Review

Cell PRESS

The gene regulatory logic of transcription factor evolution

Günter P. Wagner and Vincent J. Lynch

Department of Ecology and Evolutionary Biology, Yale University, PO Box 208106, New Haven, CT 06520-8106, USA

A growing debate in evolutionary and developmental biology concerns the relative contribution of *cis*-regulatory and protein (particularly transcription factor) changes to developmental evolution. Central to this argument are the perceived conservation of transcription factor functions and the modular architecture of *cis*-regulatory elements. In this paper, we review recent studies on transcription factor proteins that show that transcription factor genes undergo adaptive evolution and evolve novel functions that contribute to the evolution of development. Furthermore, we review experimental work that shows that transcription factor proteins are modular and can evolve with minimal pleiotropic effects. We conclude that changes in the function of proteins are likely directly contributing to developmental evolution.

Paradigms of gene regulatory evolution

The search for an explanation for the evolution of form has long motivated biologists [1], but it was not until the middle of last century that the tools to explore the molecular basis of morphological development and evolution became available. In one of the earliest studies to consider the molecular basis for morphological change, King and Wilson [2] compared the levels of morphological and protein divergence between humans and chimps and concluded that the level of protein divergence was too small to account for the anatomical differences between these two species. To reconcile the level of divergence between proteins and morphology, they proposed that morphological divergence was based mostly on changes in the mechanisms controlling gene expression and not changes in the protein-coding genes themselves.

The past 20 years have seen major advances in developmental genetics that have changed the way we approach evolutionary questions, in particular the evolution of morphological characters [3–5]. These advances have built upon the foundations of Wilson and colleagues and produced several generalizations about the relationship between genetic and phenotypic evolution. Among the most widely recognized is the concept of toolbox genes, that is that different body plans are realized with a conserved set of developmental genes, namely transcription factors and signaling molecules [4]. The second generalization, which in a way is a corollary of the first, is that toolbox genes do not change their functions during evolution, although their expression patterns can change. The last generalization to emerge is the concept that morphological evolution occurs through the gain and loss of transcription factor binding sites, a model that has come to be known as *cis*-regulatory evolution [6–8].

Like all scientific advances, these ideas have sparked a controversy about the relative importance and validity of these new generalizations [9]. Controversies play an important role in the progress of science by focusing the attention of researchers on important problems and stimulating research [10,11]. Ultimately, however, the scientific community has the responsibility to clear up the cognitive dissonance that provides the emotional fuel to the controversy and arrive at a balanced and well-supported view of reality. Here we want to contribute to this debate by proposing a pluralistic perspective on gene regulatory evolution. We begin with a short summary of the arguments for, and achievements of, the theory of *cis*-regulatory evolution (CRE) and briefly review the evidence

Glossary

***cis*-regulatory element**: a region of DNA that regulates the expression of genes located on that same strand of DNA. These usually consist of a collection of transcription factor binding sites but also includes sequence elements in the untranslated regions of the mRNA that affect translation and mRNA stability.
Domain: part of a protein that has a specific folded conformation that is relatively structurally independent from the remainder of the protein molecule. Large proteins are generally composed of many smaller domains, usually 25–500 amino acids in length. Domains are structurally, functionally and evolutionarily independent units and usually consist of smaller motifs.
Exonization: the process of generating a new exon, that is a part of the mature mRNA, either non-coding or protein-coding, from regions of the genome that were not previously expressed.
Functional equivalence: the ability of homologous genes from different species or paralogs to compensate for the function of each other in a specific context.
Functional motif: a functional element within a protein, such as small regions of proteins that perform specific functions, such as ligand binding, mediating protein–protein interactions, directing subcellular locations and so forth.
Functional specificity: the function of a protein in a specific context, usually referring to a particular function that is characteristic for that protein.
Modular architecture: independent or individualized functional units of DNA (such as regulatory elements) or proteins (such as motifs) that are connected together in such a way that individual modules can be replaced, added or deleted without affecting the proper function of the rest of the system or molecule.
Pleiotropy: the condition in which a gene or regulatory element has functions in more than one part of the body, organ, tissue or developmental period.
Short linear interaction motif (SLiM): a small functional motif, usually 3–10 amino acids long, that most often occurs in poorly structured regions of proteins (such as loops). SLiMs function in many biological processes, including subcellular localization, posttranslational modification and mediating of protein–protein interactions.
Simple sequence repeat (SSR): relatively short, microsatellite-like tandem repeats of an amino acid (codon) or nucleotide. Expansions in the length of SSRs have been implicated in several diseases, generally called repeat expansion diseases.
Structural motif: a three-dimensional structural element within a protein. Structural motifs usually consist of short stretches of amino acids and a few structural elements (such as α helices and β sheets).

Corresponding author: Wagner, G.P. (gunter.wagner@yale.edu).

0169-5347/$ – see front matter doi:10.1016/j.tree.2008.03.006 Available online 22 May 2008

put forward to support this view. This review leads us to question the validity of the generalization that transcription factor proteins are evolutionarily constrained. We present recent data showing that transcription factors evolve rapidly in response to selection, are intrinsically modular and highly evolvable and have contributed to morphological evolution. We conclude that transcription factor evolution has as much potential to contribute to developmental evolution as the evolution of *cis*-regulatory elements.

The case for the *cis*-regulatory paradigm

The case for the importance of *cis*-regulatory evolution has been eloquently made in several recent books and articles

