Int. J. Dev. Biol. 47: 675-684 (2003)

Transcriptional regulation and the evolution of development

GREGORY A. WRAY*

Department of Biology, Duke University, Durham, North Carolina, USA

ABSTRACT A growing body of evidence suggests that changes in transcriptional regulation form an important part of the genetic basis for the evolution of development. At a microevolutionary level, all the necessary conditions are present: populations harbor abundant genetic variation for differences in transcription profiles, a substantial fraction of these variants can influence organismal phenotype, and some variants have fitness consequences and are subject to natural selection. At a macroevolutionary level, the evidence is less direct but strongly suggestive: specific differences in anatomy and gene expression are often correlated, while comparisons of transcription profiles among distantly related taxa point to extensive evolutionary changes in regulatory gene networks. Understanding how transcriptional regulatory systems evolve, and what contributions these changes have made to the evolution of phenotype, represents a major challenge for Evo-Devo.

KEY WORDS: *evolution of development, Hox paradox, promoter, transcription*

Introduction

One of the central goals of evolutionary developmental biology is to understand how developmental mechanisms are modified by mutation and selection so as to change anatomy and other interesting aspects of organismal phenotype (Raff 1996; Carroll *et al.,* 2001; Wilkins 2002; West-Eberhard 2003). This is a daunting task. Genomes are enormous relative to the size of a typical mutation, most traits are polygenic, most mutations in loci relevant to a trait of interest are phenotypically neutral or unrelated to the trait of interest, most aspects of phenotype are influenced by the environment, and genetic changes accumulate over extended intervals of time. For these reasons, reconstructing the exact sequence of events that lead to a complex change in phenotype presents a significant challenge. Ideally, one would like to identify the full complement of relevant mutations, the changes in developmental mechanisms they produced, the evolutionary mechanisms that fixed these variants, and the order in which these events happened.

An increasingly powerful set of approaches can be applied to the first two goals, namely identifying the genetic and developmental bases for a change in phenotype. Quantitative genetics, genetic screens, microarray analyses, and other methods provide powerful approaches to identifying contributing loci; sequence comparisons can reveal candidate polymorphisms; and functional tests, such as RNAi and homologous recombination, allow the precise genetic basis to be distinguished from irrelevant differences. None of this is technically simple, but it is possible. Just as importantly, from an evolutionary perspective, these approaches are becoming feasible outside the handful of major model systems.

Cases where a detailed understanding exists of the genetic basis for interesting, real-world phenotypic differences are beginning to accumulate, and already some interesting trends are evident. One of the most striking is that a large fraction of cases involves changes in transcriptional regulation. In some instances, the genetic difference alters the sequence of a protein that regulates transcription (e.g., Ting *et al.,* 1998; Ronshaugen *et al.,* 2002; Enard *et al.,* 2002b), but in many cases the relevant differences lie outside coding sequences and instead influence the interaction of transcription factors with promoter regions (e.g., Crawford *et al.,* 1999; Wang *et al.,* 1999; Hamblin and Di Rienzo 2000). Evidence from a variety of sources indicates that modifications in transcriptional regulation comprise a qualitatively and quantitatively important part of the genetic basis for the evolution of diverse aspects of organismal phenotype (reviewed in Wray *et al.,* 2003).

This review considers how changes in transcriptional regulation arise, are affected by natural selection, and alter organismal phenotype. The focus is on populations and closely related species, where genetic differences and evolutionary mechanisms can be studied with greater precision, but the review ends by considering the long-term imprint that changes in transcriptional regulation have left on the evolution of developmental mechanisms.

Gene expression and developmental evolution

Developmental processes are regulated by extensive genetic interactions, many of which occur through protein:DNA and protein:protein interactions during transcriptional regulation. The evolution of transcriptional regulation is not nearly as well understood

***Address correspondence to:** Dr. Gregory A. Wray. Department of Biology, Box 90338, Duke University, Durham, NC 27708-0338, USA. Fax +1-919-660-7293. e-mail: gwray@duke.edu

0214-6282/2003/$25.00

Printed in Spain
www.ijdb.ehu.es

as the evolution of protein function (Li 1997; Wray *et al.,* 2003). Nonetheless, several lines of evidence indicate that changes in transcriptional regulation constitute an important component of the genetic basis for the evolution of development.

Mutations affecting developmental gene expression

Genetic screens for mutations that affect developmental processes often map to loci that encode transcription factors. The function of these proteins is to regulate transcription at many loci (Latchman 1998). Many hundreds of mutations in transcription factors are known that disrupt developmental processes, and involve virtually every family of transcription factor known (Cooper 1999; Davidson 2001; Wilkins 2002). The induced mutations that emerge from genetic screens are probably not generally representative of genetic variation segregating in natural populations, however, because their phenotypic consequences are typically dramatic and unlikely to prove successful in the wild. However, natural variation in the sequence of transcription factors can have phenotypic consequences and some polymorphisms are known that affect the expression of specific downstream target genes (e.g., Brickman *et al.,* 2001). Interspecific differences in the structure and function of transcription factors can also contribute to important phenotypic differences between species (e.g., Enard *et al.,* 2002b; Galant and Carroll 2002; Ronshaugen *et al.,* 2002).

Some mutants that emerge from genetic screens lie within *cis*-regulatory sequences rather than coding sequences, and disrupt or alter transcription of a nearby locus. Although the majority of well-studied induced mutations reside in coding rather than regulatory sequences, this may not represent their relative frequency or evolutionary importance. Developmental geneticists typically focus on mutations of large effect, while polymorphisms segregating in natural populations typically have more subtle phenotypic consequences. Many naturally occurring differences in *cis*-regulatory sequences are known that produce functionally important differences in organismal phenotype. Examples include polymorphisms or fixed differences that affect anatomy (Stern 1998; Wang *et al.,* 1999; Robin *et al.,* 2002), physiology (Segal *et al.,* 1999; Lerman *et al.,* 2003), behavior (Trefilov *et al.,* 2000; Fang *et al.,* 2002), host-pathogen interactions (Hamblin and Di Rienzo 2000; Bamshad *et al.,* 2002), and life history (Allendorf *et al.,* 1983; Streelman and Kocher 2002).

Interspecific differences in gene expression

Surveys of developmental gene expression among relatively close species often turn up differences in the timing, location, level, sex-specificity, inducibility, and other aspects of transcription (e.g., Paigen 1989; Wray and McClay 1989; Schiff *et al.,* 1992; Kissinger and Raff 1998; Kopp *et al.,* 2000; Brunetti *et al.,* 2001). Few studies have systematically surveyed gene expression within a clade in an attempt to estimate the frequency of changes in these components of transcription profiles among species. Many examples are known, however, suggesting that interspecific differences in developmental gene expression are not rare (Wilkins 2002; Wray *et al.,* 2003). These evolutionary differences are evident throughout development and affect the expression of genes encoding proteins of diverse functions.

