Annu. Rev. Cell Dev. Biol. 2002. 18:53–80
doi: 10.1146/annurev.cellbio.18.020402.140619

First published online as a Review in Advance on June 6, 2002

GENE CO-OPTION IN PHYSIOLOGICAL AND MORPHOLOGICAL EVOLUTION

John R. True
Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York 11794-5245, e-mail: jrtrue@life.bio.sunysb.edu

Sean B. Carroll
Laboratory of Molecular Biology and Howard Hughes Medical Institute, University of Wisconsin, Madison, 1525 Linden Drive, Madison, Wisconsin 53706, e-mail: sbcarrol@facstaff.wisc.edu

Key Words morphogenetic novelty, modularity, regulatory evolution, adaptation, gene duplication, evolvability

■ **Abstract** Co-option occurs when natural selection finds new uses for existing traits, including genes, organs, and other body structures. Genes can be co-opted to generate developmental and physiological novelties by changing their patterns of regulation, by changing the functions of the proteins they encode, or both. This often involves gene duplication followed by specialization of the resulting paralogous genes into particular functions. A major role for gene co-option in the evolution of development has long been assumed, and many recent comparative developmental and genomic studies have lent support to this idea. Although there is relatively less known about the molecular basis of co-option events involving developmental pathways, much can be drawn from well-studied examples of the co-option of structural proteins. Here, we summarize several case studies of both structural gene and developmental genetic circuit co-option and discuss how co-option may underlie major episodes of adaptive change in multicellular organisms. We also examine the phenomenon of intraspecific variability in gene expression patterns, which we propose to be one form of material for the co-option process. We integrate this information with recent models of gene family evolution to provide a framework for understanding the origin of co-optive evolution and the mechanisms by which natural selection promotes evolutionary novelty by inventing new uses for the genetic toolkit.

CONTENTS

INTRODUCTION

The story of evolution is about the appearance and disappearance of species and traits of various kinds: organs, structures, pattern elements, cell types, physiological processes, and genes. These traits can either evolve de novo or may be derived from ancestral characters. Although de novo invention certainly fueled early evolution, the broad distribution of conserved proteins and motifs across the tree of life indicates that current diversity among phyla of modern organisms has extensively involved new combinations and modifications of pre-existing molecular characters. The acquisition of new roles by ancestral characters or new characters from old ones is known as co-option. Changes at the level of genes, either in regulation or function, underlie co-option of all the types of traits listed above. As fundamental as this process has been in evolution, little is known about the mechanisms by which co-option of gene function takes place and whether particular modes of co-option are responsible for important episodes of change in the evolutionary history of complex organisms.

Relatively complex organisms such as eukaryotes have more genes than the simpler prokaryotes (Ball & Cherry 2001, Patthy 1999, Vellai & Vida 1999), suggesting that over the long course of evolution, the invention of genes with new functions underlies increases in developmental and metabolic complexity (Carroll 2001). However, comparisons among the genomes of multicellular model organisms, made possible by recent advances in whole genome sequencing, indicate that apparent increases in complexity, in terms of morphogenetic intricacy, cell-type diversity, and behavioral and physiological repertoires, from worms and flies to mice and humans have not involved a corresponding amount of gene invention. The estimated number of human genes is about 30,000 (Internationl Human Genome Sequencing Consortium 2001), which is not impressively larger than estimates for *Drosophila* (13,000), *Caenorhabditis* (20,000), and *Arabidopsis* (30,000) (reviewed by Claverie 2001, Bevan et al. 2001). Therefore, although vertebrate

genomes appear to contain some genes and protein motifs that invertebrate, plant, and fungal genomes do not possess (Brown 1999), the apparently higher complexity of vertebrates must be accounted for by inventions other than adding completely novel gene types (Duboule & Wilkins 1998). These include increases in gene family size by duplication, alternatively spliced forms of the same gene (Herbert & Rich 1999), and increased complexity and specialization of transcriptional regulation (Fickett & Wasserman 2000, Tautz 2000).

Co-option events involving evolutionary changes in gene expression, as well as the evolution of novel gene functions, may have been the central events enabling evolutionary change in developmental and metabolic systems. Although the importance of these events has been a topic of much recent discussion (von Dassow & Munroe 1999, Gerhardt & Kirchner 1997, Wagner & Altenberg 1996), we still have a very limited understanding of how co-option takes place at the mechanistic level, and perhaps even more importantly, what evolutionary forces underlie the process of genetic co-option.

Modes of Gene Co-Option

Genes can acquire new functions by changes in amino acid coding sequences (including alternative splicing), changes in spatiotemporal patterns of gene regulation, or both. The necessity to maintain the gene's original function is expected to constitute a strong constraint on that gene's evolution in most cases. There are three ways by which genes may evolve new functions, even in the presence of selection to maintain their current function (Figure 1): (*a*) Changes may occur in parts of the amino acid sequence not required for the current function; i.e., new or alternative protein domains may evolve (Figure 1*A*); (*b*) changes may occur in gene regulation, expressing a protein in novel tissues and/or developmental stages (Figure 1*B*); and (*c*) gene duplication events may occur, followed by divergence in the amino acid sequences and/or regulatory DNA sequences of the descendant paralogs (Figure 1*C*,*D*). This third mechanism was thought to be required for the evolution of most novel gene functions (Li 1980). However, there is now much evidence that the evolution of novel gene functions via co-option can take place with or without gene duplication and that the order in which duplication and adoption of novel functions takes place can vary in different evolutionary trajectories. A fourth mechanism involves cobbling novel genes together from genomic fragments. This phenomenon, of which there is an increasing number of known examples, can involve either exon shuffling (Long 2001; Gilbert 1978, 1985) or, in some cases, the fortuitous switching of DNA fragments from coding to non-coding and vice versa (e.g., Nurminsky et al. 1998, Long & Langley 1993). In this review we concentrate on co-option involving extant genes because this appears to be a major mechanism through which developmental diversity in multicellular organisms has arisen.

Most examples of co-option appear to fall into two broad categories. First, proteins have been deployed at high levels in particular tissues, strikingly, for purposes completely unrelated to their ancestral function. In some cases, the ancestral

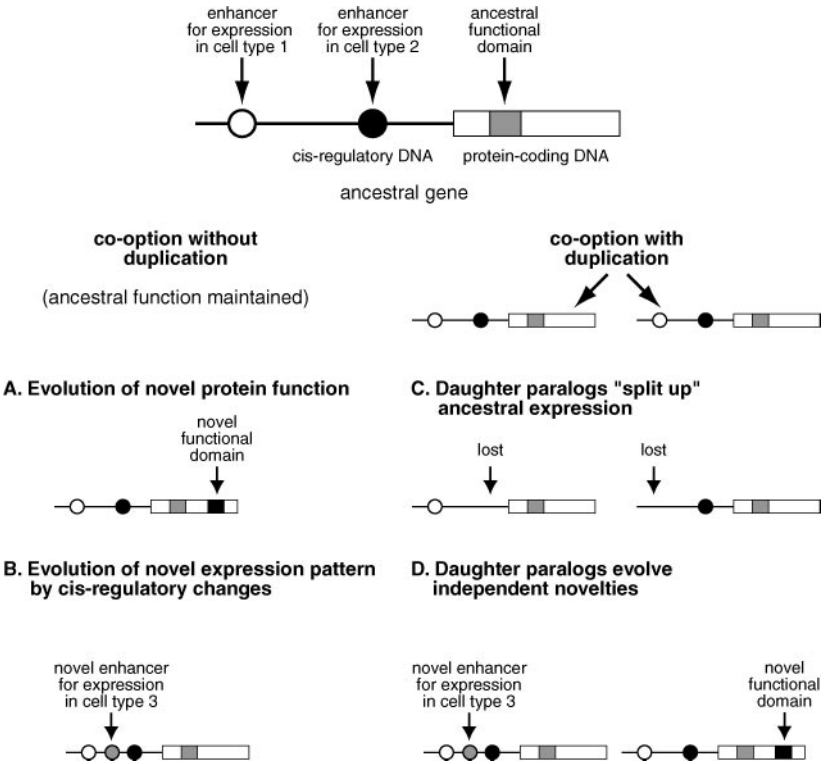


Figure 1 Modes of gene co-option. An ancestral gene can evolve to adopt a new function either by non-duplicative (*A*,*B*) or duplicative (*C*,*D*) processes. When gene duplication is not involved, genes can be co-opted for new functions by changes in the protein-coding sequence (*A*) and/or expressed in novel spatiotemporal domains by changes in *cis*-regulatory control elements or enhancers (*B*). Alternatively, gene duplication may first take place, followed by divergence in *cis*-regulatory control elements between the daughter duplicates (*C*), and/or the appearance of novel *cis*-regulatory elements or protein functional domains (*C*,*D*).

protein function has been maintained (gene sharing) (Piatigorsky et al. 1988, Piatigorsky & Wistow 1989), whereas in others, this function has been lost and only the novel function remains. In the second category, ancestral gene functions such as transcriptional regulation and components of cellular signaling pathways have been conserved, but the genes have been redeployed in different tissues and developmental stages during evolution. It is likely that major aspects of diversification in biological forms and body plans have involved such changes in the developmental genetic toolkit of transcription factors and associated signaling proteins (Carroll et al. 2001). Although there is more direct evidence of the former mode of co-option, evidence of developmental pathway co-option comes from comparative

embryology and developmental genetic analyses, which have burgeoned recently. We first discuss the major documented cases of structural gene co-option and proposed cases of developmental pathway co-option, focusing where possible on the functional and evolutionary circumstances that have been proposed for their origins. Then we briefly examine what is known about the underlying variation that may fuel co-optive evolution. Finally, we discuss the current conceptual position of co-option in evolutionary biology and point to some future research directions.

CO-OPTION OF STRUCTURAL GENES

Animal Lens Crystallins

By far, the classic and best-studied cases of co-option in animal evolution are the lens crystallins of animal eyes. Crystallins are soluble proteins constituting 30–40% of the mass of the lens and have been characterized in many vertebrates and some invertebrates. These proteins occur in highly packed, transparent arrays in the lens, and function to refract light so that it may form a focused image on the retina. Reviews of lens development can be found elsewhere (Graw 1997, Cvekl & Piatigorsky 1996, de Jong & Hendriks 1986). In brief, the vertebrate crystallins are expressed in the highly elongated lens fiber cells, which proliferate from a peripheral generative zone and differentiate forming concentric rings (somewhat like tree growth rings) around the lens embryonic nucleus as vertebrates develop and age. The differentiated cellular environment in which lens crystallins function is devoid of mitochondria and nuclei and has substantially less water than other cell types (65–70% versus 95%) (Graw 1997). Particular properties of crystallins, namely their ability to refract light, to undergo molecular packing, and to persist stably at very high concentrations for very long periods of time, are advantageous for lens function. In many cases, the functions of these co-opted genes and related molecules, both in and outside of the lens, provide clues about the selective forces and types of selectable variation involved in eye evolution.

