phenotype?

Each function in an organism requires its own subset of regulatory circuits, and since the toolkit is limited and the factors diverse, this means that evolution has produced an exquisitely complex buzz of regulatory activities within, and differences among, cells. These achieve a high degree of specificity (fingers rarely contain eyes or teeth!). But surprisingly, this logically necessary toolkit of regulatory mechanisms is functionally arbitrary in the same way that protein coding is.

As an example, **TTTTATGG** is an enhancer sequence that is bound by a TF called *cdx1*. Experiments in mouse have shown that binding of the *cdx1* TF protein to its enhancer near the *hoxc8* gene induces caudal (tail-end) development in the early embryo, by triggering a cascade of caudal-structure-related development (Fig. 4). But there is nothing about **TTTTATGG** that is chemically related to the structure of vertebrae, or tails—or even mice when it comes to it, because the same *hoxc8* gene is used in many contexts within the same animal and in many different species, including you and me, and even flies. Nor is anything about the *cdx1* protein, that induces *hoxc8,* physically related to vertebrae, tails, or mice. And nothing about *hoxc8* is physically related to vertebrae, or tails, or mice. The only thing about the *cdx1* → **TTTTATGG** → *hoxc8* regulatory cascade that is related to anything caudal is that the cascade occurs in appropriately primed cells in the caudal end of the embryo.

The arbitrariness of these mechanisms can be seen in the fact that it's becoming possible to identify gene expression patterns by screening DNA sequences on a computer to find clusters of known motifs (like **TTTATGG**). That reveals genes regulated by the corresponding TF without requiring any knowledge whatsoever of function (e.g., Michelson, 2002).

A mechanism that is functionally arbitrary is replaceable, and there are many examples of substitution of one such tool for another. Different species may use partly or entirely different regulatory mechanisms to achieve a trait they have shared since their common ancestor. Even genes expressed in the same cell at the same time can use entirely different regulatory pathways, even if those genes are descendants of a common ancestral gene and their usage has been conserved since that time. The joint expression in red blood cells of the two genes required to produce hemoglobin (alpha and beta globin) is an example.

It is not clear how often one regulatory circuit has become replaced by another in this way. But we do know that we can't characterize all the genes expressed in a given context by the same set of enhancer elements. In whatever way it happened, over evolutionary time and the complex history of cellular differentiation during development, the expression (or repression) of the thousands of genes used in a given cell is controlled by many different pathways. But one thing these share is that they are arbitrary relative to the functional characteristics of the cell.

In fact, there may be more interesting aspects of the evolution of gene regulation. The hundreds of interacting genes and enhancer sequences have natural variation just as any structural genes or traits do. This variation can affect binding efficiency, speed or duration of gene action, strength of expression, and so on. Much of the quantitative variation we see in phenotypes in natural populations is probably related to that regulatory variation. However, there are distinctions between the evolution of regulatory mechanisms and "final" phenotypes like limbs and hair.

The use of the same regulatory circuit in many different ways within the same organism (sometimes even in multiple ways during the evolution of a given organ) is a kind of pleiotropy that exposes the circuit to potentially heavy natural selection in many contexts that may be vital to the organism's survival. The more numerous these contexts the more opportunity there is for problems due to mutations in the circuit's enhancers. That multiple jeopardy could be a brutal way to keep an embryo honest, and may also explain the high degree of evolutionary conservation found in regulatory circuits (TFs or binding factors can have similar action when experimentally engineered into distant species, like between flies and mice). But there may be countervailing selection.