Some interspecific differences in the expression of transcription factors indicate dramatic changes in the regulation of gene expression. For instance, well-characterized developmental roles for even-skipped, sex-lethal, and bicoid proteins in *Drosophila* are absent in some other insects, and apparently represent evolutionarily derived regulatory functions (Patel *et al.,* 1992; Meise *et al.,* 1998; Stauber *et al.,* 1999). Conversely, some transcription factors have lost ancient regulatory functions. For example, zen and bicoid proteins derive from *Hox* cluster genes but have lost a segmental patterning function (Falciani *et al.,* 1996; Stauber *et al.,* 1999). These gains and losses of major developmental roles for transcription factors must have required many evolutionary changes in target gene interactions.

Other interspecific differences in gene expression correlate with specific phenotypic differences. For instance, the spatial extent of *Hox* gene expression matches the anatomical differences among vertebrae in tetrapods (Burke *et al.,* 1995) and among segments in crustaceans (Averof and Patel 1997). More fine-scale differences in gene expression domains correlate with the distribution of bristles in flies (Stern 1998) and with pigment in butterfly wings (Brunetti *et al.,* 2001). Although these cases are correlations, not demonstrations of a causal relationship, they identify specific changes in gene expression that can be tested for a direct role in a particular phenotypic change. A growing number of such cases provide strong evidence that changes in gene expression have played an important role in anatomical diversification (Carroll *et al.,* 2001; Davidson 2001; Wilkins 2002).

Gains and losses of transcription factors

The array of genes encoding transcription factors within the genome has changed dramatically during the diversification animals. Prior to whole-genome sequencing projects, the best-known case involved the four *Hox* cluster in vertebrates, in contrast to the single cluster (or fragmented single cluster) present in other metazoan groups. Single genes within the *Hox* cluster have also been duplicated on several occasions (e.g., Falciani *et al.,* 1996) and lost on others (e.g., Aboobaker and Blaxter 2003). With information from several genome sequences now available, it is clear that this phenomenon is not restricted to homeobox genes. The size of all of the transcription factor gene families has diverged considerably during metazoan evolution, sometimes by several fold (Lander *et al.,* 2001).

Some differences in the size of transcription factor gene families may correlate with anatomical evolution. If more complex developmental gene networks are required in order to build more complex anatomies, gene duplications may be a necessary step in the evolution of complexity (Holland 1990; Gerhart and Kirschner 1997; Force *et al.,* 1999). The most famous case involves the genome duplications that apparently occurred during early chordate evolution. The resulting infusion of many redundant regulatory loci may have provided the raw material for the diversification of transcription factor expression or activity through various mechanisms (Force *et al.,* 1999; Wray *et al.,* 2003), allowing a greater degree of anatomical complexity to evolve (Holland 1990). Duplications of single loci encoding transcription factors have also resulted in diversified regulatory functions (e.g., Ferris and Whitt 1979; Li and Knoll 1994). It seems likely that duplication of regulatory genes has been an important component in the evolution of phenotypic diversity.

Studying the evolution of transcriptional regulation

If changes in transcriptional regulation underlie many evolutionary differences in phenotype, as argued above, we need to understand how such changes become established in natural populations. Several components of this process need to be considered: how to study the evolution of transcription, the nature of mutations that alter

transcription, how selection acts on this genetic variation, and how these transcriptional differences affect developmental processes. The remaining sections of this review consider these issues in turn.

Differences between coding and regulatory sequences

Most of what we know about the evolution of function within the genome comes from protein-coding sequences. The primary analytical framework for these studies is the genetic code (Li 1997). The genetic code provides a reliable and powerful guide to proximate, or biochemical, phenotype, as manifest in codon usage bias, the K_a:K_s ratio (amino acid replacement versus silent), synonymous:nonsynonymous substitution ratio, and the presence of stop codons and frameshifts. Many tests for natural selection rely explicitly on the consistent and easily interpreted relationship that the genetic code provides between DNA and amino acid sequences.

No comparable analytic framework exists for *cis*-regulatory sequences. This is a direct consequence of they way these sequences are organized and function, which is in general much less regular than coding sequences (reviewed in: Arnone and Davidson 1997, Carey and Smale 2000, and White 2001). The functional nucleotides in *cis*-regulatory regions are short clusters (typically 4-10 bp) that bind transcription factors. These binding sites are embedded at irregular intervals within sequences that play no role in regulating transcription. Binding sites occupy no consistent position relative to the coding sequences they regulate or, in most cases, to each other. Transcription factor – DNA interactions are dually degenerate: a particular motif can often bind more than one transcription factor, and a particular transcription factor can bind to more than one motif. Whether a transcription factor actually binds to a suitable motif is strongly context-dependent, and may change across the life cycle, among cell types, and under different environmental conditions. Indeed, a suitable motif may not function at all in transcriptional regulation, even if it lies near a gene.

Identifying transcription factor binding sites

Because the organization and function of transcriptional regulatory sequences are so different from protein-coding sequences, distinct approaches are needed for analyzing their evolution (Wray *et al.,* 2003). For a variety of reasons, both gathering and analyzing evolutionary data are more challenging for *cis*-regulatory than protein-coding sequences.

Perhaps the most severe practical difficulty is that transcription factor binding sites can't be reliably identified from sequence comparisons alone (Carey and Smale 2000). Although sequence scans can identify candidate binding sites, confirmation that a particular sequence motif actually functions in regulating transcription requires direct experimental tests (Carey and Smale 2000; Li and Johnston 2001). Because such tests are labor- and cost-intensive, because regulatory sequences can lie within or many kb 5' or 3' from the locus they influence, and because binding site function can be sensitive to environmental conditions or life history stage, it is very difficult to know in practice when all the binding sites that regulate transcription of a particular locus have been identified.

As a result, nucleotides flanking a locus can't be reliably distinguished as functional and non-functional sites. Tests for selection that rely on classifying nucleotides according to function, such as the and HKA and McDonald-Kreitman tests (Hudson *et al.,* 1987; McDonald and Kreitman 1991), will lose power when sites are misclassified and may fail to detect selection when it exists (Wray *et al.,* 2003). Tests that don't rely on classifying nucleotides by function, such as Tajima's D and Fu and Li's D (Tajima 1989; Fu and Li 1993), are more dependable for *cis*-regulatory regions.

Another important consequence of the way *cis*-regulatory sequences function is that the function of a binding site is strongly context-dependent (Fry and Farnham 1999; Carey and Smale 2000). A particular binding site may function only is some cell types, for instance, and may contribute to activating transcription under some circumstances and to repressing it under others. Understanding the role of a particular binding site in producing the overall transcription profile requires biochemical and experimental tests that are labor- and cost-intensive (Carey and Smale 2000). The proximate functional consequences of sequence differences in coding sequences, in contrast, is reliably indicated by the genetic code: the amino acid sequence of the protein product can be determined from comparison of sequences alone.

Population genetics of *cis*-regulatory sequences

Any evolutionary change in transcriptional regulation must begin as a genetic polymorphism within a population. Knowing something about the nature and level of genetic variation in *cis*-regulatory sequences is therefore essential to understanding the evolution of transcriptional regulation, and in turn, developmental gene networks (Stern 2000). In part because of the technical challenges described above, far less work has been done on the population genetics of *cis*-regulatory sequences than of coding sequences and introns. During the past few years, however, this situation has begun to change dramatically, and recent studies have begun to provide some fascinating insights into the source of genetic differences that influence transcription.