ANCIENT CO-OPTION OF VERTEBRATE CRYSTALLINS Two classes of crystallins, the α and the $\beta\gamma$ families, are found in all vertebrate lenses and were probably present in the ancestral vertebrate eye (de Jong & Hendriks 1986). The common theme uniting the two families of lens-specific crystallins found throughout vertebrate eyes is that proteins involved with cellular stress responses have been co-opted to serve refractive functions in the lens. In this tissue, proteins must remain stable for long periods (i.e., the animal's lifetime) in an avascular, non-innervated environment that in many animals is exposed to high light levels. It has been speculated that functions involved with cellular stress responses and protein stability may have been pivotal in their original evolutionary recruitment in the lens (Wistow & Piatigorsky 1988, Wistow 1993, Tomarev & Piatigorsky 1996).

α crystallins possess substantial homology to the small heat shock proteins (Ingolia & Craig 1982), which are stress-induced molecular chaperones involved

in protein folding and protection against protein aggregation. This indicates that there was an ancient event in which of one of these proteins was co-opted into a refractive role in the lens in the common ancestor of vertebrates. α crystallins occur as a family of two genes, αA and αB, each encoding a 20-kDa polypeptide, 60% identical to each other, and probably representing an ancient gene duplication. Interestingly, only αA has retained lens-specific expression where it is thought to function as both a crystallin and a heat shock–induced chaperone. αB appears to be expressed ubiquitously and is thought to function both as a lens crystallin in complexes with αA and as a stress-inducible protein during heat shock. The protein-stabilizing chaperone function of α crystallins may be particularly vital in the lens because degradation and extrusion of defective proteins cannot occur in this tissue and because the light-exposed environment of the lens causes large amounts of oxidative damage to proteins (reviewed by Graw 1997).

β and γ crystallins are members of a large family of proteins (the $\beta\gamma$ superfamily) all having four copies of a structural motif known as the Greek-key. This family has undergone much more diversification than α crystallins. Six β crystallins, consisting of three acidic (βA) and three basic (βB) types, have been found in all mammals, birds, and frogs (reviewed by Graw et al. 1997). γ crystallins occur in a cluster of six very similar genes (A–F) in humans (den Dunnen et al. 1985), mice (Skow et al. 1988), and rats (den Dunnen et al. 1989) and are prominent in fish and mammal lenses, especially during embryonic and juvenile stages (Harding & Dilley 1976). The $\beta\gamma$ crystallin superfamily shows substantial sequence similarity to a superfamily of microbial stress proteins, including protein S from *Myxococcus xanthus* and spherulin 3a from *Physarum polycephalum* (Wistow 1990). These microbial proteins are expressed during the formation of dehydrated spores and are highly stable during extreme environmental conditions. Thus similar to α crystallins, $\beta\gamma$ crystallins also appear to represent ancient stress-related proteins that have been co-opted for refractory functions.

RECENT CO-OPTION OF TAXON-SPECIFIC CRYSTALLINS Analyses of crystallins found in the lenses of particular vertebrate lineages have revealed that co-option has not only marked the ancient history of eye evolution, but has also been a frequent and ongoing process within lineages. Many taxon-specific lens crystallins have been found (Figure 2). Amazingly, most are very similar or identical to known enzymes and are expressed at concentrations of up to 10–20% of soluble lens protein. This far exceeds amounts thought to be needed for enzyme activity. Many of these crystallins function as enzymes in tissues outside the lens, and as such these proteins are truly bifunctional (Wistow & Piatigorsky 1987, Wistow 1993). Detailed information on each can be found in the references in the legend of Figure 2. Here we briefly review a few cases in order to illustrate the remarkable regularity with which enzymes have been co-opted as crystallins.

The best-studied taxon-specific crystallins are those found in birds, reptiles, and fish. δ crystallin is the major lens crystallin in birds and reptiles, constituting up to 70% of the soluble protein in embryonic chick lens (Wistow & Piatigorsky 1988).

Two tandemly duplicated δ crystallin genes in chickens encode products with high degrees of homology to the enzyme arginosuccinate lyase (ASL) (Piatigorsky et al. 1988), which is involved in the detoxification of ammonia and in arginine biosynthesis. These two genes have undergone both regulatory and functional diversification. In the chicken, $\delta 1$ crystallin is expressed preferentially in the lens and has lost enzyme activity. Several specific amino acid substitutions are found in $\delta 1$ but not in $\delta 2$ or other ASLs (Piatigorsky et al. 1988), which suggests evolution of $\delta 1$, since being recruited as a crystallin, has involved selection of amino acid changes favorable for that function. $\delta 2$, on the other hand, retains ASL activity, and its expression is much lower than $\delta 1$ in the lens but much higher than $\delta 1$ in other tissues (Parker et al. 1988, Wistow & Piatigorsky 1990). Thus it appears that the original δ crystallin was encoded by a single gene in the common archosaur ancestor of birds and reptiles and had both enzyme and crystallin properties. Subsequently, in at least the lineage leading to chickens, a duplication took place, followed by loss of enzyme activity and specialization of $\delta 1$ as a crystallin.

Two taxon-specific crystallins, ε and τ, are active enzymes involved in glycolysis. This glycolytic function may be a clue to their mechanism of co-option as crystallins because differentiating lens fiber cells lose all of their organelles, including the mitochondria. Hence, glycolysis in the cytoplasm is the sole source of energy for these cells (see Wistow & Piatigorsky 1987). ε crystallin is found in birds and some reptiles and is identical to and possesses the activity of lactate dehydrogenase B (LDHB) (Wistow et al. 1987). In species not using LDHB as a lens crystallin, the enzyme is still found at low levels, suggesting that it is involved in normal lens metabolism and in specific lineages has been co-opted as a crystallin. Birds appear to possess only one ε crystallin/LDHB4 gene encoding both the crystallin and enzyme functions. Interestingly, although ε crystallin must have been recruited in the common ancestor of crocodilians and birds, it is absent in many bird species in a pattern that does not correspond in a simple way to avian phylogeny, which suggests many independent losses. It has been speculated that the NAD/NADH-binding property of LDHB may have been important in its lens role, either for energy sequestration or for glare reduction, because the absorption maximum for NADH is at a wavelength similar to the range in which bird retinas are most sensitive (Wistow et al. 1987). τ crystallin, found in many birds, fish, and reptiles, encodes the enzyme α enolase which, similar to LDHB, is involved in glycolysis. In all species studied so far, τ crystallin/α enolase is encoded by a single gene.

One mammalian-specific crystallin, ζ crystallin, has been found in guinea pigs, camels, and bovines (reviewed by Graw 1997). ζ crystallin is a member of the medium chain dehydrogenase/reductase family (MDR, which includes alcohol dehydrogenase) and has been shown to have NADPH-dependent quinone oxidoreductase activity in guinea pig but not bovines. As in other taxon-specific crystallins, a single gene appears to have acquired high lens-specific expression. Importantly, it is thought that independent co-option events may have occurred leading to the rodent, bovine, and camel ζ crystallins. Recently, a quinone oxidoreductase/ζ

crystallin was also found in the lens of the Japanese tree frog *Hyla japonica* (Fujii et al. 2001), suggesting that this enzyme may have been co-opted at least three times to serve as a lens crystallin.

Although less-well-studied, crystallins found in invertebrate eye lenses have been similarly co-opted from a variety of genes (reviewed by Tomarev & Piatigorsky 1996). These include members of the glutathione S transferase and aldehyde dehydrogenase gene families in cephalopods and a molecule similar to bacterial dinitrogenase/reductase in jellyfish (see Figure 2). Because there is much more phyletic diversity in invertebrates than vertebrates, it will be of great interest to characterize the lens crystallins from other invertebrate phyla. In particular, it will be interesting to determine whether any invertebrates or primitive chordates express α or $\beta\gamma$ crystallins in order to estimate the minimum age of co-option of these classes of crystallins found in all vertebrate eyes.

EVOLUTION OF LENS-SPECIFIC GENE REGULATION An event common to all the co-option events underlying the evolution of crystallins is the acquisition of high, tissue-specific expression in the lens. It has been speculated that one possible evolutionary trajectory for the evolution of this high-level expression of crystallins, which are ancestrally involved in stress responses, would be the ability of stress signals to induce high expression in specific organs (Cvekl & Piatigorsky 1996). Subsequently, this inducibly high level of expression would then evolve into developmentally high expression in specific organs by regulatory changes. This could conceivably occur by alterations in *cis*-regulatory DNA sequences. For example, a stress-inducible promoter or enhancer could evolve lens-specific expression by acquiring a tissue-specific enhancer either by nucleotide substitutions or by a transposition or genome rearrangement that brought the transcription unit close to an extant enhancer.

The lens expression of crystallins is thought to be mediated by tissue-specific transcription factors used in organogenesis. Studies of eye development in both vertebrates and invertebrates have revealed and characterized several such factors conserved throughout animals, most notably the paired/homeobox transcription factor Pax6, which controls the expression of a hierarchy of genes involved in eye development (reviewed by Desplan 1997). Pax6 binding sites have indeed been found in the regulatory regions of all vertebrate crystallin genes examined (Cvekl & Piatigorski 1996). Various experimental methods have demonstrated that Pax6 directly binds enhancer sequences and activates expression of the chicken αA and δ1, mouse αA and αB, and guinea pig ε (reviewed by Cvekl & Piatigorsky 1996). Interestingly, for the chicken βB1 crystallin, Pax6 has been shown to repress expression, which is consistent with non-overlapping expression patterns of chicken Pax6 and β crystallins. Chicken β1 activation may be controlled by another homeodomain transcription factor, Prox1, which is likely involved in regulating crystallin expression during early development of the lens in chicken and mouse (Tomarev et al. 1996). Prox1 is a homolog of *Drosophila* Prospero, which is required for proper cell division in the embryonic nervous system, including

neurons in the eye. Prox1 regulation, like Pax6 regulation, appears to have been involved in the recruitment of multiple lens crystallin genes. Lengler et al. (2001) have shown that Prox1 upregulates the γE and F crystallins in the developing mouse lens. Thus it appears that the acquisition of Pax6 and/or Prox1 binding sites may represent the central event in the evolution of lens-specific crystallin expression. This idea, although attractive, has not been tested. One prediction is that closely related genes that have not been co-opted for crystallin function would lack binding sites for eye-specific transcription factors.