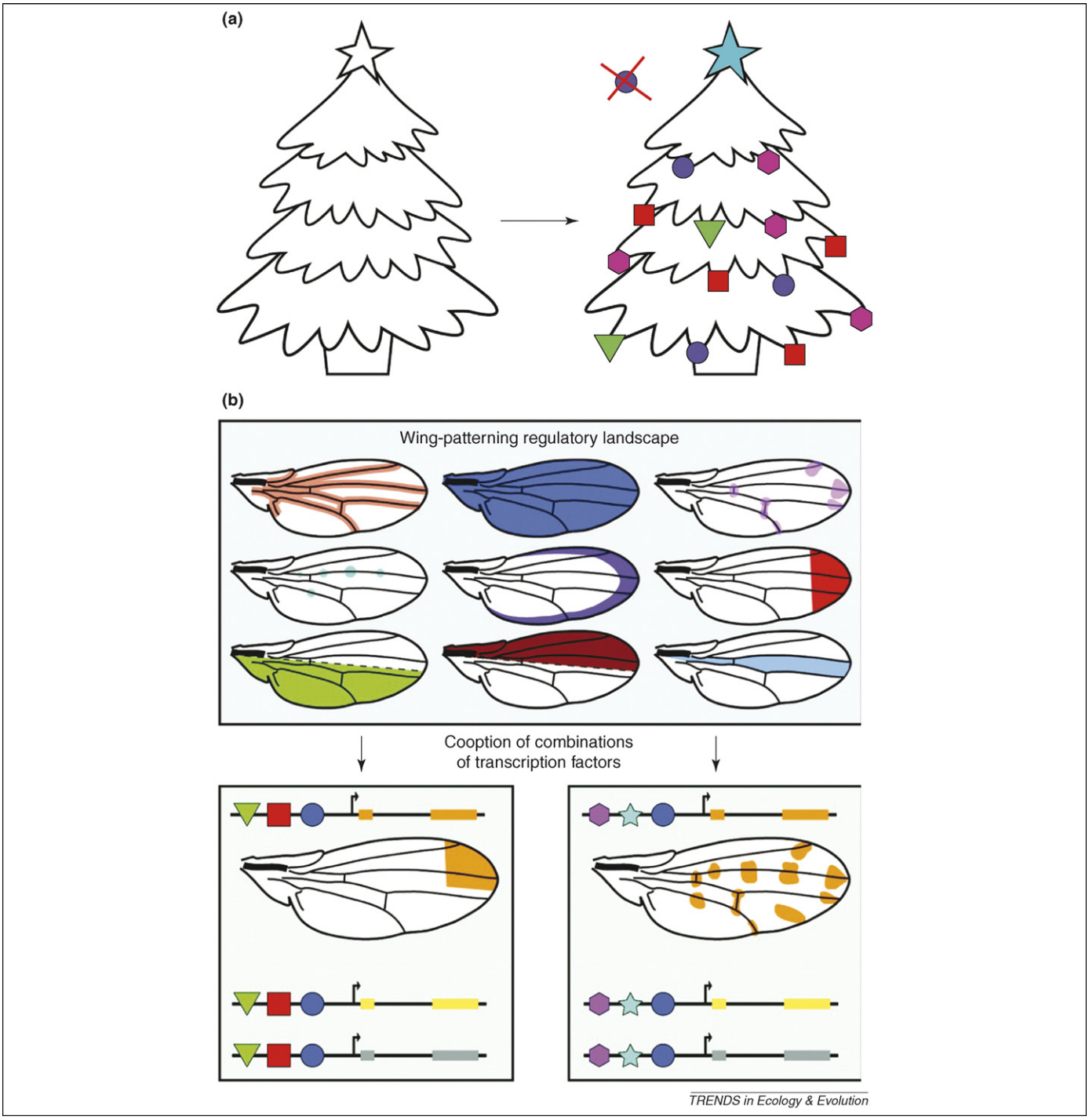

Figure 1. The 'Christmas tree model' of morphological evolution. **(a)** When decorating a Christmas tree, one is constrained to place ornaments where the tree provides branches. Although it is possible to choose where to place an ornament, the choices are limited by the structure of the tree. **(b)** Similarly, the evolution of pigment patterns in the *Drosophila* wing has been proposed to be constrained by the expression pattern of transcription factors, also called the 'regulatory landscape.' Pigment patterns are placed on the wing by hooking up the genes producing pigment to the pre-existing pattern of transcription factor expression, like the placement of an ornament on the branches of a Christmas tree (from Ref. [6]). The question that arose in the current discussion about the mode of gene regulatory evolution is whether, during evolution, the transcription factors themselves evolve to accommodate new target genes, or, in the language of the Christmas tree model, whether, in evolution, the 'branches' change to accommodate the 'ornaments.'.

[6,7,12,13]. Here we briefly summarize the main achievements of this research program to acknowledge its importance and provide the background for our critique.

The most complete studies on the role of *cis*-regulatory evolution come from the evolution of wing and abdominal pigmentation and bristle patterns in fruit flies. In a recent summary of this work, Prud'homme and colleagues [6] proposed a model of phenotypic evolution that we call the 'Christmas tree model of evolution' (Figure 1a). A Christmas tree is a structure that supports ornaments and lights on its branches. While decorating the tree, we have little control over where an ornament can be placed or what the overall structure of the tree is, but instead choose from among the existing branches when deciding where to put an ornament. Similarly, Prud'homme and colleagues show that an organ, such as a fly wing, has a 'regulatory landscape,' that is an underlying pattern of transcription factor gene expression domains that, like the branches of a Christmas tree, is used to support the activation or silencing of downstream target genes which ornament the wing. Thus, the resulting pigmentation pattern in the fly wing is preconfigured by the combinatorial possibilities of this regulatory landscape such that certain elements of the pigment pattern can be added or lost by the gain and loss of transcription factor binding sites (Figure 1b) in the regulatory regions of pigment-producing genes. The expression domains of these transcription factors are like the branches of a Christmas tree to which ornaments (pattern elements) can be hooked up.

The Christmas tree model explains the regularities of pigmentation pattern evolution and the relative ease with which specific pattern elements can be added or subtracted in evolution. It also provides an explanation of 'cryptic' (or latent) homology, a relatively obscure concept in comparative anatomy [14] (also called 'underlying synapomorph' [15]). Cryptic homology is the phenomenon in which apparently identical traits can independently appear in closely related organisms, even though there is no continuity of descent of this trait itself (examples include instances of parallelisms, reversals and atavisms). Hence, there is a potential to develop a heritable character that is not visible in the phenotype of an ancestor, but nevertheless can be induced, re-evolved or independently evolved. The conservation of the regulatory landscape underlying the evolution of complex characters, for example the *Drosophila* wing, is a perfect explanation of this otherwise obscure concept.

The Christmas tree model is supported by several detailed studies in *Drosophila*, all of which identified the causal nucleotide differences for phenotypic variation in *cis*-regulatory elements to the exclusion of transcription factor genes (i.e. in the placement of ornaments as opposed to the location of branches) [16–18]. This model, as well as the more general idea that *cis*-regulatory element evolution underlies phenotypic evolution, is further bolstered by theoretical considerations that provide a rationale for this view. Briefly, novel *cis*-regulatory elements can readily evolve from the modification of existing regulatory elements and are opportunistic, that is, any transcription factor that happens to be expressed in a cell can be used to regulate a gene given the opportunity (binding site) to do so. Most importantly, the modular architecture of *cis*-regulatory elements, with specific elements generally directing gene expression to particular spatial and temporal locations, allows mutational changes to have limited pleiotropy and thus have a high chance of being adaptive. By contrast, it is argued that mutations in broadly expressed proteins are likely to have multiple deleterious pleiotropic effects, severely limiting their role in the evolution of gene regulation and the evolution of development.

The take-home message of the *cis*-regulatory paradigm is clear: proteins are said to lack the organizational features that reduce pleiotropy and, therefore, changes in transcription factor proteins seem to be extremely unlikely to be adaptive. By contrast, *cis*-regulatory elements are modular, and therefore it is suggested that evolution of *cis*-regulatory elements is the most prevalent, if not the only, mode of developmental evolution. Below we will show that these assumptions about transcription factor proteins are wrong. In fact, transcription factors are as modular as *cis*-regulatory elements and do evolve in response to natural selection.