Levels of genetic variation influencing transcriptional regulation

Since genetic variation is the raw material upon which selection acts, it is important to know how much genetic variation influencing transcriptional regulation resides in natural populations. Several approaches can be used to address this issue, including assays of mRNA or protein abundance and RT-PCR surveys (reviewed in Rockman 2003). Extensive surveys have been carried out in phylogenetically diverse organisms, including fungi (Cavalieri *et al.,* 2000; Brem *et al.,* 2002), plants (Burstin *et al.,* 1994; Damerval *et al.,* 1994; Costa and Plomion 1999; Gerber *et al.,* 2000; de Vienne *et al.,* 2001; Schadt *et al.,* 2003), and animals (Jin *et al.,* 2001; Cowles *et al.,* 2002; Enard *et al.,* 2002a; Oleksiak *et al.,* 2002; Rockman and Wray 2002; Rifkin *et al.,* 2003; Schadt *et al.,* 2003). All of these studies point to the same basic conclusion: natural populations harbor extensive variation in gene expression that is genetically based.

An interesting implication of this general result is that selection may play a more important role than mutation in limiting the appearance of evolutionary changes in transcription, at least over the long term. The fraction of heritable variation in gene expression that affects organismal phenotype is not yet known for any organism, much less the fraction that affects fitness, although many examples of each are now known (Wray *et al.,* 2003). The population genetics of transcriptional regulation is an area about which we know remarkably little relative to its likely importance.

The genetic basis for variation in gene expression

Another important issue is whether changes in transcription require rare mutations, such as large insertions, transposition, or recombination, or whether they can arise from more common mutations, such as nucleotide substitutions, small indels, or simple sequence repeat variants. Addressing this issue requires knowing the precise genetic basis for a difference in transcription for many loci.

Currently, the only organism for which the relevant information exists is *Homo sapiens*, where well over one hundred segregating *cis*-regulatory polymorphisms have been identified (Cooper 1999; Rockman and Wray 2002). All of these cases result from common, small-scale mutations. Of these, nucleotide substitutions and indels are represented roughly in proportion to the genome as a whole, while microsatellite-based variation is slightly over-represented. Although some lethal or strongly deleterious *cis*-regulatory alleles in humans are the result of large-scale mutations, alleles segregating in populations are overwhelmingly due to ordinary mutations (Rockman and Wray 2002). These mutations alter several different aspects of transcription, including level, location (cell type), and context-dependency (response to hormonal inputs or environmental stimuli).

Thus, the most common kinds of naturally occurring mutations are capable of altering transcription to the point of producing a phenotypic impact at the whole-organism level, at least in humans. In only a few cases has the precise genetic basis for a phenotypic difference due to altered transcription been identified in other organisms (e.g., Trefilov *et al.,* 2000; Robin *et al.,* 2002; Streelman and Kocher 2002; Daborn *et al.,* 2002; Lerman *et al.,* 2003). In all of these cases, the genetic basis for the difference is due either to a small-scale mutation or to insertion of a common transposon. Changes in transcriptional regulation do not require unusual kinds of mutations, suggesting that selection, rather than mutation, is will often be the limiting factor in the evolution of gene networks.

The location of polymorphisms contributing to an expression difference

A change in the expression of a particular gene could come about in many different ways. It might reside in *cis*, within the regulatory sequences flanking the gene, or in *trans*, in any of several loci encoding a transcription factors that interact with those sequences. If the change lies *cis*, gain of affinity for an activator protein or loss of affinity for a repressor (among other possibilities) could produce similar results. If the change lies *trans*, it could affect transcription of the locus, the DNA binding affinity of the protein, or post-translational modifications that alter its activity, again producing similar results on transcription of the focal downstream gene.

Even basic information about the location of genetic variation that affects gene expression would be useful for understanding the dynamics of gene network evolution. Of particular interest is whether most differences in gene expression trace back to a few, highly pleiotropic mutations in transcription factors, or whether most of the functionally relevant genetic variation is *cis* with the loci that show differences in expression. These two extreme models have quite different implications for understanding how selection would operate on alleles that influence transcription. If most polymorphisms affecting transcription only alter expression at one locus, pleiotropy should be more confined on average and variants that improve function might become fixed because they impose few other costs. Conversely, if most polymorphisms affecting transcription alter expression at many loci, antagonistic pleiotropy is more likely and few would escape having a net negative fitness consequence even if some changes in expression were beneficial.

Unfortunately, no good estimates exist for fraction of mutations affecting transcription that lie *cis* and *trans*. Microarrays, which provide an excellent tool for revealing expression differences within and between species (e.g., Jin *et al.,* 2001; Enard *et al.,* 2002a; Oleksiak *et al.,* 2002; Rifkin *et al.,* 2003; Schadt *et al.,* 2003), have serious limitations for population genetic analyses. Except under unusual circumstances, it is not possible to determine genotypes from microarrays because of the extensive gene interactions involved in regulating transcription. An allele at one locus may affect transcription at hundreds of other loci or at none, while a difference in transcription at a particular locus may or may not be sensitive to genetic background.

As a result, other information is needed to estimate the fraction of genetic variation that lies *cis* and *trans*. By adding genotypes of parents and offspring to microarray data, it is possible to estimate a lower bound for the cis-based contribution to expression differences. Surveys of from maize, human, and mouse suggest that at least one-third to one-half of the genetic basis for intraspecific quantitative differences in transcription lie *cis* to the locus in questions (Schadt *et al.,* 2003), although a comparable survey in yeast yielded a somewhat lower estimate (Brem *et al.,* 2002).

Natural selection on *cis*-regulatory sequences

The fate of a mutation in a *cis*-regulatory region will depend on its functional impact, the genetic background, the environment, and chance. A substantial fraction of mutations in *cis*-regulatory regions probably do not alter transcription at all, because they fall outside binding sites for transcription factors, or because they don't alter protein binding kinetics, or because they have an impact on transcription but not on fitness. Such mutations will generally be invisible to selection and will accumulate by drift or hitch-hiking. Some promoter mutations, however, clearly do have an impact on organismal phenotype, and some of these are under various forms of selection.

Negative selection

Conservation of promoter sequences to a greater degree than the neutral expectation provides evidence that negative selection is operating. Variation within the *Endo16* locus of the sea urchin *Strongylocentrotus purpuratus* is lower in the promoter than in an intron that apparently lacks regulatory sequences, while nucleotides within binding sites are less polymorphic than other nucleotides in the promoter (Balhoff and Wray, unpublished). Many mutations in *cis*-regulatory regions that compromise transcription and are therefore likely to be under negative selection have been documented in humans (Cooper 1999). A form of weak negative selection operates to remove spurious transcriptional start sites throughout prokaryotic genomes (Hahn *et al.,* 2003). Some interspecific sequence comparisons also suggest conservation of promoter sequences by negative selection (Aparicio *et al.,* 1995; Tümpel *et al.,* 2002). Large-scale sequence comparisons between moderately diverged genomes suggest that approximately equal fractions of coding and non-coding nucleotides are constrained by negative selection (Shabalina and Kondrashov 1999; Onyango *et*

al., 2000; Bergman and Kreitman 2001; Frazer *et al.,* 2001; Shabalina *et al.,* 2001). Although the nature of these functional non-coding sequences is generally not known, the majority are likely to be involved in transcriptional regulation.