If *hoxc8* is used in the development of (say) 20 tissues, the organism could

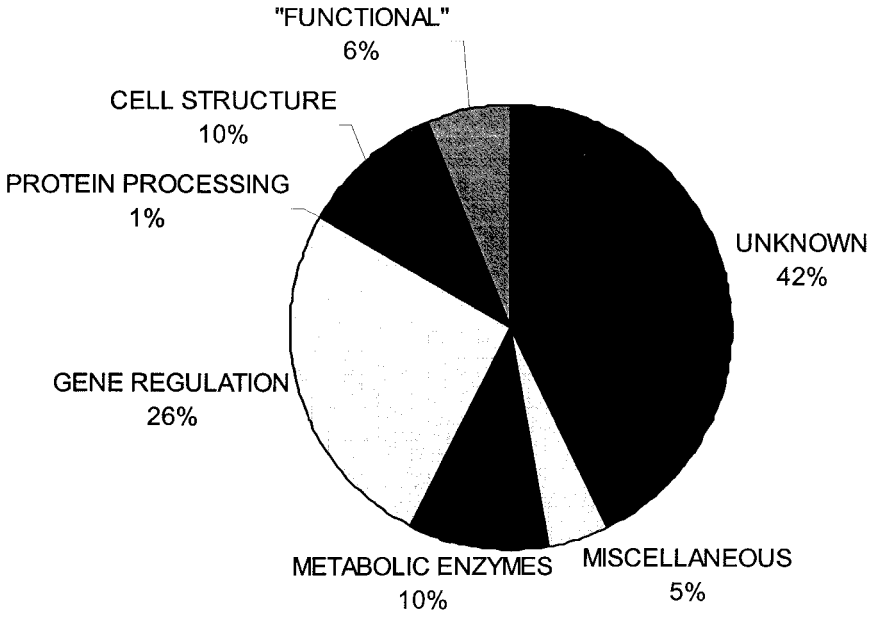

Figure 5. Rough distribution of functional types of human genes.

fail due to mutations in any of the required **TTTTATGG**'s. A form of protection that may evolve could be increases in the number of copies of an enhancer near each regulated gene, or more variation in the enhancers' sequence (that is, the TF evolves to require less precision in the enhancer sequences it binds to). If this is a common evolutionary buffering response, we should find a correlation between the number of contexts in which a TF is used and the number of copies of its enhancers near the genes it regulates and/or variation in the enhancer sequence. A recent study has found just such evidence in a screen of a bacterial genome (Sengupta et al., 2002). Note that here the "trait" being selected is the joint set of variant **TTTTATGG**'s across the genome.

Such redundancy may buffer the organism against mutation, but that may tether strange sets of straits together in the evolutionary struggle. Redundancy or slippage in enhancers may make it harder for these traits to evolve independently, because it may be more difficult for evolution to *remove* or modify a context by mutating the enhancers. Traits may thus coevolve because they are under joint constraints that have nothing to do with the function of traits themselves (flat teeth, pointy ears, long tail, . . .). This could lead to otherwise curious suites of apparent coadaptation that biologists from Cuvier to cladists have gone to great lengths to explain functionally. Ironically, such coevolution may reflect highly specific selection—but for a set of functionally *arbitrary* enhancers. That would be a bizarre twist in the nature of "adaptation" indeed!

SO WHAT IS THE MESSAGE? WHAT IS THE MEDIUM?

A lot has been going on behind the morphological scenes over the long history of evolution. We've been blithely concentrating on morphological phenotypes all this time, as if they were the real business end of life. But at the molecular level, the induction of one activity by another through protein-protein or protein-nucleic acid binding must have long preceded structural features that arose only in the last fraction of the history of life since cells evolved.

If regulation is the primary activity of life, what then is *form?* To rephrase the chicken and egg quip, maybe a person is a regulatory circuit's way of making another regulatory circuit. At least, it is gene regulation that turns a fertilized egg into an organism that can make another egg, and the turning is probably more complex a job than maintaining a body once it's been produced. Most of the information in the genome may be consumed in early stages of embryogenesis. This is ironic, because from the early Lamarckian incarnation of biological evolution, through Darwin and his acolyte Ernst Haeckel and many others, the notion of *terminal addition* has had a lot of importance in biology.

The idea of terminal addition is that by one means or another, natural selection adds modifications onto traits present in adults to render them more suited to their environments. This was the basis of many variations of Haeckel's famous biogenetic law that ontogeny recapitulates phylogeny, because with terminal addition embryos really would reflect adult stages of ancestral species. Key to this was the view that

Figure 6. The message: Fix me! Junk me!

selection doesn't act on embryos. This view is implicitly still widespread in practice, even if every biologist knows better and would say so. Much of the pervasive regulatory activity that I've been discussing occurs and is likely to be selected at embryologic as well as adult stages of life. But keep in mind that what is being selected is the action, not the nature, of functionally arbitrary genes.

Recent analysis of the entire genomes of humans and other species show that perhaps fewer than 10% of all genes are concerned with the structural aspects of the phenotypes we hold so near and dear (Figure 5). King and Wilson (1975) long ago speculated that the apparently great phenotypic differences between chimps and humans, who differ only slightly at the gene level, might be due mainly to slight differences in the regulation of developmental timing. This notion seems ever more likely to be true rather than just speculative.