Challenges to the primacy of the *cis*-regulatory paradigm

We have briefly discussed some of the conceptual assumptions of the *cis*-regulatory paradigm. However, what remains to be done is to seriously interrogate the growing body of data implicating *cis*-regulatory element evolution as the primary source of phenotypic evolution to determine what the limits of this model of evolution are. Although numerous empirical results have been used in support of the *cis*-regulatory paradigm, perhaps none has been as influential as the finding that transcription factor proteins from organisms as divergent as flies and humans can be interchangeable (functionally equivalent) in certain experimental contexts. Before proceeding to a critique of the model, we will thus first review the argument that transcription factors remain functionally equivalent during evolution.

One of the main arguments for the conserved nature of transcription factors is the high level of protein sequence conservation reported for parts of the molecule. Indeed, highly conserved domains and motifs characterize transcription factor families and many other proteins; the homeodomain, a 60 amino acid DNA-binding domain common in many developmentally important transcription factors such as Hox genes, is the most prominent example. However, the homeodomain is composed of just 60 amino acids, which is only a small part of most transcription factors which are typically several hundred amino acids long. Sequence conservation outside of these conserved motifs can be extremely low. Thus, transcription factors are more accurately described as having islands of conservation in a sea of divergence. For example, the Hox protein Ultrabithorax (UBX) from velvet worm (*Akanthokara kaputensis*) and fruit flies (*Drosophila sp.*) differ in their homeodomains by only two substitutions, but amino acids outside of the homeodomain are not conserved except for two or three small motifs (Figure 2). Although it is possible that amino acids outside of the homeodomain are functionally irrelevant and only the homeodomain is func-

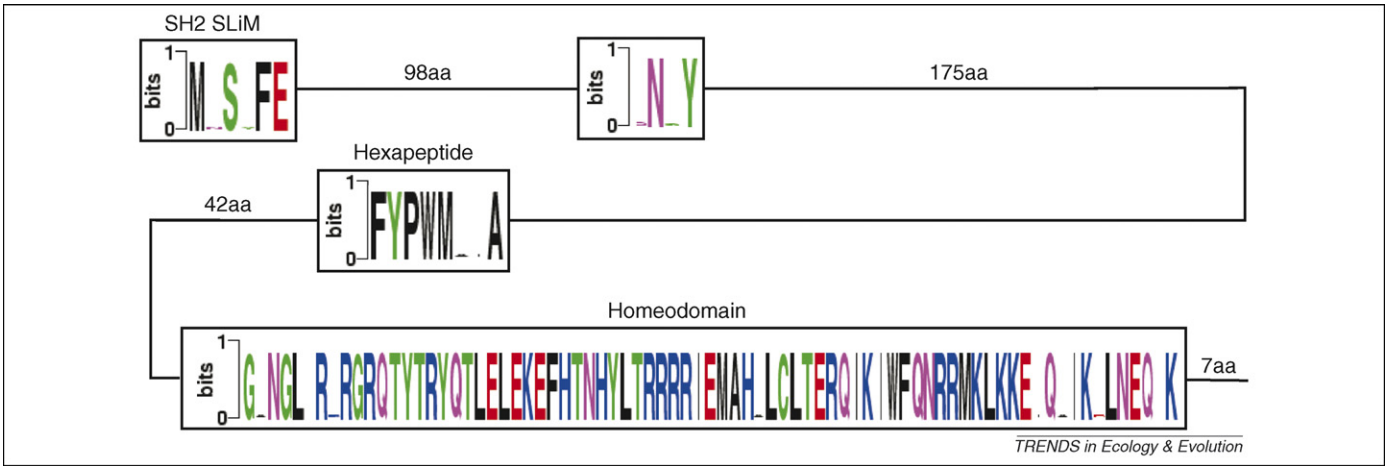

Figure 2. Islands of conserved motifs in a transcription factor. Sequence logo of the homeodomain protein Ubx fruit fly *Drosophila melanogaster* (P02834), water flea *Daphnia magna* (AE96992), brine shrimp *Artemia franciscana* (AE97000) and velvet worm *Akanthokara kaputensis* (AB92412). Sequence logos are graphical representations of multiple sequence alignments displayed on a single line. The height of each amino acid (displayed in single-letter notation) indicates the level of conservation of that position (shown as a bit score). Note that sequence similarity (bit score > 0) is limited to an SH2-like short linear motif (SH2 SLiM), an XNXY motif of unknown function, the hexapeptide short linear motif and the homeodomain. Conserved motifs are boxed and intervening regions with no conservation (bit score = 0) are shown as a line with the number of aligned amino acids between conserved motifs shown above the lines. GenBank accession numbers are given for each gene.

tionally important, we think this is unlikely for several reasons. First, the 'conserved' domain is itself variable at a broader phylogenetic scale. Second, there is evidence for adaptive evolution of amino acid residues outside the conserved domains, which shows that these residues must have some functional relevance, and lastly orthologous transcription factors have been shown to have divergent functions.

Even though the homeodomain is highly conserved, it is far from invariant between lineages and paralogs. Many of the amino acid differences between Hox gene homeodomains, for example, are likely to be involved in protein–protein interactions [19] and have signatures of functional differentiation [20], indicating that even so-called conserved domains can have different functions. The rate and pattern of sequence evolution in transcription factors is highly variable through evolutionary time and mosaic in the protein, with conserved regions intermixed with variable regions. In some cases, amino acid divergence is associated with signs of positive darwinian selection, such as elevated d_N/d_S ratios [21–24], providing evidence that the amino acid changes have beneficial functional effects. Accelerated rates of evolution and positive darwinian selection of transcription factor genes have been associated with biologically significant processes such as domestication [25], adaptive radiation [26], gene duplication [21–24] and morphological innovations [27].

Although protein conservation is often cited in support of the *cis*-regulatory paradigm, the most powerful arguments in its favor have been claims of functional conservation of transcription factors between divergent organisms. The classical approach to test for functional equivalence of transcription factors is to express a homologous gene from one species in another, usually in a model organism, and record its regulatory effects. This was first done by McGinnis and collaborators [28] with human *HoxD-4*, a semi-ortholog of the *Drosophila* homeotic gene *Dfd*. These authors showed that the auto-activation function of *Dfd*, a specific regulatory function of *Dfd* in *Drosophila*, can be replaced by the human *HoxD4*. This was the first evidence that some of the functions of the proteins in the Hox4 paralog group remained conserved over long periods of time. However, this conservation does not extend to all functions of transcription factors. To our knowledge, the first evidence of functional nonequivalence between transcription factor genes was the *Drosophila tinman* gene compared to its vertebrate homolog *Nkx2.5* [29], which both are involved in heart development in their respective species. Whereas *tinman*-null flies have *tinman*-dependent functions restored when injected with *tinman* mRNA, *Nkx2.5* only recovered a subset of *tinman*-specific functions. Thus, *tinman* and *Nkx2.5* are only partially equivalent and therefore must have distinct sets of target genes [29].