Positive selection

If a new allele confers a fitness advantage, it may, depending on demography and chance, sweep through the population and replace the ancestral allele. A promoter allele of *Cyp6G1* in *D. melanogaster* that confers resistance to some pesticides is currently increasing in frequency in populations around the world (Daborn *et al.,* 2002). Some promoter alleles in *hsp70* in *D. melanogaster* reduce transcription in response to thermal stress and are probably under local directional selection (Lerman *et al.,* 2003). Several examples of positive selection on promoters in pathogens are also known (e.g., Buckwold *et al.,* 1997; Montano *et al.,* 2000). Conversely, some variants in human promoters confer resistance to infection by particular pathogens (e.g., Hamblin and DiRienzo 2000; Thursz 2001). Although these examples all concern adult physiology, there is no reason in principle why positive selection could not act on transcriptional regulation in embryos as well.

Balancing selection

Several *cis*-regulatory polymorphisms appear to be under balancing selection, meaning that no single allele confers the highest fitness under all circumstances. An example is *LDH* in the fish *Fundulus heteroclitus*, where two promoter alleles appear to be maintained by differences in water temperature (Crawford *et al.,* 1999; Segal *et al.,* 1999). Other likely cases of balancing selection on promoter alleles include *CCR5* and *IL4* in humans (Bamshad *et al.,* 2002; Rockman *et al.,* 2003) and *Endo16* in sea urchins (Balhoff and Wray, unpublished). Balancing selection is in principle more likely for gene products that carry out multiple functions, a common situation for developmental regulatory genes.

Stabilizing selection

Several cases have been documented where promoter sequences at orthologous loci are divergent, yet direct very similar transcription profiles (e.g., Wu and Brennan 1993; Tamarina *et al.,* 1997; Piano *et al.,* 1999; Ludwig *et al.,* 2000). In the more subtle cases, one or a few binding sites required for correct transcription in one species are absent in another. A more dramatic case is *Endo16* in camarodont sea urchins: the promoter is highly similar over a short (~180 bp) region but unalignable over the remaining >2 kb of functional sequence – yet drives nearly identical transcription profiles (Romano and Wray 2003). A plausible interpretation of such cases is that stabilizing selection is operating to maintain a consistent transcription profile, but that binding site turnover results in different arrangements of functional sequences and different protein:DNA interactions (Ludwig *et al.,* 2000; Wray *et al.,* 2003). In some situations, it may not matter precisely how transcription is regulated, so long as mRNA is produced at approximately the right time and place during development and in sufficient quantities. Given that a cell contains dozens of different transcription factors at any given time that are capable of regulating transcription and given the binding sites they interact with can evolve over relatively short evolutionary timescales (Stone and

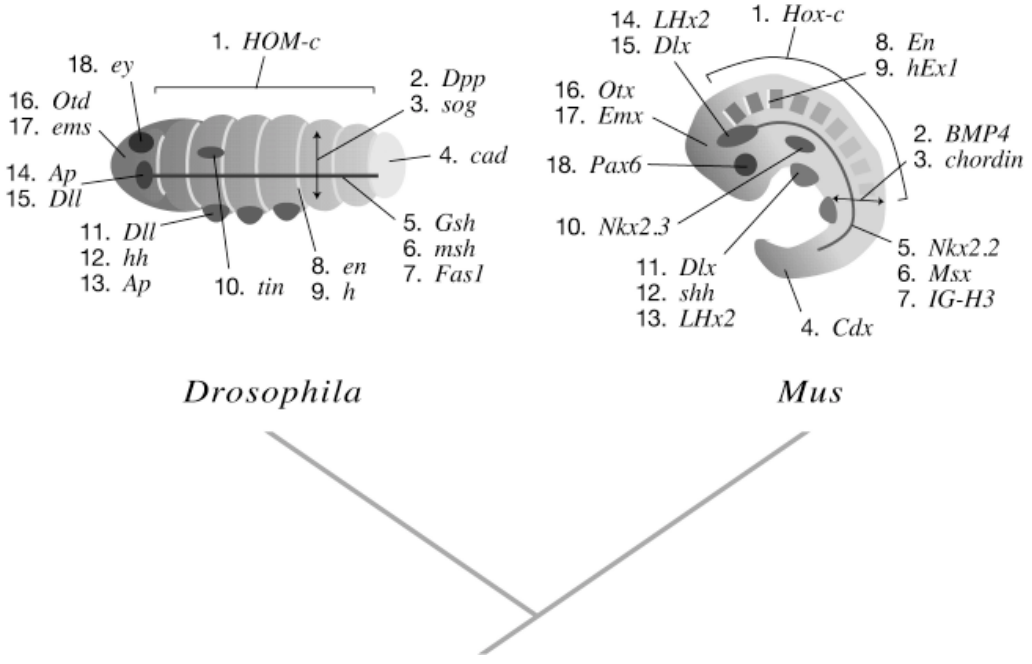

Fig. 1. The *Hox* Paradox. *Many gene encoding key developmental regulatory proteins are expressed in superficially similar domains during embryonic development in* Drosophila melanogaster *and* Mus musculus *(for reviews, see: Gerhart and Kirschner 1997; Carroll* et al., *2001). Some of the most widely discussed cases are shown here. Orthologous genes are numbered; if a gene is expressed in two different locations that are similar, it is numbered twice. Similar sites of expression are indicated by lines leading to structures. Despite the many similarities in gene expression domains, these embryos give rise to adults that are anatomically very different and contain few structures for which there is clear evidence of common ancestry. This disconnection between similar gene expression and dissimilar anatomy implies extensive changes in the organization of gene networks (see text).*

Wray 2001), there are probably many ways to achieve a particular change in transcription. Studies estimating rates of binding site turn-over (Dermitzakis and Clark 2002; Dermitzakis *et al.,* 2003; Costas *et al.,* 2003) have found that individual binding sites are gained and lost over $10^6 – 10^7$ year timescales despite conservation of transcription profiles, making this a highly dynamic process.

The *Hox* paradox and the evolution of gene networks

The evidence reviewed above suggests that genetic variation influencing gene expression is abundant within populations and that the processes of natural selection operate efficiently to sort this variation. But what about the longer term? The final, and most difficult, issue is whether changes in transcriptional regulation have played an important role in large-scale phenotypic changes, such as the origin and diversification of body plans. The available evidence is largely indirect, but comes from a growing number of cases.