Comparative studies also suggest a role for promoter evolution in crystallin co-option. Whereas some crystallins use the same promoter for lens and non-lens expression (e.g., duck δ1, δ2, and ε crystallins) (Hodin & Wistow 1993), many crystallin genes appear to have evolved TATA promoters for lens-specific expression. Many δ crystallins and the duck τ crystallin have TATA promoters (reviewed by Piatigorsky & Zelenka 1992), whereas the human orthologs of these genes, which have not been co-opted for lens expression, have the GC-rich promoters characteristic of more ubiquitously expressed or housekeeping genes (Kim & Wistow 1993, Abramson et al. 1991). Of particular interest is the guinea pig ζ crystallin gene, which also functions as a quinone oxidoreductase in the lens and other tissues. This gene has both a GC-rich promoter, which is used in non-lens expression, and a TATA promoter, which is used in the lens (Gonzalez et al. 1994). While more data are awaited, it is apparent that no single common mechanism was utilized to evolve lens-specific expression.

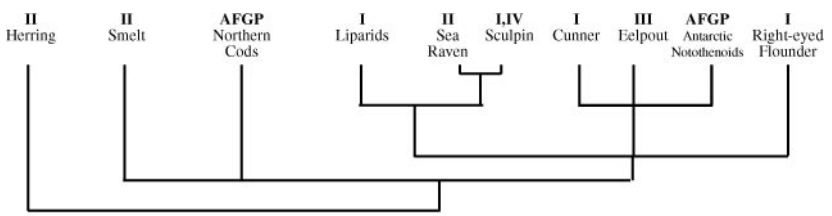
NATURAL SELECTION AND CRYSTALLIN GENE CO-OPTION Light is one of the most important environmental variables driving the ecology of organisms. Does the rich co-option history of the lens crystallins contain clues about the selective forces involved in major events in the evolution of animal lineages? Wistow (1993) suggests that crystallin gene family diversification has been adaptively important for major radiations of aquatic vertebrates onto land. The eye of vertebrates evolved in animals living in an optically dense medium, water, in which highly dehydrated lenses with high refractive indices were needed, similar to those of modern fish. γ crystallins are predominant in the high-refraction regions of both fish and mammalian lenses, as well as in the hard, dehydrated regions of rodent lenses (Wistow & Piatigorsky 1988). The novel land environments that ancestral terrestrial vertebrates encountered may have favored softer lenses with lower refractive indices, important for the ability to focus at a distance through air. This type of change is seen in birds, where virtually all γ crystallins have been replaced with the δ, ε, and τ enzyme crystallins, which may have contributed to changes in refractive index (Wistow 1993). Some mammalian lineages have similarly replaced γ crystallins, possibly because of similar ecological forces, but with a variety of different taxon-specific crystallins (see Figure 2). Humans also show reduced lens expression levels of four of their six γ crystallins (Brakenhoff et al. 1990). Although experimental data are desirable in order to test these selective hypotheses, it certainly appears that a variety of independent solutions have been found by

many terrestrial vertebrates to fulfill the specific demands this lifestyle imposes upon vision.

Antifreeze Proteins

Organisms inhabiting sub-freezing aquatic or terrestrial environments express various small antifreeze proteins (AFPs) that function to adsorb and halt ice crystal growth. These proteins have a great diversity of structures and thus are thought to have originated many times in evolution. Echoing the crystallin saga, natural selection has favored the co-option of several different proteins toward the ice-binding function. Two broad types of AFPs are known, those with repetitive amino acid motifs and those without repetitive primary structure. Both types are thought to bind to various faces of small ice crystals to inhibit further addition of water molecules to the ice lattice, although the specific mechanisms may be quite diverse (reviewed by Fletcher et al. 2001, Knight 2000).

TELEOST AFPs The best studied AFPs are in teleost fish, various lineages of which possess one of five independently evolved and structurally diverse AFPs (Figure 3), found at high levels in the blood, skin, and gills (reviewed by Fletcher et al. 2001, Cheng 1998). One class, the antifreeze glycoproteins (AFGPs) of the unrelated Antarctic notothenoids and North Atlantic cod, shows evidence for independent co-option from unrelated genes (Chen et al. 1997a). Both AFGPs are present as gene families, and they share a repetitive three-amino acid repeat motif, Ala-Ala-Thr, of varying repeat number. The disaccharide galactosyl-*N*-galactosamine is attached to each Thr, and the repeats are punctuated by a spacer sequence that is the substrate of a protease, which processes pre-AFGP polyproteins to their mature



AFP type	Ancestral protein function
I	unknown
II	C-type lectin
III	unknown
IV	apolipoprotein
AFGP	Trypsinogen-type serine protease

Figure 3 Evolution of teleost antifreeze proteins (AFPs). Type I and II AFPs and APGPs are thought to have been derived multiple independent times, whereas types III and IV AFPs appear to be uniquely derived (Fletcher et al. 2001) (Figure adapted from Fletcher et al. 2001).

form. Although they share the amino acid repeat motif and have remarkably similar size distributions of mature peptides, the AFGPs of notothenoids and northern cod have been shown to be unrelated by virtue of their complete lack of similarity in the 5′ and 3′ ends of the genes, spacer sequences, and in intron-exon structure (Chen et al. 1997a). The notothenoid AFGP is closely related to the trypsinogen class of pancreatic enzymes, and its Ala-Ala-Thr repeats originated by duplication and amplification of a conserved, partially noncoding, nine-nucleotide sequence at the beginning of exon 2. Most AFGPs have since lost the remainder of exon 2, as well as exons and introns 3–5 of trypsinogen, but have retained exon 6, as well as exon 1, which encodes the signal peptide for extracellular transport, vital for both trypsinogen and AFGP function (Chen et al. 1997b). Surprisingly, a chimeric AFGP-trypsinogen protein has been isolated from one species, the giant Antarctic toothfish, *Dissostichus mawsoni*, which still contains exons 2–6 of the trypsinogens. This chimeric gene, which may be present in many species of notothenoids, is thought to represent an evolutionary intermediate between trypsinogens and AFGPs (Cheng & Chen 1999).

The AFGPs of northern cod do not bear trypsinogen-like motifs, and the genes from which they may have been derived are thus far unknown (Chen et al. 1997b). Furthermore, cod and notothenoids are highly unrelated groups, so it is unlikely that their common ancestor possessed an AFGP of this sort. Molecular phylogenetic evidence strongly suggests that in both cases, AFGPs enabled these fishes to colonize cold water environments. The Antarctic notothenoids radiated 7–15 million years ago (mya), roughly corresponding to the mid-Miocene cooling of the Antarctic Ocean. Importantly, the notothenoids are by far the most species-rich of any Antarctic teleosts (Eastman 1993), which suggests that the recent invention of AGFPs may have been the key adaptation enabling the invasion of this lineage into new ecological niches (Chen et al. 1997b). The evolution of AFGP in northern cod, as well as four other classes of structurally diverse AFPs found in other northern hemisphere teleost lineages, is believed to have taken place even more recently, some 1–2 mya, corresponding to the cooling of the Arctic Oceans (reviewed by Cheng 1998). The extremely different structures of AFPs in northern hemisphere teleost lineages also suggest that the strong selective pressures for survival in ice-laden waters were able to co-opt a variety of proteins to serve as antifreezes. Two of these AFPs show relatedness to C-type lectins and apolipoproteins, respectively, whereas the origins of cod AFGPs and two other AFP classes remain unknown (see Figure 3).

Further evidence for rapid evolution of AFPs comes from analyses of closely related species. Two different species of North Atlantic sculpins have been shown to express completely different classes of AFPs (Deng et al. 1997, Hew et al. 1980). AFP copy number and tissue-specific regulation also appear to evolve rapidly. Closely related species of winter flounder have a widely different copy number of particular AFPs in their genomes. The shallow coastal species *Pleuronectes americanus* has 30–40 copies of AFPs, whereas its deep water congener *P. ferrugenia*, which encounters little ice in its habitat, has only about one third this copy number (Scott et al. 1988).

INSECT AND PLANT AFPs Insect and plants living in freezing terrestrial environments have also invented antifreeze proteins that are unrelated to any teleost AFP. Insect AFPs protect against ice crystal growth at substantially lower temperatures than teleost AFPs (Graether et al. 2000), consistent with the more extreme freezes experienced in terrestrial conditions. AFPs have recently been characterized from a beetle and a moth, and both possess a repetitive β-helix structure, a highly ordered motif that was thought to be very rare in nature (Knight 2000). The first plant AFP was discovered recently in the ryegrass *Lolium perenne* and possesses a repetitive seven–amino acid repeat motif (Sidebottom et al. 2000). Taken together, the diversity of AFPs strongly suggests that AFP protein co-option has been widespread and rapid and, similar to crystallins, provides further evidence that under particular and strong selective conditions, co-option events utilizing a diverse group of gene precursors have frequently occurred during evolution.

Cytoskeletal Gene Diversification and Major Evolutionary Transitions

Diversification of gene families is a hallmark of genome evolution in higher organisms, and gene duplication often facilitates evolutionary co-option of genes for new functions. The crystallin and AFP data give some indications of the selective forces underlying particular episodes of co-option of functionally specific genes. Can information on distribution and expression of other types of gene families provide information on evolutionary mechanisms underlying other, phyletically deeper, radiations of organisms?

The cytoskeleton of all eukaryotic cells functions dynamically to mediate a huge repertoire of movements, shape changes, interactions, and signaling processes. In multicellular organisms, many large gene families of cytoskeletal proteins have arisen, often resulting in the expression of specific family members in particular cell types. An intriguing possibility is that specialization of cytoskeletal components may have been a crucial factor in the evolutionary diversification of multicellular organisms.