One of the best investigated cases of functional nonequivalence of orthologous transcription factor genes is *Ubx* in fruit fly *Drosophila melanogaster*, velvet worm *Acanthokara kaputensis* [30,31] and brine shrimp *Artemia* [32]. Carroll and colleagues [30,31] compared the *in vivo* activity of *Ubx* in the Onychophoran velvet worm (*O-Ubx*) to *Drosophila Ubx1a* (*D-Ubx1a*). Like *D-Ubx*, *O-Ubx* can transform an antenna into a leg and forewing into a haltere. But other typical effects of *D-Ubx* cannot be reproduced by *O-Ubx*. For example, ectopic *D-Ubx* expression transforms thoracic cuticle into abdominal cuticle, but *O-Ubx* was not able to cause this transformation. Similar results were obtained by McGinnis and collaborators in a comparison of *D-Ubx* and *Artemia Ubx* [32].

Functional divergence is not limited to orthologous genes between species; functional differences between the homeodomains of paralogous Hox genes within a species has also been demonstrated [33,34]. Zhao and Potter replaced the *HoxA-11* homeobox with the *HoxA-13* homeobox in the mouse and found that the HoxA-$11^{(A13Hd)}$ protein could functionally replace the *HoxA-11* homeodomain in the development of the vertebrae, ribs, kidney and male reproductive tract, but development of the female reproductive tract was abnormal. These results

were followed up with homeodomain swap experiments between *HoxA-11*, *HoxA-10* and *HoxA-4* that showed similar patterns of incomplete equivalence. The pattern emerging from these studies is that not even the homeodomain is functionally interchangeable between different Hox gene paralogs. The more divergent paralogs were nonequivalent in more tissues than the more closely related homeodomains. In addition, the transcription factor function showed a lower degree of equivalence in evolutionarily younger organs (such as the uterus and vagina) than in more ancient organs (e.g. the body axis).

Expanding the realm of the thinkable

Why has the evidence that transcription factors can change their functions after species diverge and genes duplicate not made a more substantial impact on evolutionary thought? We suspect that this omission results from the powerful logic of the *cis*-regulatory paradigm: mutations in protein-coding regions (particularly transcription factors) are argued to have highly pleiotropic effects, and therefore will be less likely a source of adaptive variation than mutations in *cis*-regulatory regions. However, as pointed out some years ago [35], this line of reasoning confounds the pleiotropic functions of genes and the pleiotropic effects of mutations: mutations in highly pleiotropic genes, that is those expressed in many tissues, need not have functional effects in every tissue the gene is expressed in [35]. For example, if some function of protein X is dependent on cofactor Y, a mutation in protein X that abolishes interaction with Y will only have observable effects in tissues that coexpress both genes. Examining protein structure and function in light of modularity indicates that transcription factor proteins are, in fact, highly modular and follow a combinatorial logic that is similar to that of *cis*-regulatory elements.

Escaping negative pleiotropy: alternative splicing

Much of the conceptual basis of the *cis*-regulatory paradigm depends on the modular structure of regulatory elements that reduces pleiotropic effects of mutations in regulatory regions. A particularly interesting feature of proteins is that motifs and domains tend to be encoded by a single exon, suggesting that alternative splicing of nonconstitutive (constant) exons can provide a means of escaping negative pleiotropy. Alternative splicing is increasingly recognized as a widespread mechanism that enables multiple structurally and functionally distinct proteins to be generated from a single transcript [36]. Tissue-specific alternative splicing is a potentially powerful mechanism to increase protein diversity and escape the negative consequences of pleiotropy, in a manner similar to the modular structure of regulatory elements.

Although alternative splicing produces proteins with different structural architectures, it does not necessarily lead to different functional specificities. However, several studies have found that alternative splicing does alter protein function, from minor functional tweaking to complete changes in function [37]. One of the most dramatic examples of a change in function is the human transcription factor AML1, which can act either as an activator or a repressor depending on isoform structure [37]. Splice patterns are also poorly conserved. For example, comparison of human/mouse ortholog gene pairs indicates that 80–90% of alternative spliced transcripts either have novel species-specific exons or species-specific splice variants. Several studies have demonstrated that species-specific exons are common in rodents [38,39] and humans [40] and originate at particularly high rates ($\sim 2.710^{-3}$ per gene per million years in rodents). The rate of exonization is dramatically higher than both the nucleotide substitution rate and gene duplication rate, suggesting novel exon formation can play an important role in generating phenotypic diversity.

Escaping negative pleiotropy: short linear motifs

Whereas alternative splicing can act to reduce the negative consequences of pleiotropy and lead to species-specific exons and isoforms, alternative splicing *per se* does not explain the divergence of transcription factor protein functions. To explain functional divergence of orthologs (after lineages split) and paralogs (after genes duplicate), we need to consider how proteins, particularly transcription factors, function. Although enzymes function by catalyzing reactions, transcription factors function by binding DNA and assembling protein complexes that recruit the transcriptional machinery. These functions are dependent on protein–DNA and protein–protein interactions, which are generally thought to be mediated by secondary structural motifs and large domain–domain contacts. However, a growing number of protein–protein interactions are being identified that are mediated by short linear motifs (SLiMs). The key feature of linear motifs is their small size, usually just 3–10 amino acids long with only 2 or 3 amino acids absolutely required for the interaction. Interestingly, SLiMs occur most often in poorly structured (unordered) regions of proteins, suggesting these motifs are relatively free from structural constraints [41] and thus can freely respond to functional needs.

SLiMs are also particularly evolvable. Their small size and lax sequence specificity means new functional linear motifs appear and disappear more easily than domains and structural motifs. Just a single mutation is often enough to convert a nonfunctional stretch of amino acids into a functional SLiM, giving these motifs a high degree of evolutionary plasticity and modularity [41]. A recent review [41] examined patterns of conservation in experimentally determined linear motifs across eukaryotes and found that whereas domain architecture was well conserved across species, linear motifs were poorly conserved between lineages. Linear motifs are also likely to evolve independently in unrelated proteins because their small size and low sequence specificity means any random mutation has a high probability of generating a new SLiM [41].