The Hox *Paradox*

The broad phylogenetic distribution of key developmental regulatory genes among animal phyla, so much an accepted fact today, was initially quite surprising. Not only are the body plans of animals in different phyla starkly distinct (Raff 1986; Nielsen 2002), but so too are the cell biological contexts within which embryonic patterning occurs in different phyla (Davidson 1991; Gerhart and Kirschner 1997). Yet today we know that most families of developmental regulatory genes are widely distributed throughout the animal kingdom, and that a substantial fraction of these genes are expressed in superficially similar domains within embryos of distantly related phyla (Gerhart and Kirschner 1997; Carroll *et al.,* 2001).

All of this is now so familiar that it is easy to overlook a basic question: How is it possible for the same genes, expressed in similar spatial domains, to produce such different kinds of animals? Spatial similarities in gene expression between *Drosophila* and mouse embryos include dozens of genes involved in patterning the brain and limbs, polarizing the anteroposterior and dorsoventral axes, specifying cell fate in sense organs and the heart, and establishing segmental organization (Fig. 1) (for reviews, see: DeRobertis and Sasai 1996; Carroll *et al.,* 2001; Davidson 2001). Yet an adult fly and mouse are so anatomically distinct that very few unambiguously homologous structures can be identified between them (Raff 1996; Nielsen 2002; Wilkins 2002).

Thus, strikingly similar gene expression in embryos produces strikingly dissimilar adults. This broad-scale evolutionary dissociation between cause and effect is the *Hox* Paradox (Wray 2002), so named because the *Hox* cluster has been emblematic of the phenomenon of conserved developmental genes and expression domains.

Resolving the Hox *Paradox*

Paradoxes aren't real, of course, but an apparent paradox is a clear sign of faulty logic, inaccurate information, or missing information. Three explanations have been advanced to resolve the *Hox* Paradox, corresponding to each of these possibilities. These three explanations are not incompatible and all are probably true to some extent; the real challenge is deciding what proportion of the similarities in gene expression among phyla each explanation accounts for.

The first approach is to deny that adult anatomy is really so different after all (DeRobertis and Sasai 1996; Panganiban *et al.,* 1997; Holland and Holland 1999). This approach interprets topologically similar expression of homologous genes to be a highly reliable indicator of anatomical homology. The underlying assumption is that gene expression is inherently more evolutionarily conservative than anatomy, and therefore provides a more reliable index of common evolutionary origins. This explanation has been criticized on several grounds (Dickinson 1988; Abouheif *et al.,* 1997; Duboule and Wilkins 1998; Wray and Lowe 2000; Erwin and Davidson 2002; Wilkins 2002). Not only are numerous exceptions now known to the underlying assumption of conservatism in gene expression, but there are also clear phylogenetic discontinuities in the presence of anatomical structures that are homologous by this criterion (Wray and Lowe 2000; Erwin and Davidson 2002; Wilkins 2002). Although it may apply in some cases, this explanation does not provide a satisfactory general resolution to the *Hox* Paradox.

The second approach is to discount the extent of the paradox. According to this view, similarities in embryonic gene expression have been overstated: most cases are either vague ("anterior" is not a very convincing similarity on its own), coincidental (nearly every transcription factor is expressed somewhere within the central nervous system), or present for reasons that have nothing to do with a common evolutionary origin (serially homologous structures or presence of the same cell type might correlate with similar expression domains in clearly non-homologous structures). A sociological bias may also contribute to an overemphasis on similarities in gene expression: gene expression profiles that are dissimilar are difficult to interpret and harder to publish in prominent journals. These reasons may explain part of the *Hox* Paradox, but not all of it. Some of the resemblances are detailed, specific, and compatible with phylogenetic continuity; these similarities still require an explanation.

The third approach to resolving the *Hox* Paradox posits widespread evolutionary changes in developmental gene networks. According to this view, similarities in embryonic gene expression in anatomically distinct organisms could arise in two ways. One possibility is that the expression domains really are homologous, but that the downstream targets of these genes have changed extensively, resulting in very different anatomical outcomes. The other possibility is that similar expression domains are not homologous but instead the result of evolutionary convergence. This would require changes in upstream regulators to produce topologically similar expression domains, and could also involve differences in downstream targets. Under either scenario, an orthologous gene expressed in a similar domain could regulate highly divergent developmental processes that produce distinct anatomies and even structures that are not homologous (Wray and Lowe 2000; Davidson 2001).

Testing hypotheses for similar gene expression

Fortunately, it is often possible to distinguish between alternative explanations for similarity in gene expression by gathering data from additional loci and taxa (Abouheif *et al.,* 1997; Wray and Lowe 2000; Wilkins 2002). One test is that putatively homologous gene expression domains should be present in taxa representing intermediate phylogenetic positions (Fig. 2, top). The rationale is that if similar expression in two extant species is indeed due to homology, this (by definition) means that it was also present in their latest common ancestor. Similar expression might subse-

quently be lost in some descendants, but should be present in most surviving lineages. This prediction is testable by examining living representatives of appropriate groups. A second test is that the expression domains of functionally related genes should be similar to the one hypothesized to have conserved expression (Fig. 2, bottom). The rationale is that gene products don't act in isolation but rather as components of interacting complexes, where the expression of a single component would not be functional; as such, all or most interactors in one species should be co-expressed in the other. This prediction can be tested by examining multiple genes that contribute to the same developmental process. Neither of these tests is infallible, but they provide a much more reliable guide to distinguishing homology from convergence than the more usual fly-mouse comparison.

The expression of transcription factors responsible for patterning the nervous system illustrates the power of increased sampling of both taxa and loci. Several authors have interpreted similarities in the expression of these genes in *Drosophila* and mouse as evidence for a conserved well-organized central nervous system (DeRobertis and Sasai 1996; Holland and Holland 1999) and even specific brain regions (Hirth *et al.,* 2003) in the latest common bilaterian ancestor. Lowe *et al.,* (2003) tested this hypothesis by examining the expression of 22 neural patterning genes in the hemichordate *Saccoglossus kowalevski*, which occupies a phylogenetic position intermediate between arthropods and chordates (Brusca and Brusca 2002; Nielsen 2002). Their results provide strong support for conservation of gene expression domains, but also demonstrate that these expression domains are not functionally tied to a conserved nervous system anatomy. The hemichordate nervous system is largely composed of a diffuse network ectodermal nerve net with short dorsal and ventral nerve cords but no obvious brain (Knight-Jones 1952). By sampling many genes from a taxon occupying a key phylogenetic position, this study provides convincing evidence for third possibility mentioned in the previous section. The dramatic decoupling of conserved regulatory gene expression domains despite enormous differences in anatomy could only happen if the set of downstream target genes were quite different in hemichordates and vertebrates. Whether the nerve-net organization seen in living hemichordates represents the ancestral condition for bilaterians or a derived condition that evolved from a strongly centralized nervous system can only answered by sampling more taxa.