ACTINS AND THE RADIATION OF LAND PLANTS One cytoskeletal gene family whose taxonomic distribution shows a striking correspondence with a major macroevolutionary radiation is plant actins. In *Arabidopsis*, there are eight functional actin genes, all of which show strong tissue-specific expression, three in vegetative structures such as flower, leaf, stem, and root, and the other five in germline structures, including pollen, pollen tube, and early embryonic tissues (reviewed by Meagher et al. 1999b). Importantly, at the amino acid level, actins of each class are more closely related to each other (up to 12% divergence) than they are to actins in the other tissue classes (about 28% divergent) (McDowell et al. 1996), suggesting ancient specialization followed by more recent diversifications of vegetative

and germline-specific gene families. Analysis of the plant actin gene tree and the taxonomic distribution of tissue-specific expression indicates that the evolutionary split between vegetative- and germline-specific actins occurred 350–500 mya. This split corresponds roughly to the divergence of leaves from reproductive structures in early vascular plants. Additional evidence of this correspondence between the gene phylogeny and this major event in plant macroevolution is suggested by the occurrence of an asparagine residue at position 79 exclusively in two pollen-specific acins, Act4 and Act12. These pollen-specific actins are found in diverse angiosperms and in the most recently derived gymnosperms (Kandasamy et al. 1999). Meagher et al. (1999b) have proposed that this class of pollen-specific actins arose about 220 mya in the late gymnosperm ancestor of derived gymnosperms and all angiosperms. Intriguingly, another family of plant cytoskeletal proteins, the profilins, also shows some evidence for an ancient split into ubiquitously expressed and pollen-specific forms (Huang et al. 1996). Actins and other cytoskeletal proteins have the ability to exhibit a quality that Meagher et al. (1999a) refer to as "isovariant dynamics," in which the milieu of polymerization and other protein-protein interactions in the cytoskeleton and the structures it contacts may be highly specific to the particular assemblages and concentrations of monomers in each cell type. This ability to mediate cell-type-specific dynamics and behavior may have been important in the evolution of specific functions and responses to developmental and physiological signals.

CYTOSKELETAL PROTEIN DIVERSIFICATION IN METAZOANS There is also much diversity and cell-type specificity in animal actins (Kabsch & Vandekerckhove 1992), as well as in actin binding and crosslinking proteins (Dubreuil 1991), and molecular motor proteins (Vale & Milligan 2000). This suggests that the cytoskeleton may represent a site where many co-option events important for cell type diversity have occurred during evolution.

Links between cytoskeletal protein co-option and muscle evolution have been uncovered. During the radiation of animal phyla, muscle-specific actins appear to have been derived independently in molluscs, arthropods, and chordates (Mounier et al. 1992, Carlini et al. 2000). Also, Groger et al. (1999) have isolated a novel tropomyosin, Tpm2, specifically expressed in the striated muscle of the hydrozoan cnidarian *Podocoryne carnea*. Tpm2 is structurally very divergent from striated muscle-specific tropomyosins in higher animals and is also divergent from Tpm1, a ubiquitously expressed *Podocoryne* tropomyosin. This divergence suggests either that the common ancestor of higher animals and cnidarians possessed a striated muscle-specific tropomyosin or that these tropomyosins were recruited independently in two different primitive metazoan lineages. Importantly, cnidarians are believed to be diploblasts, lacking mesoderm, whereas in higher animals, muscle cells derive from the mesoderm. It will be interesting to determine whether other cytoskeletal components show similar specificity in diploblast muscle versus triploblast mesoderm.

CO-OPTION OF DEVELOPMENTAL REGULATORY GENES AND CIRCUITS

The examples we have discussed thus far have involved evolutionary co-option of structural or enzyme-encoding genes that lie at the termini of regulatory circuits. It is possible that multiple genes connected by regulatory linkages can be co-opted as a unit to serve a novel function. The number of genetic changes involved in this type of evolution is unknown, but it is tempting to speculate that novel expression of just one or a few regulatory genes may result in the redeployment of entire sets of genes in novel contexts. The evolution of some morphological novelties is correlated with the co-option of regulatory circuits or sets of regulatory genes.

Butterfly Eyespots

One of our favorite examples of morphological novelties and gene co-option is presented by the origin and diversification of butterfly eyespots. These striking pattern elements, composed of concentric rings of pigmented scales, evolved in the Lepidoptera, presumably from simpler spot patterns (Nijhout 1991). These elements are believed to serve in predator avoidance to distract strikes from birds or lizards away from the more vulnerable main body of the animal (Brakefield & Reitsma 1991).

Pioneering studies by Nijhout (1980a,b) demonstrated that eyespots develop from a region with a classic organizer function that acts as a signaling center. Transplantation of this organizer, termed the focus, induces eyespot formation at ectopic sites (Nijhout 1980b, French & Brakefield 1995). Surveys of the expression of a variety of transcription factors and components of signaling pathways that regulate pattern formation in *Drosophila* appendages have revealed several proteins deployed in novel and striking patterns in the establishment of the focus and in response to focal signaling that appear to be the product of multiple co-option events. (Figure 4*A*). For example, the Distal-less (Dll) protein is stably expressed in all of the cells that form the focus (Brakefield et al. 1996). The Dll protein has been well-studied in regard to its requirement for appendage formation in insects (French & Daniels 1994), and this role appears to be conserved throughout arthropods and perhaps other animals (Panganiban et al. 1995, 1997). However, the deployment of Dll in spots in developing butterfly wings is novel. Differences in eyespot number, size, and position all correlate with Dll expression, and it was recently found that genetic variation at the *Dll* locus in *Bicyclus anynana* is associated with differences in eyespot size between selected lines, suggesting that Dll plays a direct role in regulating eyespot formation (Beldade et al. 2002).

How is the novel pattern of Dll expression induced in the focus? It appears that an entire regulatory circuit has been co-opted in butterflies that induces gene expression in the focus (Figure 4*A*, *left*). Four gene products known to share regulatory connections in *Drosophila* are expressed in and around the focus. These include the Hedgehog (Hh) signaling ligand, one of its co-receptors Patched (Ptc), a

downstream transcription factor Cubitus interruptus (Ci), and a target gene *Engrailed* (*En*). Not only are all four of these components expressed in patterns suggestive of playing roles in focal establishment, these patterns are also both novel and violate compartmental (i.e., anterior or posterior) restrictions in the wing that might be expected to be strictly conserved. One factor that might be important in enabling so many novel gene expression patterns to evolve in butterflies is the co-option of entire circuits through recruitment of key individual components. We have speculated that the original recruitment of this circuit during eyespot evolution did not involve genetic changes at each of these loci but instead probably entailed changes at one or two key loci that led to deployment of Hh expression, which in turn induced changes in Ptc, Ci, and En expression through existing regulatory linkages (Keys et al. 1999). Recruitment of entire regulatory circuits through co-option of key individual genes would be a genetic shortcut in the evolution of morphological novelties.

A second factor important in the origin and diversification of wing color patterns is their uncoupling from earlier regulatory steps in wing formation. Color pattern formation can be viewed as a terminal addition that may be less constrained, because genetic and developmental changes affecting this process need not impact earlier stages, which may have greater effects on butterfly anatomy and fitness. Evidence supporting this idea comes from two observations. First, mutations have been identified that selectively affect eyespot patterns without affecting other parts of the wing or body (Brakefield et al. 1996, Brunetti et al. 2001). Changes in eyespot number or color scheme, for example, can arise without pleiotropic effects. Secondly, comparisons of gene expression in eyespot fields between species suggest that the regulatory linkages in eyespot development are unusually flexible. For example, two transcription factors, Spalt (Sal) and En, have been found to be expressed in concentric rings within eyespot fields (Figure 4*A*) (Brunetti et al. 2001). However, species with different eyespot color schemes deploy these proteins in different patterns; no transcription factor denotes any particular color or ring identity across species.

It is particularly noteworthy that proteins such as Dll, En, and Sal, which play such conserved roles in insect development, are deployed in such diverse ways in the widely varying eyespots of butterflies. The genetic regulatory system that governs eyespot development appears to be remarkably plastic and relatively unconstrained compared with that acting in the formation of body parts and tissues. This may, in turn, facilitate the occurrence of subsequent gene co-option events.

Hox Genes and Tetrapod Limb Evolution

One of the best-studied morphological novelties from both a paleontological and developmental standpoint is the tetrapod limb, which evolved from the paired fins of vertebrates over a span of about 10-million years in the Late Devonian (Shubin et al. 1997). The development of tetrapod limbs shares many regulatory features; one of the most important is the spatiotemporal deployment of *Hox* genes in the

developing limb buds. The so-called posterior *Hox* genes of paralog groups 9–13 are utilized in patterning posterior axial morphology. Subsequent to this, in developing limb buds, the *HoxA*9-13 and *HoxD*9-13 genes are expressed again during regional specification of the limb field. This later deployment of the *Hox* genes appears to be the product of evolutionary co-option (reviewed by Zakany & Duboule 1999, Herault et al. 1999, Morgan & Tabin 1994).

The tetrapod limb is composed of three distinct elements along its proximodistal axis. Whereas the two most proximal elements have homologs in fish fins and exhibit similar patterns of Hox expression, the most distal element, the autopod, which will form the hand/feet and digits, is a novel structure (Figure 4*B*). Interestingly, Sordino et al. (1995) have shown that a third spatiotemporal phase of posterior *HoxD* gene deployment characteristic of tetrapod autopods is absent in bony fish, suggesting that this regulatory program and limb element arose after the divergence of tetrapods from bony fish.

This later phase of Hox expression in the limb bud is unique to tetrapods and suggests that a novel regulatory circuit evolved in the tetrapod lineage. Recently, Spitz et al. (2001) have shown that whereas the *cis*-regulatory DNA sequences that control axial expression of *Hox* genes and early deployment in the proximal limb bud are nested within the *HoxD* complex, the sequences involved in expression in the distal limb are located at remote positions outside of the complex. This evidence indicates that the evolution of a novel limb-specific *cis*-regulatory control region played a key role in the co-option of *Hox* genes during the evolution of this critical vertebrate novelty.