Perhaps the most dramatic example of a functional change in a transcription factor is the *Drosophila* Hox/HOM gene *fushi tarazu* (*Ftz*). Löhr and colleagues [42] ectopically expressed *Ftz* from the flour beetle *Tribolium* and grasshopper *Schistocera* in fruit flies to assess their potential to cause homeotic transformations and regulate segmentation. Whereas *Ftz* from the flour beetle and grasshopper functioned as homeotic genes, the *Drosophila* gene had no homeotic functions and instead functioned in seg-

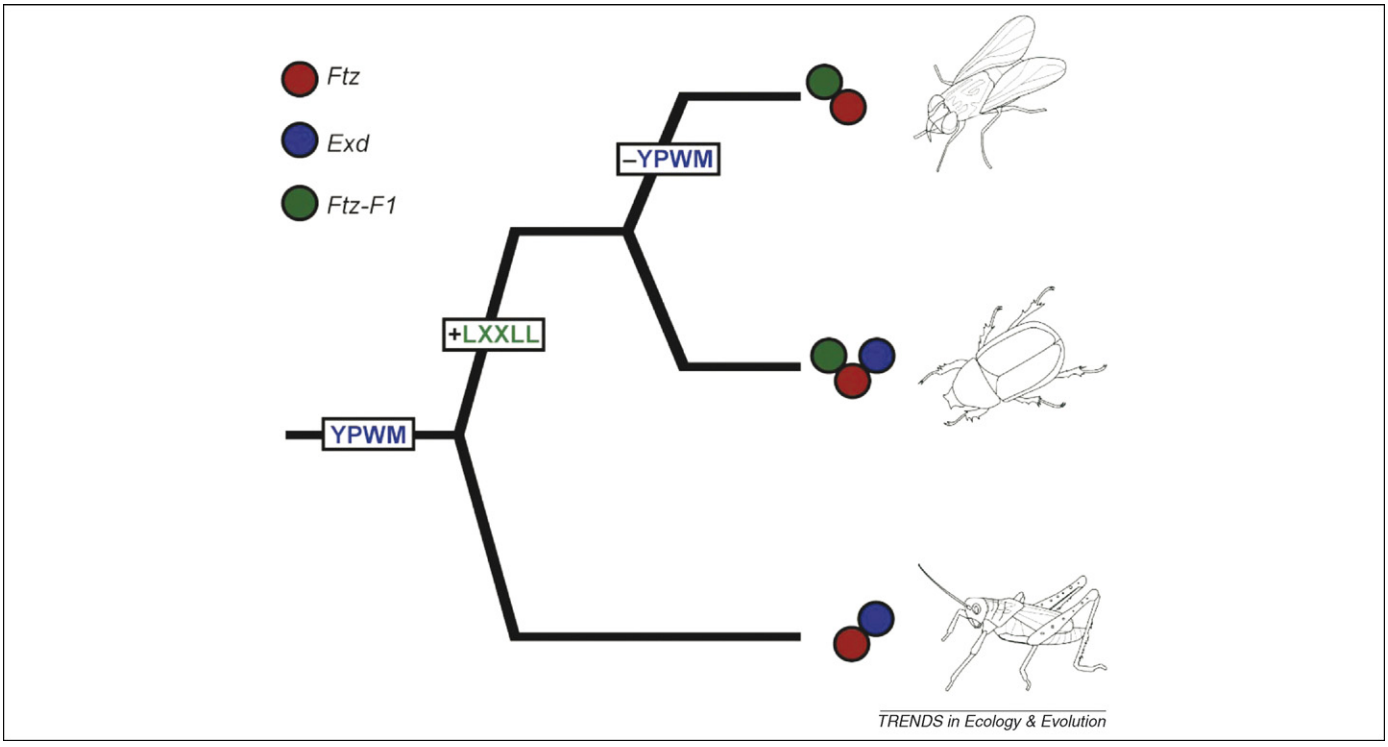

Figure 3. Short linear motifs (SLiMs) in evolution of the homeodomain protein Fushi tarazu (FTZ) [43]. Ancestrally, FTZ had a YPWM SLiM that is necessary for interaction with the extradentical protein (EXD). The FTZ–EXD interaction is required for the homeotic function of FTZ, is ancestral for insects and is still observed in grasshopper FTZ. The protein evolved an LXXLL SLiM for interaction with FTZ-F1 in the stem lineage of flies and beetles. A protein that has both SLiMs is found in *Tribolium* beetles. In the lineage leading to *Drosophila*, the EXD interaction SLiM was lost and only the Ftz-F1 SLiM retained. Interaction between FTZ and FTZ-F1 is necessary for segmentation function, whereas interaction between FTZ and EXD is required for homeotic function. This example shows how the evolution of a novel function in a transcription factor can occur with minimal pleiotropic effects via the gain and loss of SLiMs.

mentation (Figure 3). This change in *Drosophila Ftz* function is dependent on the ability of FTZ protein to interact with the cofactor FTZ-F1, an interaction mediated by the SLiM LXXLL. The FTZ/FTZ-F1 interaction motif LXXLL is present in *Drosophila* and *Tribolium* FTZ, but not *Schistocerca*. Loss of homeotic function in *Drosophila* Ftz, by contrast, is dependent on loss of the SLiM YPWM, which mediates the interaction of Ftz with extradenticle (EXD) [43]. Remarkably, the beetle FTZ protein has both the Ftz-F1 interaction motif and the EXD motif and has