At the time, McLuhan's phrase that the medium is the message appealed to a lot of people, but to me he seemed just to be playing cute word games with our commercialized culture. It's true that selling a product often uses arbitrary tactics having little if anything to do with the product itself. McLuhan's notion may also be highly relevant to biology. Phenotypes and the regulatory networks that make them are connected, perhaps like ads and cars, but their arbitrary nature and substitutability allows traits and regulation to go their separate ways over evolutionary time. While biologists have been paying most of their attention to the end products, manipulating the organism to get its products to market may be the main engine of evolution.

Speaking of engines, I certainly felt manipulated by advertising (Fig. 6). My '60s experience was no triumph, because it was a real Triumph: no girls, and after a few thousand miles—no engine, either! If you ever owned one, you'll know what I mean.

NOTES

I would welcome comments on this column: kmw4@psu.edu

I thank Anne Buchanan for critically reading and purging this manuscript.

TO READ

Most things discussed here including basic descriptions of the DNA-RNA-protein coding system can be easily explored by web searching.

Carroll SB, Grenier JK, Weatherbee SD. 2001. From DNA to diversity: molecular genetics and the evolution of animal design. Malden, MA: Blackwell Science.

Davidson, Eric H. 2001. Genomic regulatory systems development and evolution. San Diego, Academic Press.

Hall BK. 1999. Evolutionary developmental biology. Dordrecht, Kluwer Academic.

Kay L. 2000. Who wrote the book of life? A history of the genetic code. Palo Alto, Stanford Press.

King M-C, Wilson A. 1975. Evolution at two levels in humans and chimpanzees. Science 188: 107–16.

Michelson A. 2002. Deciphering genetic regulatory codes: A challenge for functional genomics. Proc Nat Acad Sci USA 99:546–548.

Sengupta A, Djordjevic M, Shraiman B. 2002. Specificity and robustness in transcription control networks. Proc Nat Acad Sci USA 99:2072–2077.

Shashikant C, Bieberich C, Belting H-G, Wang J, Borbely M, Ruddle F. 1995. Regulation of *Hoxc-8* during mouse embryonic development: identification and characterization of critical elements in early neural tube expression. Development 121:4339–4347.

Sownward J. 2001. The ins and outs of signaling. Nature 411:759–762.

Weiss K, Fullerton SM. 2000. Phenogenetic drift and the evolution of genotype-phenotype relationships. Theor Pop Biol 57:187–195.

Wolpert L, Beddinton R, Brockes J, Jessell T, Lawrence P, Meyerowitz E. 1998. Principles of development. Oxford, Current Biology.

Books Received

- Peacock, J.L. (2002) *The Anthropological Lens: Harsh Light, Soft Focus*. Xvii + 156pp. New York: Cambridge University Press. ISBN 0-521-80838-3 (cloth) $55.00.
- Marks, K. (2002) *What It Means to Be 98% Chimpanzee: Apes, People, and Their Genes*. 325pp. Berkeley: University of California Press. ISBN 0-520-22615-1 (cloth) $27.50.
- Schwartz, J.H. and Tattersall, I. (2002) *The Human Fossil Record: Volume One—Terminology and Craniodental Morphology of Genus* Homo *(Europe)*. Xi + 388pp. New York: Wiley-Liss. ISBN 0-471-31927-9 (cloth) $125.00.
- Salzano, F.M. and Bortolino, M.C. (2002) *The Evolution and Genetics of Latin American Populations*. xvi + 512pp. New York: Cambridge University Press. ISBN 0-521-65275-8 (cloth) $90.00.
- Calvin, W.H. (2002) *A Brain For All Seasons. Human Evolution and Abrupt Climate Change*. 341pp. Chicago: University of Chicago Press. ISBN 0-226-09201-1 (cloth) $25.00.
- Larsen, C.S. (2000) *Skeletons in Our Closet: Revealing Our Past Through Bioarchaeology*. xvii + 248pp. New Jersey: Princeton University Press. ISBN 0-691-09284-2 (paper) $18.95.
- Clark, J.D. (2001) *Kalambo Falls Prehistoric Site. Volume III*. xix + 701pp. New York: Cambridge University Press. ISBN 0-521-20071-7 (cloth) $375.00.
- Leonard, W.R. and Crawford, M.H. (eds.) (2002) *Human Biology of Pastoral Populations*. xi + 314pp. New York: Cambridge University Press. ISBN 0-521-78016-0 (cloth) $80.00.
- Beckerman, S. and Valentine, P. (eds.) (2002) *Cultures of Multiple Fathers: The Theory and Practice of Partible Paternity in Lowland South America*. vii + 291pp. Gainesville: University of Florida Press. ISBN 0-8130-2456-0 (cloth) $59.95.