Evolution of Complex Leaves in Plants

Co-option of a developmental circuit containing a homeobox transcription factor has also been involved in the transition from simple to complex leaves, which are a major adaptive feature of angiosperm body plans (reviewed by Bharathan & Sinha 2001, Byrne et al. 2001) (Figure 4*C*). KNOX1, a member of the Knotted homeobox protein family, is expressed in the apical meristem in association with the maintenance of indeterminant growth of the primary axis. In the leaf primordium, growth changes from indeterminant to determinant, which requires downregulation of KNOX1 expression, as indicated by the observation that inappropriate expression of KNOX1 in leaf primordia leads to increased leaf ramification (Sinha et al. 1993, Chuck et al. 1996). In the developing compound leaves of tomato, leaflet primordia form as distal bulges on the main leaf primordium, and expression of KNOX1 re-appears in these regions (Hareven et al. 1996). Although the precise role of KNOX1 in this process is not understood, it appears that a growth control pathway involving KNOX1, originally used in the primary axis of the plant, has been co-opted in the evolution of compound leaves. Complex leaves evolved several independent times in angiosperms (Goliber et al. 1999, Bharathan & Sinha 2001), but there are intriguing differences in the developmental underpinnings of different cases. In the pea, compound leaf development is not associated with KNOX1 expression. Instead, a circuit involving FLORICAULA/LEAFY, a

transcription factor required for determination of the floral meristem, appears to have been co-opted (Gourlay et al. 2000).

Sonic Hedgehog and the Evolution of Feathers

Co-option of a cell signaling module involving the morphogen Sonic hedgehog (Shh), an ortholog of arthropod Hedgehog, appears to underlie the evolution of avian feathers. Feathers are morphogenetically complex but are homologous with the much simpler scales of reptiles and are thought to have evolved from a scale-like epidermal appendage (Prum 1999, Chuong et al. 2000, Brush 2000). This derivation of complex feathers from simple scales involved a series of morphogenetic novelties involving expression of the Shh circuit that was hierarchically contingent; that is, novelties appearing later in development (and in evolution) modified earlier appearing morphogenetic changes. Shh functions in specifying polarity in many organ systems (Hammerschmidt et al. 1997), and its signaling is mediated by the *TGF-beta/dpp* family member, bone morphogenetic protein 2 (Bmp2). M. Harris (personal communication) has found that the development of epidermal appendages in archosaurs (birds, dinosaurs, and alligators) utilizes the integrative signaling of these morphogens to pattern and control appendage morphogenesis in a conserved manner. Epidermal placodes, which are the anlagen of both reptile and bird scales as well as bird feathers, express Shh and Bmp2 in posterior and anterior domains, respectively, reflecting the inherent polarity of these structures (Figure 4*D*). Furthermore, Bmp2 negatively regulates Shh expression in both the early placode stage of scales and in feather outgrowth, which indicates that these signaling pathways are functionally integrated and that this relationship is conserved in different developmental contexts (M. Harris, personal communication). At the conical bud stage of feather development, Shh and Bmp2 are expressed in largely nonoverlapping longitudinal stripes prepatterning the developing structures that will form the feather barbs (Figure 4*D*). Finally, the dynamic expression of Shh during late feather morphogenesis becomes even more complex and corresponds with observable phenotypic variation in feather-branching patterns (Figure 4*D*). M. Harris (personal communication) has found that within this novel developmental context, Shh and Bmp2 signaling is required for barb morphogenesis and that the function of these two pathways is integrated and similar to that seen in initial patterning of the placodes. The dynamic expression of the Shh circuit during feather morphogenesis suggests that feathers evolved from scale-like epidermal precursors by co-opting this signaling pathway, perhaps multiple times.

VARIATION IN GENE EXPRESSION: THE RAW MATERIAL OF CO-OPTION

We have seen that gene co-option has been a widespread and important process for both ancient and recent evolution. The ubiquity of these phenomena suggests that the variation fueling co-option may be present and recognizable as variation

within species and between closely related species. Alternatively, major genetic changes, such as gene duplications and transpositions, may be singular, atypical events, but when occurring, may persist long enough to be of adaptive value. These hypotheses are not mutually exclusive, and if either is true, then one would expect that regimes of gene expression might display some flexibility, such that proteins are occasionally or even routinely expressed in tissues in which they are not required. This kind of variation should be neutral with respect to individual fitness and may be present either transiently, as its underlying allelic variation is subjected to genetic drift, or for long periods, in cases where gene regulatory elements mediating required gene expression in some tissues also cause pleiotropic expression in tissues where the gene product is not required.

One of the first investigations of cryptic variation in gene expression patterns came from the seminal work of Dickinson and colleagues on enzyme expression in *Drosophila*. (Dickinson 1980; reviewed in Dickinson 1988). In a survey of the expression patterns of 6 enzymes in 14 different tissue types in 27 species, Dickinson found, surprisingly, that most enzymes showed substantial levels of variation in tissue-specific expression. Roughly 30% of the enzyme expression traits (i.e., expression of a particular enzyme in a particular tissue) showed at least one departure from a standard pattern in at least one species. Indeed, the majority of species pairs differed in expression in at least one tissue for each enzyme. The functional significance of this unexpectedly high level of variation among closely related species is unknown, but Dickinson (1980, 1988, 1991) speculated that at least some variants might be neutral. No obvious correlations could be found between expression pattern variation and ecological divergence among these species.

Several other studies in *Drosophila* suggest that this type of variation may be common. A similar variation in several enzyme expression patterns has been reported in the *Drosophila melanogaster* species group (Morton & Singh 1985, Ross et al. 1994, Wang et al. 1996), in midgut amylase expression in the *D. obscura* and *willistoni* groups (Powell et al. 1980), and aldehyde oxidase expression in imaginal discs of Hawaiian *Drosophila* (Ahearn & Kuhn 1981). Interestingly, Thorpe et al. (1993) found that the degree of enzyme pattern variation was about 60% higher in Hawaiian *Drosophila* than among species in the *D. virilis* group. Most of this difference was because of extremely high levels of variation in two enzymes, alcohol dehydrogenase and aldehyde oxidase I, in the Hawaiian species. Thorpe et al. (1993) speculated that more rapid and recent evolution in the Hawaiian species may have contributed to their relatively high rate of variation. Population bottlenecks and founder effects, which probably have played a critical role in the last 15 my during the radiation of Drosophilids in Hawaii, may have resulted in the chance fixation of many variants that were not crucial to fitness in various lineages. It will be of particular interest to determine if similar levels of variation occur in other types of genes, including those encoding transcription factors and signaling proteins, which are central to the regulation of pattern formation in metazoans.

THE EVOLUTIONARY BIOLOGY OF CO-OPTION

Co-option is a long-recognized phenomenon in evolution (reviewed by Ganfornina & Sanchez 1999). Several recent authors have suggested that a relationship exists between the toolkit of gene and pathway functions and the "evolvability" or adaptive potential of a taxon or lineage (Wagner & Altenberg 1996, Gerhardt & Kirschner 1997, Kirschner & Gerhardt 1998, von Dassow & Munro 1999). The implication then is that co-option is the mode by which elements of the toolkit are utilized; i.e., co-option is how evolvability is manifested. This framework, while providing a metaphorical image of natural selection operating a switchboard of genes, pathways, and cell types, has not, however, been explicit as to either the developmental or population level mechanisms underlying co-option. If the mode and tempo of morphological change are generally influenced by the richness of the developmental genetic toolkit, then variation in developmental process within species must be the raw material necessary for trajectories of co-option.

Variation in Developmental Systems

Developmental evolutionists have yet to address the extent of variation in developmental systems or the importance of this variation in evolution (Stern 2000, von Dassow & Munro 1999). However, it is clearly possible from the co-option case studies and the *Drosophila* enzyme expression surveys discussed above that ample variation for the evolution of novel gene functions may be routinely present in many organisms. This variation can occur in the spatiotemporal expression of genes, owing to variation in *cis*-regulatory DNA elements, in the deployment of *trans*-acting regulatory proteins (which may themselves be due to *cis*-regulatory variation), or in polymorphisms in protein sequences. Importantly, in order for these variations to become subject to co-option, they must either be selectively neutral or, at worst, only slightly deleterious (such that they could become fixed in a population by genetic drift or linkage to advantageous alleles). These selective constraints on variation are much better understood for protein-coding sequences, which typically are highly conserved within species, than for regulatory DNA. An important remaining question then, is how much leakiness or noise is permissible in gene expression patterns. This permissiveness is expected to depend on the function of the gene product. We have seen in the case of crystallins and *Drosophila* metabolic enzymes that substantial variation may occur in the expression patterns of structural proteins. It is unknown, however, whether developmental regulatory proteins show the same degree of variability. The cases of apparent co-option of developmental pathways discussed above may have resulted from a relatively small number of changes in the expression of regulatory proteins. We have schematized two possible mechanisms by which this might occur in Figure 5. If selectively neutral expression of a developmental regulatory protein exists after a critical window of function (Figure 5*A*), then subsequent changes bringing target genes under control of the regulator (e.g., by changes in *cis*-regulatory DNA) in those tissues have the

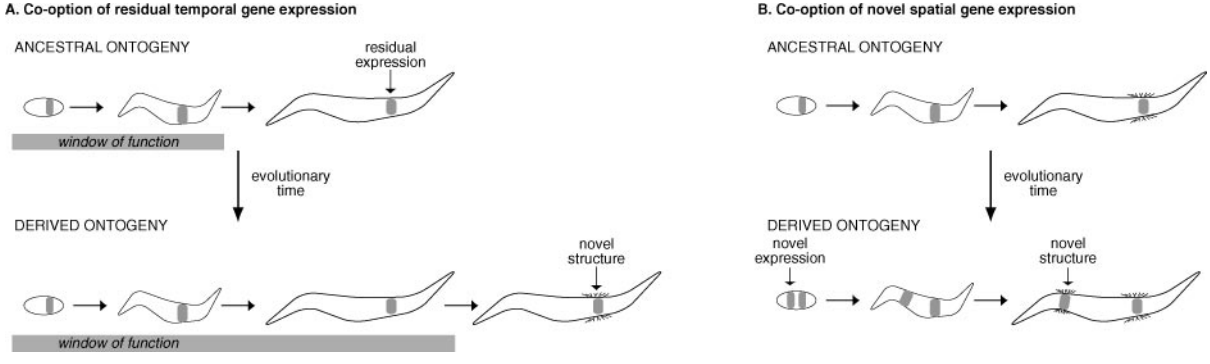


Figure 5 Two modes by which neutral variation in expression of a developmental regulatory gene may be co-opted to promote novel morphogenetic features. *A*. Gene expression that persists after a critical period in which the function of the gene is required may be utilized during evolution to regulate novel morphogenetic processes. This results in terminal addition of a novel feature and the lengthening of the critical period of gene function. *B*. A novel, initially neutral, expression domain may arise as a pleiotropic effect of evolutionary change in another developmental program. This novel expression may be co-opted to promote morphogenesis of an ancestral feature in the novel region of the body.

potential to produce morphogenetic novelty. This is an example of terminal addition, mentioned above in the case of butterfly eyespots. Alternatively, if novel spatial expression of the gene arises in regions where it was not originally required, traits may arise in novel locations. This novel spatial gene expression could originate, for example, as a pleiotropic side-effect of an adaptive genetic change (Figure 5*B*). If targets of the regulatory protein are themselves regulatory proteins with the ability to activate further genes, then multigenic cascades of novel expression may arise. It is expected that in most cases changes in expression of developmental genes would be deleterious, but given the evidence we have described above, it seems likely that rare co-options of this kind do occur.