Some other cases of similar regulatory gene expression in arthropods and vertebrates may be the result of convergence rather than conservation. For instance, the many similarities in gene expression during limb development between these groups (Panganiban *et al.,* 1997) are most reasonably interpreted as convergent rather than homologous, since representatives of groups lying between these taxa lack comparable expression patterns, and since the evolutionary lineages separating arthropods and chordates lacked limbs (Shubin *et al.,* 1997; Davidson 2001). The same argument holds for the few similarities in expression that have been found in the genes responsible for segmentation among arthropods, annelids, and chordates (Holland *et al.,* 1997; Stollewerk *et al.,* 2003). Despite these similarities, the majority of genes involved in segmental patterning in one group are clearly not involved in this process in the others (e.g., Davis and Patel 1999; Wilkins 2002). In addition, the comparative anatomy of both living and fossil taxa provides clear evidence of a phylogenetic discontinuity in segmentation among these phyla, (Brusca and Brusca 2002; Erwin and Davidson 2002). The limited similarities that do exist in gene expression during segmentation are likely coincidental or convergent.

In most cases where the data from multiple taxa are available, similarities in developmental gene expression between distantly related animals fall into one of these two broad categories. Either the expression domain really is conserved, but the anatomy is clearly not, or the expression domain is more likely convergently (or perhaps even coincidentally) similar. In either case, extensive changes in gene expression are highly likely (Wray and Lowe 2000; Davidson 2001; Wilkins 2002).

Rewiring developmental gene networks

Gathering direct evidence of changes in gene regulatory networks is quite difficult. Identifying some of the target genes

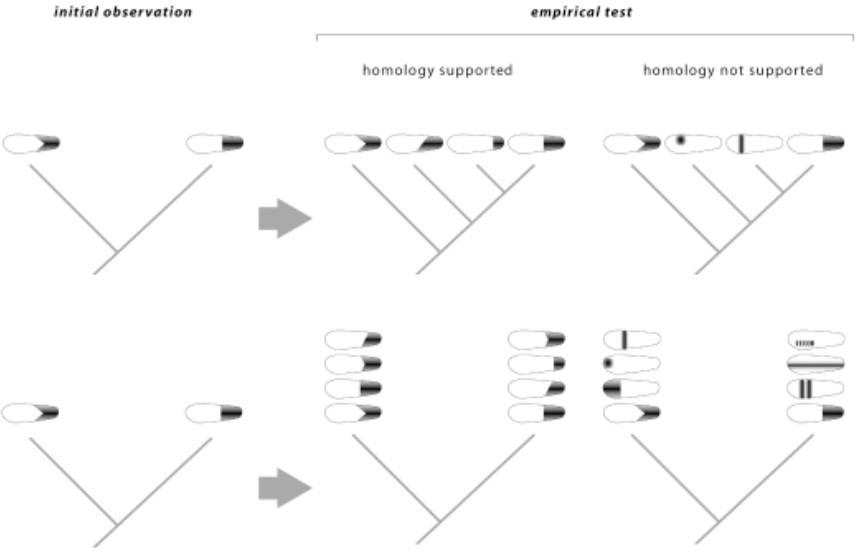

Fig. 2. Testing hypotheses of homology in gene expression domains. *Similar features of any kind in two different taxa could be the result of a common origin (homology) or independent origins (parallelism or convergence). When comparing gene expression among distantly related animals, as in Fig. 1, these possibilities are difficult to distinguish. Two simple tests can be used to discriminate among these alternative evolutionary histories. Sampling intermediate taxa (top row of cladograms) tests for phylogenetic continuity: homology predicts retention of similar expression domains in all or most representatives of intermediate taxa, while independent origins predict that the expression domain will be present in two separate clades and lacking in representatives of other taxa. Sampling additional genes whose products interact with the one initially observed to have similar expression domains tests for functional congruence: homology predicts that most interactors will be expressed in the same location while independent origins predicts that interactors differ between the two species. Neither test is infallible, but they provide considerable power when many taxa and genes are examined.*

regulated by a particular transcription factor is fairly straightforward in a few model organisms, but confidently identifying them all is currently not possible. Comparisons of downstream targets among taxa are therefore problematic, because false negatives are difficult to avoid. False positives, at least in an evolutionary sense, may also be difficult to detect. Since transcription factor binding sites can evolve very quickly (Hancock *et al.,* 1999; Stone and Wray 2001; Dermitzakis and Clark 2002; Dermitzakis *et al.,* 2003; Costas *et al.,* 2003), the fact that an orthologous gene is the downstream target of a particular transcription factor in two different taxa could represent either conservation or independent gains of the interaction. Until it becomes possible to catalogue the complete set of downstream target genes in multiple taxa with low false positive and negative error rates, direct evidence for extensive changes in downstream connections will be difficult to gather.

Fortunately, upstream connections may be more tractable. Part of the reason is that most eukaryotic transcription factors probably directly regulate hundreds or even thousands of downstream genes but are themselves directly regulated by only about 5-20 genes (Liang and Biggin 1998; Wray *et al.,* 2003). Furthermore, any gene in the genome could, in principle, belong to the set of direct downstream targets of a particular transcription factor, but less than a thousand (those encoding transcription factors and co-factors) could directly regulate its transcription.

In some cases, it is possible to rule out similar upstream connections in developmental gene regulatory networks. Embryonic transcription of the *Hox* complex of *Drosophila* is largely regulated by the products of the gap loci (Carroll *et al.,* 2001). In vertebrates, there is little evidence that the orthologues of the gap genes directly regulate the *Hox* complex; most of these genes are not even expressed at the right time and place (Davidson 2001; Wilkins 2002). Thus, the somewhat similar expression of the *Hox* complex in nested domains along the anteroposterior axis is apparently regulated by a largely non-overlapping set of upstream gene products in two different phyla.

Unfortunately, comparable data are simply not available for most developmental regulatory genes. As methods for reconstructing gene networks become more powerful, it will be quite interesting to learn how often the upstream regulators and downstream targets of key developmental loci change during the course of evolution.

Summary

Differential gene expression lies at the heart of development, and changes in the regulation of gene expression is a central component in the evolution of developmental mechanisms. Much less is currently known about the evolution of *cis*-regulatory sequences than of protein-coding sequences, in part because of technical problems and in part because of neglect. Nonetheless, the available evidence suggests that extensive genetic variation capable of altering transcription is present in natural populations, that selection operates efficiently on this variation, and that it has an impact on important aspects of organismal phenotype. Although direct evidence that changes in transcription have played a role in major anatomical differences among phyla remains elusive, a large body of indirect evidence suggests that developmental gene networks can differ significantly and may be uncoupled from particular anatomical features. The genetic basis for interesting evolutionary changes in developmental mechanisms may lie to a significant extent in transcriptional regulation.

References

ABOOBAKER, A.A. and BLAXTER, M.L. (2003) *Hox* gene loss during dynamic evolution of the nematode cluster. *Current Biol.* 13: 37-40.

ABOUHEIF, E.H., AKAM, M., DICKINSON, W.J., HOLLAND, P.W.H., MEYER, A., PATEL, N.H., RAFF, R.A., ROTH, V.L. and WRAY. G.A. (1997) Homology and developmental genes. *Trends Genet.* 13: 432-433.

ALLENDORF, F. W., KNUDSEN, K.L. and LEARY, R.F. (1983) Adaptive significance of differences in the tissue-specific expression of a phosphoglucomutase gene in rainbow trout. *Proc. Natl. Acad. Sci. USA* 80: 1397-1400.