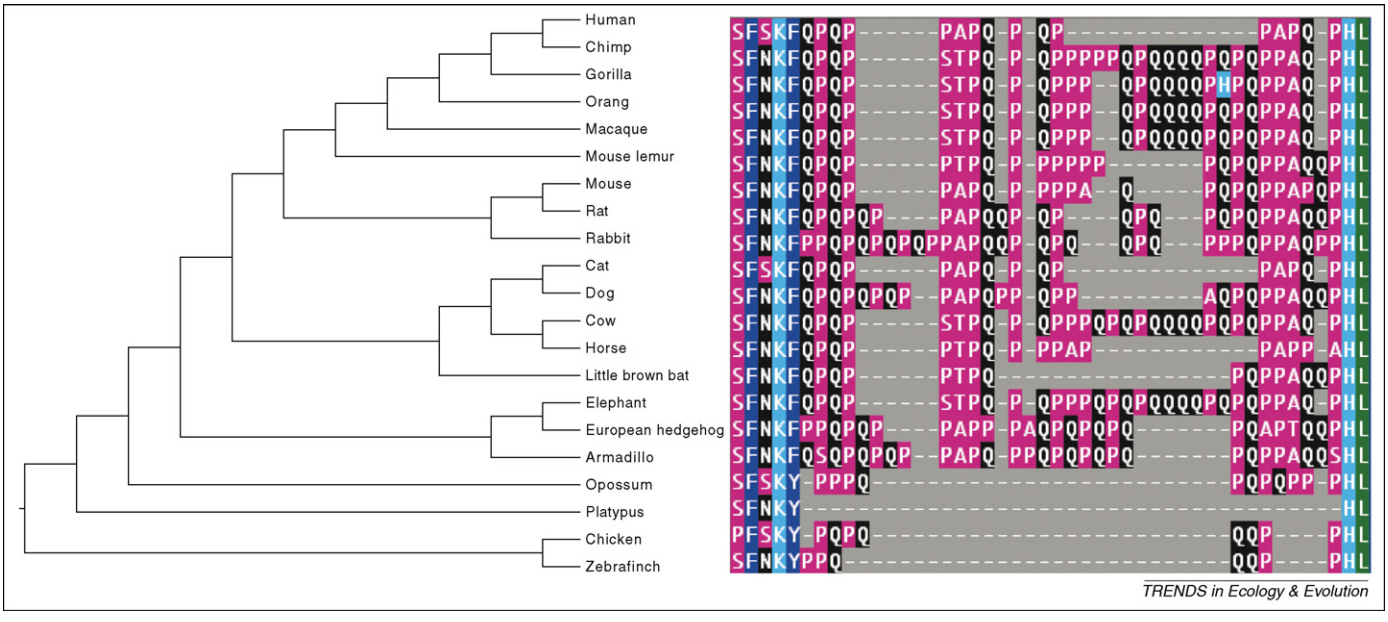

Figure 4. Rapid evolution of a simple sequence repeat (SSR) through gain and loss of amino acids in the ALX4 QP repeat of amniotes. Amino acids are colored according to physicochemical property. This region has been experimentally shown to mediate a protein–protein interaction with LEF-1 [49]. Note that repeats evolve rapidly and convergent gains and losses are common, suggesting SSRs are relatively free from negative pleiotropy. For example, the human lineage has lost 12 amino acids in two deletion events since the human–chimp divergence ~6 MYA.

both homeotic function and weak segmentation potential. These results suggest a stepwise model to explain the shift in *Drosophila* FTZ function: acquisition of a novel protein–protein interaction site, leading to a derived function that coexists with an ancestral function, followed by loss of the ancestral function via loss of the ancestral protein–protein interaction motif. This powerful example demonstrates that change in transcription factors need not have the kinds of deleterious pleiotropic effects usually attributed to changes in proteins, and actively contributes to developmental divergence between species.

Escaping negative pleiotropy: simple sequence repeats

Like linear motifs, simple sequence repeats (SSRs) are evolutionarily labile and often variable in length between species, and have therefore been called 'evolutionary knobs' that fine-tune transcription factor function. SSRs are microsatellite-like amino acid repeats that are particularly abundant in proteins that regulate gene expression and evolve rapidly [44–47]. For example, the rate of amino acid insertion and deletion in the glutamine-proline (QP) repeat of the *Alx-4* gene (also discussed below) is 2.7 indels per 100 million years, nearly 3× the rate of duplicate gene fixation, 5× the average nucleotide substitution rate and 10× the rate of novel exon formation (Figure 4). SSRs have also recently been implicated in generating morphological divergence. In a study of SSR variation in 17 developmental genes between 92 dog breeds, Fondon and Garner [48] found extraordinarily high levels of variation in the length of SSRs between different breeds. Furthermore, they found evidence that repeats were fixed in breeds more rapidly than expected under drift, and that these repeats were more diverse than expected under a neutral model. Although most of the variation was minor changes in repeat length, of usually two or three amino acids, five genes (*Six-3*, *HoxA-7*, *Runx-2*, *HoxD-8* and *Alx-4*) were found to have large expansions or contractions in SSRs in coding regions [48].

Although the function of most of these repeats is unknown, previous developmental and biomedical studies in mice and humans have suggested that mutations in Alx-4 can result in phenotypic effects. The Great Pyrenees dog breed, for example, is usually homozygous for the *Alx-4* Δ17aa mutation. An official characteristic of this breed is an extra toe on both hind feet (bilateral rear first digit polydactyly). All four Great Pyrenees with bilateral polydactyly examined by Fondon and Garner [48] were homozygous for the *Alx-4* Δ17aa mutation, whereas the single individual lacking extra dewclaws and the other 88 breeds had neither polydactyly nor the 17 amino acid deletion in *Alx-4*. This form of polydactyly is similar to that observed in *Alx-4* knockout mice. Amazingly, deletion of the QP repeat, which is reduced in size in the Great Pyrenees, in mice specifically abolished the ability of ALX-4 to bind with the cofactor LEF-1 and drive target gene expression in the limb bud [49]. These data indicate repeat length variation can have functional effects and suggest that changes in repeat length can alter gene expression by modulating protein–protein interactions between transcription factors.

Will a change in the length of simple amino acid repeats always lead to negative pleiotropy? Biomedical studies of repeat expansion diseases, a class of genetic diseases caused by expansion and contraction of SSRs, suggests that SSRs might have extremely few pleiotropic effects. For example, expansion of a polyalanine repeat in *HoxD-13* by 7–14 residues causes synpolydactyly, a dominant developmental limb deformity characterized by duplication of fingers and webbing between fingers [50–53]. No other organs or tissues are affected. Combined with the data that SSRs can mediate protein functions, these results suggest that changes in the length of repeats can have specific functional consequences without globally affecting protein function. Thus, one of the primary postulates of the *cis*-regulatory paradigm, namely that proteins are strongly constrained in their potential to contribute to morphological change because of negative pleiotropy, is not supported by data from the evolution of SLiMs and SSRs.

Conclusions and future directions

In this article, we have focused on the role of transcription factor proteins in the evolution of gene regulation as relevant to developmental evolution. This focus on transcription factors is not meant to advocate a narrow focus on transcription factors but was dictated by the limits of what can be covered in this short article. There are many other levels at which gene expression is regulated and which can contribute to developmental evolution (see Box 1). It is thus unlikely that any particular mechanism has exclusive claim on the evolution of gene regulation [36]. Rather, it is important to determine what kinds of phenotypic changes are caused preferentially by one rather than another mechanism. We conclude in agreement with Ref. [7] that asking specific questions about mechanisms is more productive than asking statistical questions, namely whether one kind of change is more frequent than another. For instance, it might be that transcription factor proteins are more likely to evolve if they acquire new target genes rather than when the level of regulation of an existing target gene is modified by natural selection. Acquiring new target genes might be rare in evolution, but it is nevertheless an important evolutionary event.