Gene Duplication and Selective Conflict

Most of the cases of co-option so far discussed have occurred in the context of diversification of multigene families, in which different members have evolved tissue and developmental stage-specific functions. The structure of *cis*-regulatory regions controlling expression of different gene family members lends important dynamics to the evolution of multigene families. In particular, the modular organization of gene regulatory regions (von Dassow & Munro 1999, Dynan 1989, Davidson 2001, Carroll et al. 2001) is an important factor in decreasing the pleiotropic effects of genetic variants and contributes to the ability of gene regulatory regions to be specialized for specific tissues. When a gene is duplicated, one of the copies may be free to adopt new expression patterns and functions, which lead to novel traits. Alternatively, the sibling duplicates may "split up" the original function between them (see Figure 1*C*,*D*). Force et al. (1999) and Lynch & Force (2000) recently modeled this process, which they call subfunctionalization, and found that it may allow gene duplicates to persist longer in evolutionary time than was previously thought (Li 1980). Importantly, the probability of subfunctionalization (as opposed to destruction of one of the duplicates) increases with the size and regulatory complexity (i.e., number of enhancers with distinct spatiotemporal expression) of genes relative to their coding regions (Force et al. 1999). Additionally, the expected time to subfunctionalization is expected to decrease with regulatory complexity of the gene (Force et al. 1999), thus decreasing the time window in which a gene duplicate may be lost due to redundancy and increasing the subsequent persistence time of that gene. This propensity of subfunctionalization to extend the life of gene duplicates is important for the evolution of novel gene functions by co-option. Because the appearance of new functions depends on rare beneficial mutations, the longer time a gene duplicate persists, the greater chance it has of eventually evolving a new function, either by regulatory or structural mutation (Force et al. 1999).

Recent reviews of co-option (Ganfornina & Sanchez 1999, Eizinger et al. 1999) have pointed out that gene duplication has important selective implications for gene function. In particular, a single shared or co-opted bifunctional gene is expected to experience selective conflict or differential selection pressures from the two

distinct functions. Duplication followed by specialization of the two paralogous daughter genes for the two functions relieves this conflict. There is ample evidence of both shared multifunctional genes and specialization following duplication, which suggests that selective pressures and conflicts may vary widely among co-opted systems. Nevertheless, the apparent facility with which gene duplication events have fueled gene co-option indicates that they have often been available for selection to act upon. A recent theoretical analysis by Lynch et al. (2001) estimates that the probability of permanent retention of a gene duplicate, in the absence of selection against it, is surprisingly high. This probability is no lower than half the fixation rate of any neutral allele, and possibly much higher if new, advantageous mutations are common. So we are left with a paradox: Comparative developmental and genomic data provide vast evidence that gene families have diversified and specialized into adaptively important roles, and theory suggests that once present, gene duplications have substantial probabilities of survival. However, the mechanisms by which duplication and transposition bring about co-option of novel gene functions have thus far been hidden from view because functionally important polymorphisms involving these events are difficult to identify. The next phase of evolutionary developmental biology research must address this paradox by investigating the levels, causes, and consequences of microevolutionary variation in developmental systems.

CONCLUSION

Our goals herein have been to summarize the empirical data on case studies of co-option and to examine the molecular events and evolutionary contexts in which genes have evolved new functions. It is clear that gene co-option has been rampant in the evolution of complex, multicellular organisms, and phylogenetic patterns strongly suggest that many of these events are associated with major changes in ecology and life history. Thus the processes of co-option and regulatory diversification, rather than invention of new genes, explain to a great degree why total gene number in the most complex organisms, such as mammals, does not greatly exceed that of simpler organisms such as flies and worms. Furthermore, the apparent facility and regularity with which some types of co-option have taken place, such as crystallins and antifreeze proteins, indicate that in some cellular and developmental contexts the ingredients for co-option have been plentiful during evolution. This strongly implies that variation, in the form of both noise in spatiotemporal gene expression patterns and transient functional redundancy of duplicated genes, has been readily available for natural selection to act upon.

The present convergence of evolutionary developmental biology and genome level analyses provides the stage upon which a deeper understanding of gene functional and regulatory evolution can be played out. In order to better understand the evolutionary and functional mechanisms surrounding co-option, studies of the phenotypic effects and extent of developmental variation in natural populations are

needed. This variation is expected to encompass both *cis*-regulatory and protein-coding DNA. Quantitative and molecular genetic approaches must be integrated and combined with transgenic techniques to understand genotype-phenotype relationships. Specialization among members of multigene families has already been widely implicated as central to co-optive evolution. Comparative genomics can provide an indication of how both small- and large-scale genomic changes fuel the evolution of novelties. Integrating this information across taxa will indicate the propensity for genetic and genomic experimentation to occur in extant organisms and will provide clues about the genomic and selective circumstances surrounding ancestral co-option events.

ACKNOWLEDGMENTS

We thank Barry Williams, Geeta Bharathan, and Thomas Merritt for many valuable comments and discussions during preparation of this manuscript. We are also very grateful to Matthew Harris for sharing unpublished data. LeeAnn Olds and Craig Brunetti provided images and assistance in the preparation of figures. J.R.T was supported by a postdoctoral fellowship from the National Institutes of Health. S.B.C. is an Investigator of the Howard Hughes Medical Institute.

The *Annual Review of Cell and Developmental Biology* is online at http://cellbio.annualreviews.org

LITERATURE CITED

Abramson RD, Barbosa P, Kalumuck K, O'Brien WE. 1991. Characterization of the human argininosuccinate lyase gene and analysis of exon skipping. *Genomics* 10:126–32

Ahearn JN, Kuhn DT. 1981. Aldehyde oxidase distribution in picture-winged Hawaiian *Drosophila*: evolutionary trends. *Evolution* 35:635–46

Ball CA, Cherry JM. 2001. Genome comparisons highlight similarity and diversity within the eukaryotic kingdom. *Curr. Opin. Chem. Biol.* 5:86–89

Beldade P, Brakefield PM, Long AD. 2002. Contribution of *Distal-less* to quantitative variation in butterfly eyespots. *Nature* 415: 315–18

Bevan M, Mayer K, White O, Eisen JA, Preuss D, et al. 2001. Sequence and analysis of the *Arabidopsis* genome. *Curr. Opin. Plant Biol.* 4:105–10

Bharathan G, Sinha R. 2001. The regulation of compound leaf development. *Plant Physiol.* 127:1–5

Brakefield PM, Gates J, Keys D, Kesbeke F, Wijngaarden PJ, et al. 1996. Development, plasticity and evolution of butterfly eyespot patterns. *Nature* 384:236–42

Brakefield PM, Reitsma N. 1991. Phenotypic plasticity, seasonal climate and the population biology of *Bicyclus* butterflies (Satyridae) in Malawi. *Ecol. Entomol.* 16:291–304

Brakenhoff RH, Aarts HJ, Reek FH, Lubsen NH, Schoenmakers JG. 1990. Human gamma-crystallin genes. A gene family on its way to extinction. *J. Mol. Evol.* 216:519–32

Brown TA. 1999. *Genomes.* New York: Wiley & Sons

Brunetti C, Selegue JE, Monteiro A, French V, Brakefield PM, et al. 2001. The generation and diversification of butterfly eyespot colour patterns. *Curr. Biol.* 16:1578–85

Brush AH. 2000. Evolving a protofeather and feather diversity. *Am. Zool.* 20:631–39

Byrne M, Timmermans M, Kidner C, Martienssen R. 2001. Development of leaf shape. *Curr. Opin. Plant Biol.* 4:38–43

Carlini DB, Reece KS, Graves JE. 2000. Actin gene family evolution and the phylogeny of coleoid cephalopods (Mollusca: Cephalopoda). *Mol. Biol. Evol.* 17:1353–70

Carroll SB. 2001. Chance and necessity: the evolution of morphological complexity and diversity. *Nature* 409:1102–9

Carroll SB, Grenier JK, Weatherbee SD. 2001. *From DNA of Diversity: Molecular Genetics and the Evolution of Animal Design.* Malden, MA: Blackwell Science

Chen L, DeVries AL, Cheng CH. 1997a. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc. Natl. Acad. Sci. USA* 94:3811–16

Chen L, DeVries AL, Cheng CH. 1997b. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proc. Natl. Acad. Sci. USA* 94:3817–22

Cheng CH. 1998. Evolution of the diverse antifreeze proteins. *Curr. Opin. Genet. Dev.* 8: 715–20

Cheng CH, Chen L. 1999. Evolution of an antifreeze glycoprotein. *Nature* 401:443–44

Chuck G, Lincoln C, Hake S. 1996. KNAT1 induced lobed leaves with ectopic meristems when overexpressed in *Arabidopsis*. *Plant Cell* 8:1277–89

Chuong CM, Chodankar R, Widelitz RB, Jiang TX. 2000. Evo-devo of feathers and scales: building complex epithelial appendages. *Curr. Opin. Genet. Dev.* 10:449–56

Claverie JM. 2001. Gene number. What if there are only 30,000 human genes? *Science* 291:1255–57

Cvekl A, Piatigorsky J. 1996. Lens development and crystallin gene expression: many roles for Pax-6. *BioEssays* 18:621–30

Davidson EH. 2001. *Genomic Regulatory Systems: Development and Evolution.* London: Academic

de Jong WW, Hendriks W. 1986. The eye lens crystallins: ambiguity as evolutionary strategy. *J. Mol. Evol.* 24:121–29

den Dunnen JT, Jongbloed RJ, Geurts van Kessel AH, Schoenmakers JG. 1985. Human lens gamma-crystallin sequences are located in the p12-qter region of chromosome 2. *Hum. Genet.* 70:217–21

den Dunnen JT, van Neck JW, Cremers FP, Lubsen NH, Schoenmakers JG. 1989. Nucleotide sequence of the rat gamma-crystallin gene region and comparison with an orthologous human region. *Gene* 78:201–13