APARICIO, S., MORRISON, A., GOULD, A., GILTHORPE, J., CHAUDHURI, C., RIGBY, P., KRUMLAUF, R. and BRENNER, S. (1995) Detecting conserved regulatory elements with the model genome of the Japanese pufferfish, *Fugu rubipes*. *Proc. Natl. Acad. Sci. USA* 92: 1684-1688.

ARNONE, M. I. and DAVIDSON, E. H. (1997) The hardwiring of development: Organization and function of genomic regulatory systems. *Development* 124: 1851-1864.

AVEROF, M. and PATEL, N. H. (1997) Crustacean appendage evolution associated with changes in *Hox* gene expression. *Nature* 388: 682-686.

BAMSHAD, M. J., MUMMIDI, S., GONZALEZ, E., AHUJA, S.S., DUNN, D.M., WATKINS, W.S., WOODING, S., STONE, A.C., JORDE, L.B., WEISS, R.B. and AHUJA, S.K. (2002) A strong signature of balancing selection in the 5' cis-regulatory region of *CCR5*. *Proc. Natl. Acad. Sci. USA* 99: 10539-10544.

BERGMAN, C. M. and KREITMAN, M. (200) Analysis of conserved noncoding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences. *Genome Res* 11: 1335-1345.

BRICKMAN, J. M., CLEMENTS, M., TYRELL, R., MCNAY, D., WOODS, K., WARNER, J., STEWART, A., BEDDINGTON, R. S. P. and DATTANI, M. (2001) Molecular effects of novel mutations in *Hesx1/HESX1* associated with human pituitary disorders. *Development* 128: 5189-5199.

BREM, R.B., YVERT, G., CLINTON, R. and KRUGLYAK, L. (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296: 752-755.

BRUNETTI, C.R., SELEGUE, J.E., MONTEIRO, A., FRENCH, V., BRAKEFIELD, P.M. and CARROLL, S.B. (2001) The generation and diversification of butterfly eyespot color patterns. *Current Biol.* 11: 1578-1585.

BRUSCA, R.C. AND BRUSCA. G.J. (2002) The Invertebrates, 2nd ed. Sinauer Associates, Sunderland MA.

BUCKWOLD, V.E., XU, Z.C., YEN, T.S.B. and OU, J.H. (1997) Effects of a frequent double-nucleotide basal core promoter mutation and its putative single-nucleotide precursor mutations on hepatitis B virus gene expression and replication. *J. Gen. Virol.* 78: 2055-2065.

BURKE, A.C., NELSON, C.E., MORGAN, B.A. and TABIN, C. (1995) *Hox* genes and the evolution of vertebrate axial morphology. *Development* 121: 333-346.

BURSTIN, J., DE VIENNE, D., DUBREUIL, P. and DAMERVAL, C. (1994) Molecular markers and protein quantities as genetic descriptors in maize. I. Genetic diversity among 21 inbred lines. *Theor. Appl. Genet.* 89: 943-950.

CAREY, M. and SMALE, S.T. (2000) Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

CARROLL, S.B., GRENIER, J.K., and WEATHERBEE, S.D. (2001) From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design. Blackwell Science, Inc., Malden, MA.

CAVALIERI, D., TOWNSEND, J.P. and HARTL, D.L. (2000) Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. *Proc. Natl. Acad. Sci. USA* 97: 12369-12374.

COOPER, D. N. (1999) Human Gene Evolution. Academic Press, Inc., San Diego.

COSTA, P. and PLOMION, D.L. (1999) Genetic analysis of needle proteins in maritime pine. 2. Variation in protein accumulation. *Silvae Genet.* 48: 146-150.

COSTAS, J., CESARES, F. and VIERA, J. (2003) Turnover of binding sites for transcription factors involved in early *Drosophila* development. *Gene* 310: 215-220.

COWLES, C.R., HIRSHHORN, J.N., ALTSHULER, D. and LANDER, E.S. (2002) Detection of regulatory variation in mouse genes. *Nat. Genet.* 32: 432-437.

CRAWFORD, D.L., SEGAL, J.A. and BARNETT, J.L. (1999) Evolutionary analysis of TATA-less proximal promoter function. *Mol. Biol. Evol.* 16: 194-207.

DABORN, P.J., YEN, J.L., BOGWITZ, M.R., GOFF, G.L., FEIL, E., JEFFERS, S., TIJET, N., PERRY, T., HECKEL, D., BATTERHAM, P., FEYEREISEN, R., WILSON, T.G. and FFRENCH-CONSTANT, R.H. (2002) A single P450 allele associated with insecticide resistance in *Drosophila*. Science 297: 2253-2225.

DAMERVAL, C., MAURICE, A., JOSSE, J.M. and DE VIENNE, D. (1994) Quantitative trait loci underlying gene product variation: A novel perspective for analyzing regulation of genome expression. Genetics 137: 289-301.

DAVIDSON, E.H. (1991) Spatial mechanisms of gene regulation in metazoan embryos. *Development* 113: 1-26.

DAVIDSON, E.H. (2001) Genomic Regulatory Systems: Development and Evolution. Academic Press, San Diego.

DAVIS, G.K. and PATEL, N.H. (1999) The origin and evolution of segmentation. *Trends Genetics* 15: M68-M72.

DE VIENNE, D., BOST, B., FIEVET, J., ZIVY, M. and DILLMANN, C. (2001) Genetic variability of proteome expression and metabolic control. *Plant. Physiol. Biochem.* 39: 271-283.

DERMITZAKIS, E.T. and CLARK, A.G. (2002) Evolution of transcription factor binding sites in mammalian gene regulatory regions: Conservation and turnover. *Mol. Biol. Evol.* 19: 1114-1121.

DERMITZAKIS, E.T., BERGMAN, C.M. and CLARK, A.G. (2003) Tracing the evolutionary history of *Drosophila* regulatory regions with models that identify transcription factor binding sites. *Mol. Biol. Evol.* 20: 703-714.

DEROBERTIS, E.M. and SASAI, Y. (1996) A common plan for dorsoventral patterning in Bilateria. *Nature* 380: 37-40.

DICKINSON, W. J. (1988) On the architecture of regulatory systems: Evolutionary insights and implications. *BioEssays* 8: 204-208.

DUBOULE, D. and WILKINS, A.S. (1998) The evolution of 'bricolage'. *Trends Genet.* 14: 54-59.

ENARD, W., KHAITOVICH, P., KLOSE, J., ZÖLLNER, S., HEISSIG, F., GIAVALISCO, P., NIESELT-STRUWE, K., MUCHMORE, E., VARKI, A., RAVID, R., DOXIADIS, G.M., BONTROP, R.E. and PÄÄBO, S. (2002a) Intra- and interspecific variation in primate gene expression patterns. *Science* 296: 340-343.

ENARD, W., PRZEWORSKI, M. , FISHER, S.E., LAI, C.S.L., WIEBE, V., KITANO, T., MONACO, A.P. and PÄÄBO, S. (2002b) Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* 418: 869-872.

ERWIN, D.H. and DAVDISON, E. H. (2002) The last common bilaterian ancestor. *Development* 129:3021-3032.