The main challenge to the field is that there are two approaches to the study of transcription factor evolution that rarely connect: the study of transcription factor sequence evolution and the experimental study of transcription factor function. Although the study of transcription factor sequence evolution has provided ample evidence that transcription factor proteins are subject to adaptive evolution, that is, amino acid sequence changes driven by directional natural selection, the functional importance of these changes is rarely investigated. By contrast, the experimental study of transcription factor function has provided strong evidence for functional nonequivalence of homologous transcription factors, but it is not clear what the driving evolutionary forces were that caused these differences. This state of affairs leaves many of the most important questions unanswered. For example, what is the biological and evolutionary context in which transcription factors change their functional specificities? It is clear that transcription factors do not always change, so under what circumstances do transcription factors become targets of natural selection and change? We suggest that such a

Box 1. Alternate levels of gene expression regulation

Gene regulation occurs at many levels. In this review, we have focused on the regulation of gene transcription (transcriptional regulation), but in principle evolutionary changes in gene expression can (and likely do) occur at many other levels. These non-transcriptional mechanisms of gene regulation have been called 'alternative regulatory levels' [36].

Regulatory non-coding RNA

Several classes of regulatory non-coding RNAs have been identified, including microRNAs (miRNA) and small interfering RNAs (siRNA). microRNAs are a large family of small, ~22 nucleotide long, non-coding RNAs that have independently evolved in metazoans and plants as regulators of gene expression. In mammals, miRNAs might regulate the expression of ~30% of protein-coding genes and participate in many cellular processes, including direct regulation of gene transcription through binding enhancers and promoters and posttranslational repression through mRNA targeting and degradation [56–58].

Regulatory elements in untranslated regions

In addition to regulating translation, *cis*-regulatory elements in 5′- and 3′-untranslated regions (UTRs) can modulate mRNA stability and translational efficiency [36]. For example, the orthodenticle-related 2 gene (*Otx2*), which encodes a homeodomain-containing transcription factor that controls brain morphogenesis, contains a vertebrate-specific 140 bp *cis*-regulatory element in its 3′ UTR that is essential for normal polyribosome complex formation and translation [36]. Mutation of this conserved element results in severe defects in head development in transgenic mice.

Alternative splicing

Alternative exon splicing is increasingly recognized as a widespread mechanism that enables multiple structurally and functionally distinct proteins to be generated from a single transcript to regulate gene expression [59]. Alternative splicing is a potentially powerful mechanism to increase protein diversity and appears to be common. For example, nearly 50% of mouse genes are alternatively spliced, whereas 18% and 14% utilize either alternative start or stop sites, respectively [60], and at least 50% and potentially as much as 80% of human genes are alternatively spliced [61,62]. Mutations affecting alternative splicing patterns can lead to novel isoforms of many proteins in addition to transcription factors, and thus novel and species-specific functions can evolve in enzymes, cell signaling molecules and receptors, contributing to phenotypic evolution.

Epigenetic gene regulation

Gene expression can also be regulated by the state of chromatin, such as transcriptionally active euchromatin and silent heterochromatin. Gene silencing in heterochromatin (and in imprinted regions as well) is associated with hypermethylation of DNA and covalent modifications of histones which prevent access to enhancers and promoters by the transcriptional machinery [63]. Several lines of evidence suggest that maintenance of methylation patterns at CG sequences is responsible for the formation of heritable activity states, termed epialleles [63]. Changes in the sequence of DNA in these regions between species might lead to differential chromatin states between species, resulting in different patterns of gene expression.

situation might occur whenever an evolutionary change calls for the origin of novel protein–protein interactions to regulate a novel target gene. By contrast, the mechanistic consequences of adaptive amino acid changes in transcription factors deserve more study. Do the changes directly affect specific transcription factor functions, such as DNA binding, protein–protein interactions or protein–RNA interactions, or do they reflect unspecific adaptive trends such as maintenance of protein stability and folding kinetics in changing environments? Perhaps the transcription factors that cooperate in a gene regulatory network are coadapted [54], which might explain the evolutionary stability of some gene regulatory networks.

We envision that a research program addressing these questions could start with identifying episodes of adaptive evolution in transcription factor proteins, and then proceed to investigating the functional consequences of these changes. This might lead to a fusion of comparative and experimental approaches as envisioned by Dean and Thornton as the 'functional synthesis' in evolutionary biology [55]. Work in this area is challenging, but not much more so than the exemplary work on the evolution of *cis*-regulatory elements [4,7,12].

Acknowledgements

Research in the G.P.W. laboratory is funded by grants from the National Science Foundation, the Alexander von Humboldt Foundation and the John Templeton Foundation. The opinions expressed in this report are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

References

1 Nyhart, L. (1995) *Biology Takes Form: Animal Morphology and the German Universities, 1800–1900,* University of Chicago Press
2 King, M.C. and Wilson, A.C. (1975) Evolution at two levels in humans and chimpanzees. *Science* 188, 107–116
3 Wilkins, A.S. (2002) *The Evolution of Developmental Pathways,* Sinauer Associates
4 Carroll, S.B. *et al.* (2001) *From DNA to Diversity,* Blackwell Science
5 Raff, R. (1996) *The Shape of Life,* Chicago University Press
6 Prud'homme, B. *et al.* (2007) Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci. U. S. A.* 104, 8605–8612
7 Wray, G.A. (2007) The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* 8, 206–216
8 Wray, G.A. *et al.* (2003) The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20, 1377–1419
9 Hoekstra, H.E. and Coyne, J.A. (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution Int. J. Org. Evolution* 61, 995–1016
10 Kuhn, T.S. (1962) The Structure of Scientific Revolutions, *University of Chicago Press*
11 Laubichler, M.D. and Davidson, E.H. (2008) Boveri's long experiment: sea urchin merogenes and the establishment of the role of nuclear chromosomes in development. *Dev. Biol.* 314, 1–11
12 Davidson, E.H. (2006) *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution,* Academic Press
13 Carroll, S.B. (2005) Evolution at two levels: on genes and form. *PLoS Biol.* 3, e245
14 DeBeer, G.R. (1971) *Homology an Unsolved Problem,* Oxford University Press
15 Saether, O. (1979) Underlying synapomorphies and anagenetic analysis. *Zool. Scr.* 8, 305–312
16 Gompel, N. *et al.* (2005) Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* 433, 481–487
17 Jeong, S. *et al.* (2006) Regulation of body pigmentation by Abdominal-B protein and its gain and loss in *Drosophila* evolution. *Cell* 125, 1387–1399
18 Prud'homme, B. *et al.* (2006) Repeated morphological evolution through *cis*-regulatory changes in a pleiotropic gene. *Nature* 440, 1050–1053
19 Sharkey, M. *et al.* (1997) Hox genes in evolution: protein surfaces and paralog groups. *Trends Genet.* 13, 145–151
20 Lynch, V.J. *et al.* (2006) Adaptive evolution of Hox-gene homeodomains after cluster duplications. *BMC Evol. Biol.* 6, 86
21 Fares, M.A. *et al.* (2003) Selection on coding regions determined Hox7 genes evolution. *Mol. Biol. Evol.* 20, 2104–2112
22 Hughes, A.L. (1999) *Adaptive Evolution of Genes and Genomes,* Oxford University Press