Deng G, Andrews DW, Laursen RA. 1997. Amino acid sequence of a new type of antifreeze protein, from the longhorn sculpin *Myoxocephalus octodecimspinosis*. *FEBS Lett.* 402:17–20

Desplan C. 1997. Eye development: governed by a dictator or a junta? *Cell* 91:861–64

Dickinson WJ. 1980. Evolution of patterns of gene expression in Hawaiian picture-winged *Drosophila*. *J. Mol. Evol.* 16:73–94

Dickinson WJ. 1988. On the architecture of regulatory systems: evolutionary insights and implications. *BioEssays* 8:204–8

Dickinson WJ. 1991. The evolution of regulatory genes and patterns in *Drosophila*. *Evol. Biol.* 24:127–73

Duboule D, Wilkins A. 1998. The evolution of 'bricolage.' *Trends Genet.* 14:54–59

Dubreuil RR. 1991. Structure and evolution of the actin crosslinking proteins. *BioEssays* 13:219–26

Dynan WS. 1989. Modularity in promoters and enhancers. *Cell* 58:1–4

Eastman JT. 1993. *Antarctic Fish Biology: Evolution in a Unique Environment.* San Diego: Academic

Eizinger A, Jungblut B, Sommer RJ. 1999. Evolutionary change in the functional specificity of genes. *Trends Genet.* 15:197–202

Fickett JW, Wasserman WW. 2000. Discovery and modeling of transcriptional regulatory regions. *Curr. Opin. Biotechnol.* 11:19–24

Fletcher GL, Hew CL, Davies PL. 2001. Antifreeze proteins of teleost fishes. *Annu. Rev. Physiol.* 63:359–90

Force A, Lynch M, Pickett FB, Amores A,

Yan Y, et al. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–45

French V, Brakefield PM. 1995. Eyespot development on butterfly wings: the focal signal. *Dev. Biol.* 168:112–23

French V, Daniels G. 1994. The beginning and the end of insect limbs. *Curr. Biol.* 4:34–37

Fujii Y, Kimoto H, Ishikawa K, Watanabe K, Yokota Y, et al. 2001. Taxon-specific zeta-crystallin in Japanese tree frog (*Hyla japonica*) lens. *J. Biol. Chem.* 276:28134–39

Ganfornina MD, Sanchez D. 1999. Generation of evolutionary novelty by functional shift. *BioEssays* 21:432–39

Gerhardt J, Kirschner M. 1997. *Cells, Embryos, and Evolution.* Oxford, UK: Blackwell Sci.

Gilbert W. 1978. Why genes in pieces? *Nature* 271:501

Gilbert W. 1985. Genes-in-pieces revisited. *Science* 228:823–24

Goliber T, Kessler S, Chen J-J, Bharathan G, Sinha N. 1999. Genetic, molecular, and morphological analysis of compound leaf development. *Curr. Top. Dev. Biol.* 43:260–90

Gonzalez P, Hernandez-Calzadilla C, Rao PV, Rodriguez IR, Zigler JSJ, et al. 1994. Comparative analysis of the zeta crystallin/quinone reductase gene in guniea pig and mouse. *Mol. Biol. Evol.* 11:305–15

Gourlay CW, Hofer JM, Ellis TH. 2000. Pea compound leaf architecture is regulated by interactions among the genes UNIFOLIATA, chochleata, afila, and tendril-less. *Plant. Cell* 12:1279–94

Graether SP, Kuiper MJ, Gagne SM, Walker VK, Jia Z, et al. 2000. Beta-helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature* 406: 325–28

Graw J. 1997. The crystallins: genes, proteins and diseases. *Biol. Chem.* 378:1331–48

Groger H, Callaerts P, Gehring WJ, Schmid V. 1999. Gene duplication and recruitment of a specific tropomyosin into striated muscle cells in the jellyfish *Podocoryne carnea*. *J. Exp. Zool.* 285:378–86

Hammerschmidt M, Brook A, McMahon AP. 1997. The world according to hedgehog. *Trends Genet.* 13:14–21

Harding JJ, Dilley KJ. 1976. Structural proteins of the mammalian lens: a review with emphasis on changes in development, aging and cataract. *Exp. Eye Res.* 22:1–73

Hareven D, Gutfinger T, Parnis A, Eshed Y, Lifschitz E. 1996. The making of a compound leaf: genetic manipulation of leaf architecture in tomato. *Cell* 84:735–44

Herault Y, Beckers J, Gerard M, Duboule D. 1999. *Hox* gene expression in limbs: colinearity by opposite regulatory controls. *Dev. Biol.* 208:157–65

Herbert A, Rich A. 1999. RNA processing and the evolution of eukaryotes. *Nat. Genet.* 21:265–69

Hew CL, Fletcher GL, Ananthanarayanan VS. 1980. Antifreeze proteins from the shorthorn sculpin, *Myoxocephalus scorpius*: isolation and characterization. *Can. J. Biochem.* 58: 377–83

Hodin J, Wistow G. 1993. 5′-RACE PCR of mRNA for three taxon-specific crystallins: for each gene one promoter controls both lens and non-lens expression. *Biochem. Biophys. Res. Commun.* 190:391–96

Huang S, McDowell JM, Weise MJ, Meagher RB. 1996. The *Arabidopsis* profilin gene family. Evidence for an ancient split between constitutive and pollen-specific profilin genes. *Plant. Physiol.* 111:115–26

Ingolia TD, Craig EA. 1982. Four small *Drosophila* heat shock proteins are related to each other and to mammalian alpha-crystallin. *Proc. Natl. Acad. Sci. USA* 79:2360–64

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921

Janssens H, Gehring WJ. 1999. Isolation and characterization of drosocrystallin, a lens crystallin gene of *Drosophila melanogaster*. *Dev. Biol.* 207:204–14

Kabsch W, Vanderckhove J. 1992. Structure and function of actin. *Annu. Rev. Biophys. Biomol. Struct.* 21:49–76

Kandasamy MK, McKinney EC, Meagher RB. 1999. The late pollen-specific actins in angiosperms *Plant J.* 6:681–91

Keys DN, Lewis DL, Seleague JE, Pearson BJ, Goodrich LV, et al. 1999. Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science* 283:532–34

Kim RY, Wistow GJ. 1993. Expression of the duck alpha-enolase/tau-crystallin gene in transgenic mice. *FASEB J.* 7:464–69

Kirschner M, Gerhart J. 1998. Evolvability. *Proc. Natl. Acad. Sci. USA* 95:8420–27

Knight CA. 2000. Structural biology. Adding to the antifreeze agenda. *Nature* 406:249–51

Lengler J, Krausz E, Tomarev S, Prescott A, Quinlan RA, et al. 2001. Antagonistic action of Six3 and Prox1 at the gamma-crystallin promoter. *Nucleic Acids Res.* 29:515–26

Li WH. 1980. Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fishes. *Genetics* 95:237–58

Long M. 2001. Evolution of novel genes. *Curr. Opin. Genet. Dev.* 11:673–80

Long M, Langley CH. 1993. Natural selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila*. *Science* 260:91–95

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459–73

Lynch M, O'Hely M, Walsh B, Force A. 2001. The probability of preservation of a newly arisen gene duplicate. *Genetics* 159:1789–804

McDowell JM, Huang S, McKinney EC, An YQ, Meagher RB. 1996. Structure and evolution of the actin gene family in *Arabidopsis thaliana*. *Genetics* 142:587–62

Meagher RB, McKinney EC, Kandasamy MK. 1999a. Isovariant dynamics expand and buffer the responses of complex systems: the diverse plant actin gene family. *Plant Cell* 11:995–1006

Meagher RB, McKinney EC, Vitale AV. 1999b. The evolution of new structures: clues from plant cytoskeletal genes. *Trends Genet.* 15: 278–84

Morgan BA, Tabin C. 1994. Hox genes and growth: early and late roles in limb bud morphogenesis. *Dev. Biol. Suppl.* 181–86

Morton RA, Singh RS. 1985. Biochemical properties, homology, and genetic variation of Drosophila 'nonspecific' esterases. *Biochem. Genet.* 23:959–73

Mounier N, Gouy M, Mouchiroud D, Prudhomme JC. 1992. Insect muscle actins differ distinctly from invertebrate and vertebrate cytoplasmic actins. *J. Mol. Evol.* 34: 406–15

Nijhout HF. 1980a. Ontogeny of the color pattern on the wings of *Precis coenia* (Lepidoptera: Nymphalidae). *Dev. Biol.* 80:275–88

Nijhout HF. 1980b. Pattern formation on lepidopteran wings: determination of an eyespot. *Dev. Biol.* 80:267–74

Nijhout HF. 1991. *The Development and Evolution of Butterfly Wing Patterns.* Washington, DC: Smithsonian Inst. Press

Nurminsky DI, Nurminskaya MV, De Aguiar D, Hartl DL. 1998. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature* 396:572–75

Panganiban G, Irvine SM, Lowe C, Roehl H, Corley LS, et al. 1997. The origin and evolution of animal appendages. *Proc. Natl. Acad. Sci. USA* 94:5162–66

Panganiban G, Sebring A, Nagy L, Carroll S. 1995. The development of crustacean limbs and the evolution of arthropods. *Science* 270:1363–66

Parker DS, Wawrousek EF, Piatigorsky J. 1988. Expression of the delta-crystallin genes in the embryonic chicken lens. *Dev. Biol.* 126:375–81

Patthy L. 1999. Genome evolution and the evolution of exon-shuffling—a review. *Gene* 238:103–14

Piatigorsky J, O'Brien WE, Norman BL, Kalumuck K, Wistow GJ, et al. 1988. Gene sharing by delta-crystallin and argininosuccinate lyase. *Proc. Natl. Acad. Sci. USA* 85:3479–83

Piatigorsky J, Wistow GJ. 1989. Enzyme/crystallins: gene sharing as an evolutionary strategy. *Cell* 57:197–99

Piatigorsky J, Zelenka PS. 1992. Transcriptional regulation of crystallin genes: *cis* elements, *trans*-factors, and signal transduction systems. *Adv. Dev. Biochem.* 1:211–56

Powell JR, Rico M, Andjelkovic M. 1980. Population genetics of *Drosophila* amylase. *Evolution* 34:209–13

Prum RO. 1999. Development and evolutionary origin of feathers. *J. Exp. Zool.* 285:291–306

Ross J, Fong PP, Cavener DR. 1994. Coordinate evolution of the *cis*-acting regulatory elements and developmental expression of the *Drosophila Gld* gene in seven species from the melanogaster subgroup. *Dev. Genet.* 15: 38–50

Scott GK, Davies PL, Kao MH, Fletcher GL. 1988. Differential amplification of antifreeze protein genes in the peuronectinae. *J. Mol. Evol.* 27:29–35

Shubin N, Tabin C, Carroll S. 1997. Fossils, genes and the evolution of animal limbs. *Nature* 388:639–48

Sidebottom C, Buckley S, Pudney P, Twigg S, Jarman C, et al. 2000. Heat-stable antifreeze protein from grass. *Nature* 406:256

Sinha N, Williams RE, Hake S. 1993. Overexpression of the maize homeobox gene, *KNOTTED-1*, causes a switch from determinate to indeterminate cell fates. *Dev. Genet.* 7:787–95

Skow LC, Donner ME, Huang SM, Gardner JM, Taylor BA, et al. 1988. Mapping of mouse gamma crystallin genes on chromosome 1. *Biochem. Genet.* 26:557–70

Sordino P, Hoeven FVD, Duboule D. 1995. *Hox* gene expression in teleost fins and the origin of vertebrate digits. *Nature* 375:678–82

Spitz F, Gonzales G, Peichel C, Vogt TF, Duboule D, et al. 2001. Large scale transgenic and cluster deletion analysis of the *HoxD* complex separate an ancestral regulatory module from evolutionary innovations. *Genes Dev.* 15:2209–14

Stern DL. 2000. Evolutionary biology. The problem of variation. *Nature* 408:529–31

Tautz D. 2000. Evolution of transcriptional regulation. *Curr. Opin. Genet. Dev.* 10:575–79

Thorpe PA, Loye J, Rote CA, Dickinson WJ. 1993. Evolution of regulatory genes and patterns: relationships to evolutionary rates and to metabolic functions. *J. Mol. Evol.* 37:590–99

Tomarev SI, Piatigorsky J. 1996. Lens crystallins of invertebrates—diversity and recruitment from detoxification enzymes and novel proteins. *Eur. J. Biochem.* 235:449–65

Tomarev SI, Sundin O, Banerjee-Basu S, Duncan MK, Yang JM, et al. 1996. Chicken homeobox gene *Prox1* related to *Drosophila prospero* is expressed in the developing lens and retina. *Dev. Dyn.* 206:354–67

Vale RD, Milligan RA. 2000. The way things move: looking under the hood of molecular motor proteins. *Science* 288:88–95

Vellai T, Vida G. 1999. The origin of eukaryotes: the differences between prokaryotic and eukaryotic cells. *Proc. R. Soc. London Ser. B* 266:1571–77

von Dassow G, Munro E. 1999. Modularity in animal development and evolution: elements of a conceptual framework for EvoDevo. *J. Exp. Zool.* 285:307–25

Wagner GP, Altenberg L. 1996. Complex adaptations and the evolution of evolvability. *Evolution* 50:967–76

Wang D, Marsh JL, Ayala FJ. 1996. Evolutionary changes in the expression pattern of a developmentally essential gene in three Drosophila species. *Proc. Natl. Acad. Sci. USA* 93:7103–7

Wistow G. 1990. Evolution of a protein superfamily: relationships between vertebrate lens crystallins and microorganism dormancy proteins. *J. Mol. Evol.* 30:140–45

Wistow G. 1993. Lens crystallins: gene recruitment and evolutionary dynamism. *Trends Biochem. Sci.* 18:301–6

Wistow G, Piatigorsky J. 1987. Recruitment of enzymes as lens structural proteins. *Science* 236:1554–56

Wistow GJ, Mulders JW, de Jong WW. 1987. The enzyme lactate dehydrogenase as a structural protein in avian and crocodilian lenses. *Nature* 326:622–24

Wistow GJ, Piatigorsky J. 1988. Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. *Annu. Rev. Biochem.* 57:479–504

Wistow GJ, Piatigorsky J. 1990. Gene conversion and splice-site slippage in the argininosuccinate lyases/delta-crystallins of the duck lens: members of an enzyme superfamily. *Gene* 96:263–70

Zakany J, Duboule D. 1999. *Hox* genes in digit development and evolution. *Cell Tissue Res.* 296:19–25

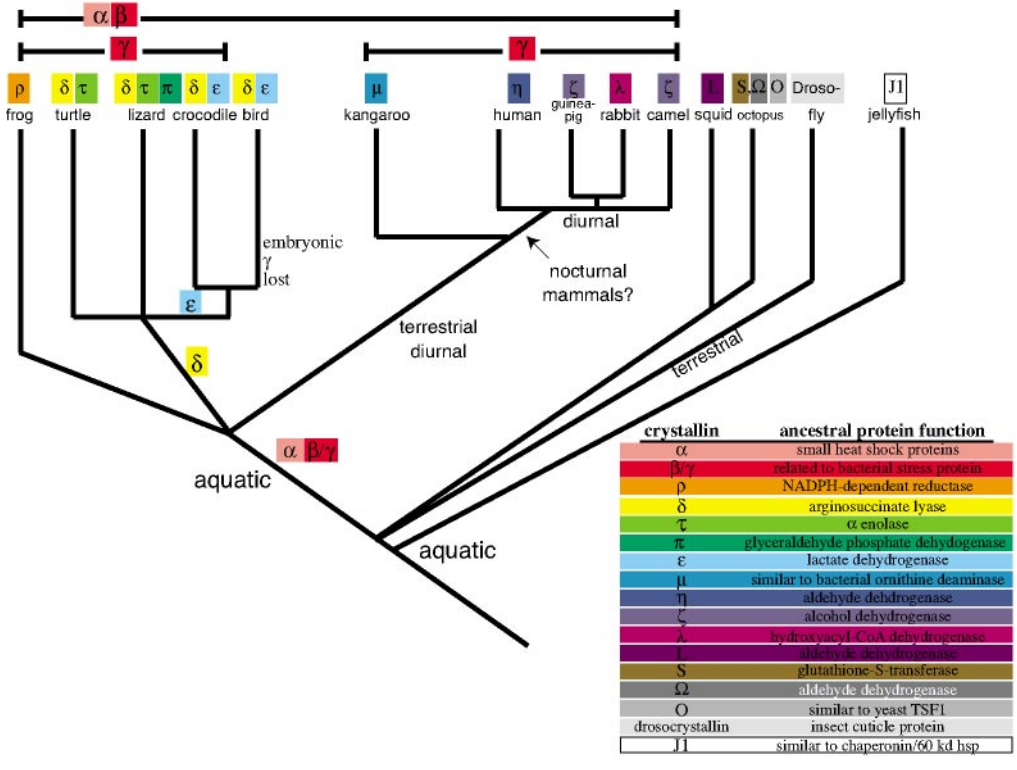


Figure 2 Animal lens crystallin diversity. For ubiquitous vertebrate crystallins (α, β, γ), bars across top indicate distribution (embryonic γ crystallin expression is absent in birds). Taxon-specific crystallins are shown only for the taxa in which they have been found and their complete distributions are not necessarily indicated. Two novel crystallins, J2 and J3, have also been found in jellyfish (Tomarev & Piatigorsky 1996). Figure adapted from Wistow 1993, with further information from Tomarev & Piatigorsky 1996 and Janssens & Gehring 1999.

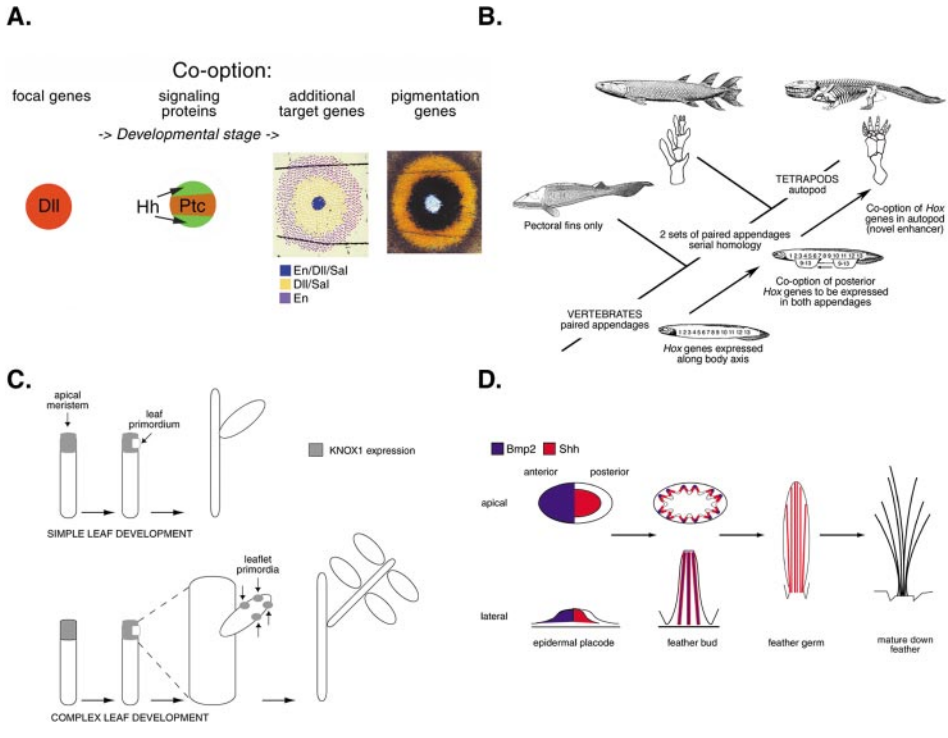


Figure 4 Co-option of developmental circuits in the evolution of novelties in multicellular organisms. See text for descriptions. (*A*) Butterfly eyespots are the developmental products of many co-option events that include the Hedgehog (Hh) signaling pathway, Distal-less (Dll), Spalt (Sal), Engrailed (En), and Patched (Ptc), as well as pigmentation enzymes. (*B*) Co-option of Hox A9-13 and D9-13 paralog sets during regional specification of the limb field in fishes and tetrapods. (*C*) Co-option of KNOX1, a plant homeobox transcription factor, in the evolution of complex leaves. (*D*) Co-option of the Sonic hedgehog (Shh) and bone morphogenetic protein 2 (Bmp2) signaling circuit during feather evolution.