23 Malaga-Trillo, A.M. (2001) Genome duplications and accelerated evolution of Hox genes and cluster architecture in teleost fishes. *Am. Nat.* 41, 676–686
24 Crow, K.D. *et al.* (2006) The "fish specific" Hox cluster duplication is coincidental with the origin of teleosts. *Mol. Biol. Evol.* 23, 121–136
25 Purugganan, M.D. *et al.* (2000) Variation and selection at the CAULIFLOWER floral homeotic gene accompanying the evolution of domesticated *Brassica oleracea*. *Genetics* 155, 855–862
26 Barrier, M. *et al.* (2001) Accelerated regulatory gene evolution in an adaptive radiation. *Proc. Natl. Acad. Sci. U. S. A.* 98, 10208–10213
27 Lynch, V. *et al.* (2004) Adaptive evolution of HoxA-11 and HoxA-13 at the origin of the uterus in mammals. *Proc. Biol. Sci.* 271, 2201–2207
28 McGinnis, N. *et al.* (1990) Human Hox-4.2 and Drosophila deformed encode similar regulatory specificities in *Drosophila* embryos and larvae. *Cell* 63, 969–976
29 Ranganayakulu, G. *et al.* (1998) Divergent roles for NK-2 class homeobox genes in cardionenesis in flies and mice. *Development* 125, 3037–3048
30 Grenier, J.K. and Carroll, S.B. (2000) Functional evolution of the Ultrabithorax protein. *Proc. Natl. Acad. Sci. U. S. A.* 97, 704–709
31 Galant, R. and Carroll, S.B. (2002) Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* 415, 910–913
32 Ronshaugen, M. *et al.* (2002) Hox protein mutation and macroevolution of the insect body plan. *Nature* 415, 914–917
33 Zhao, Y. and Potter, S.S. (2001) Functional specificity of the Hoxa 13 homeobox. *Development* 128, 3197–3207
34 Zhao, Y. and Potter, S.S. (2002) Functional comparison of the Hoxa 4, Hoxa 10, and Hoxa 11 homeoboxes. *Dev. Biol.* 244, 21–36
35 Stern, D.L. (2000) Evolutionary developmental biology and the problem of variation. *Evolution Int. J. Org. Evolution* 54, 1079–1091
36 Alonso, C.R. and Wilkins, A.S. (2005) The molecular elements that underlie developmental evolution. *Nat. Rev. Genet.* 6, 709–715
37 Lopez, A.J. (1995) Developmental role of transcription factor isoforms generated by alternative splicing. *Dev. Biol.* 172, 396–411
38 Nekrutenko, A. (2004) Identification of novel exons from rat–mouse comparisons. *J. Mol. Evol.* 59, 703–708
39 Wang, W. *et al.* (2005) Origin and evolution of new exons in rodents. *Genome Res.* 15, 1258–1264
40 Zhang, X.H.F. and Chasin, L.A. (2006) Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proc. Natl. Acad. Sci. U. S. A.* 103, 13427–13432
41 Neduva, V. and Russell, R.B. (2005) Linear motifs: evolutionary interaction switches. *FEBS Lett.* 579, 3342–3345
42 Löhr, U. *et al.* (2001) *Drosophila fushi tarazu*: a gene on the border of homeotic function. *Curr. Biol.* 11, 1403–1412
43 Löhr, U. and Pick, L. (2005) Cofactor-interaction motifs and the cooption of a homeotic Hox protein into the segmentation pathway of *Drosophila melanogaster*. *Curr. Biol.* 15, 643–649
44 Mar Albà, M. *et al.* (1999) Amino acid reiterations in yeast are overrepresented in particular classes of proteins and show evidence of a slippage-like mutational process. *J. Mol. Evol.* 49, 789–797
45 Karlin, S. and Burge, C. (1996) Trinucleotide repeats and long homopeptides in genes and proteins associated with nervous system disease and development. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1560–1565
46 Alba, M.M. and Guigo, R. (2004) Comparative analysis of amino acid repeats in rodents and humans. *Genome Res.* 14, 549–554
47 Young, E.T. *et al.* (2000) Trinucleotide repeats are clustered in regulatory genes in *Saccharomyces cerevisiae*. *Genetics* 154, 1053–1068
48 Fondon, J.W., III and Garner, H.R. (2004) Molecular origins of rapid and continuous morphological evolution. *Proc. Natl. Acad. Sci. U. S. A.* 101, 18058–18063
49 Boras, K. and Hamel, P.A. (2002) Alx4 binding to LEF-1 regulates N-CAM promoter activity. *J. Biol. Chem.* 277, 1120–1127
50 Anan, K. *et al.* (2007) Morphological change caused by loss of the taxon-specific polyalanine tract in Hoxd-13. *Mol. Biol. Evol.* 24, 281–287
51 Goodman, F.R. *et al.* (1997) Synpolydactyly phenotypes correlate with size of expansions in HOXD13 polyalanine tract. *Proc. Natl. Acad. Sci. U. S. A.* 94, 7458–7463
52 Kjaer, K.W. *et al.* (2002) HOXD13 polyalanine tract expansion in classical synpolydactyly type Vordingborg. *Am. J. Med. Genet.* 110, 116–121
53 Zhao, X. *et al.* (2007) Mutations in HOXD13 underlie syndactyly type V and a novel brachydactyly-syndactyly syndrome. *Am. J. Hum. Genet.* 80, 361–371
54 Fraser, H.B. (2006) Coevolution, modularity and human disease. *Curr. Opin. Genet. Dev.* 16, 637–644
55 Dean, A.M. and Thornton, J.W. (2007) Mechanistic approaches to the study of evolution: the functional synthesis. *Nat. Rev. Genet.* 8, 675–688
56 Chapman, E.J. and Carrington, J.C. (2007) Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.* 8, 884–896
57 Sempere, L.F. *et al.* (2006) The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J. Exp. Zoolog. B Mol. Dev. Evol.* 306, 575–588
58 Hertel, J. *et al.* (2006) The expansion of the metazoan microRNA repertoire. *BMC Genomics* 7, 25
59 Blencowe, B.J. (2006) Alternative splicing: new insights from global analyses. *Cell* 126, 37–47
60 Sharov, A.A. *et al.* (2005) Genome-wide assembly and analysis of alternative transcripts in mouse. *Genome Res.* 15, 748–754
61 Johnson, J.M. *et al.* (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302, 2141–2144
62 Xu, Q. *et al.* (2002) Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.* 30, 3754–3766
63 Vaillant, I. and Paszkowski, J. (2007) Role of histone and DNA methylation in gene regulation. *Curr. Opin. Plant Biol.* 10, 528–533