FALCIANI, F., HAUSDORF, B., SCHRÖDER, R., AKAM, M., TAUTZ, D., DENELL, R., and BROWN, S. (1996) Class 3 *Hox* genes in insects and the origin of *zen*. *Proc. Natl. Acad. Sci. USA* 93: 8479-8484.

FANG, S., TAKAHASHI, A. and WU, C.-I. (2002) A mutation in the promoter of *desaturase 2* is correlated with sexual isolation between *Drosophila* behavioral races. *Genetics* 162: 781–784.

FERRIS, S.D. and WHITT, G.S. (1979) Evolution of the differential regulation of duplicate genes following polyploidization. *J. Mol. Evol.* 12: 267-317.

FORCE, A., LYNCH, M., PICKETT, F.B., AMORES, A., YAN, Y.-L. and POSTLETHWAIT, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151: 1531-1545.

FRAZER, K.A., SHEEHAN, J.B., STOKOWSKI, R.P., CHEN, X., HOSSEINI, R., CHENG, J.F., FODOR, S.P., COX, D.R. and PATIL, N. (2001) Evolutionarily conserved sequences on human chromosome 21. *Genome Res.* 11: 1651-1659.

FRY, C.J. and FARNHAM, P.J. (1999) Context-dependent transcriptional regulation. *J. Biol. Chem.* 274: 29583-29586.

FU, Y.-X. and LI, W.-H. (1993) Statistical tests of neutrality of mutations. *Genetics* 133: 693-709.

GALANT, R. and CARROLL, S. B. (2002) Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* 415: 910-913.

GERBER, S., FABRE, F. and PLANCHON, C. (2000) Genetics of seed quality in soybean analysed by capillary gel electrophoresis. *Plant Sci.* 152: 181-189.

GERHART, J. and KIRSCHNER, M. (1997) Cells, Embryos, and Evolution. Blackwell Science, Inc., Malden, MA.

HAHN, M.W., STAJICH, J.E. and WRAY, G.A. (2003) The effects of selection against spurious transcription factor binding sites. *Mol. Biol. Evol.* 20: 901-906.

HAMBLIN, M.T. and DI RIENZO, A. (2000) Detection of the signature of natural selection in humans: Evidence from the Duffy blood group locus. *Amer. J. Hum. Gen.* 66: 1669-1679.

HANCOCK, J., SHAW, P., BENNETON, F. and DOVER, G. (1999) High sequence turnover in the regulatory regions of the developmental gene *hunchback* in insects. *Mol. Biol. Evol.* 16: 253-265.

HIRTH, R., KAMMERMEIER, L., FREI, E., WALLDORF, U., NOLL, M. and REICHERT, H. (2003) An urbilaterian origin of the tripartite brain: developmental genetic insights. *Development* 130: 2385-2373.

HOLLAND, L.Z., KENE, M., WILLIAMS, N.A. and HOLLAND, N.D. (1997) Sequence and embryonic expression of the amphioxus *engrailed* gene (*AmphiEn*): the metameric pattern of transcription resembles that of its segment-polarity homolog in *Drosophila*. *Development* 124: 1723-1732.

HOLLAND, N.D. and HOLLAND, L.Z. (1999) Amphioxus and the utility of molecular genetic data for hypothesizing body part homologies between distantly related animals. *Amer. Zool.* 39: 630-640.

HOLLAND, P.W.H. (1990) Homeobox genes and segmentation: co-option, co-evolution, and convergence. *Semin. Dev. Biol.* 1: 135-145.

HUDSON, R.R., KREITMAN, M. and AGUADE, M. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* 116: 153-159.

JIN, W., RILEY, R.M., WOLFINGER, R.D., WHITE, K.P., PASSADOR-GURGEL, G. and GIBSON, G. (2001) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat. Gen.* 29: 389-395.

KISSINGER, J.C. and RAFF, R.A. (1998) Evolutionary changes in sites and timing of actin gene expression in embryos of the direct- and indirect-developing sea urchins, *Heliocidaris erythrogramma* and *H. tuberculata*. *Dev. Genes. Evol.* 208: 82-93.

KNIGHT-JONES, E. (1952) On the nervous system of *Saccoglossus cambrensis* (Enteropneusta). *Philos. Trans. R. Soc. Lond.* B 236: 315-354.

KOPP, A., DUNCAN, I. and CARROLL, S.B. (2000) Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature* 408: 553-559.

LANDER, E.S., L.M. LINTON, B. BIRREN (and 239 co-authors). (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

LATCHMAN, D.S. (1998) Eukaryotic Transcription Factors. Academic Press, San Diego.

LERMAN, D.N., MICHALAK, P., HELIN, A.B., BETTENCOURT, R.B. and FEDER, M.E. (2003) Modification of heat-shock gene expression in *Drosophila melanogaster* populations via transposable elements. *Mol. Biol. Evol.* 20: 135-144.

LI, Q.M. and JOHNSTON, S.A. (2001) Are all DNA binding and transcription regulation by an activator physiologically relevant? *Mol. Cell. Biol.* 21: 2467-2474.

LI, W.-H. (1997) Molecular Evolution. Sinauer Associates, Sunderland, Mass.

LI, X. and NOLL, M. (1994) Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different *cis*-regulatory regions. *Nature* 367: 83-87.

LIANG, Z. and BIGGIN, M.D. (1998) Eve and ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in *Drosophila*. *Development* 125: 4471-4482.

LOWE, C.J., WU, M., SALIC, A., EVANS, L., LANDER, E., STANGE-THOMANN, N., GUBER, C.E., GERHART, J. and KIRSCHNER, M. (2003) Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell* 113: 853-865.

LUDWIG, M.Z., BERGMAN, C., PATEL, N.H. and KREITMAN, M. (2000) Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403: 564-567.

MCDONALD, J.H. and KREITMAN, M. (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351: 652-654.

MEISE M., HILFIKERKLEINER, D., DUBENDORFER, A., BRUNNER, C., NOTHIGER, R. and BOPP, D. (1998) *Sex-lethal*, the master sex-determining gene in *Drosophila*, is not sex-specifically regulated in *Musca domestica*. *Development* 125: 1487-1494.

MONTANO, M.A., NIXON, C.P., NDUNG'U, T., BUSSMANN, H., NOVITSKY, V.A., DICKMAN, D. and ESSEX, M. (2000) Elevated tumor necrosis factor-alpha activation of human immunodeficiency virus type 1 subtype C in Southern Africa is associated with an NF-kappaB enhancer gain-of-function. *J. Infect. Dis.* 181: 76-81.

NIELSEN, C. (2002) Animal Evolution: Interrelationships of the Animal Phyla. 2nd ed. Oxford University Press, Oxford.

OLEKSIAK, M.F., CHURCHILL, G.A. and CRAWFORD, D.L. (2002) Variation in gene expression within and among natural populations. *Nat. Genet.* 32: 261-266.

ONYANGO, P., MILLER, W., LEHOCZKY, J., LEUNG, C.T., BIRREN, B., WHEELAN, S., DEWAR, K. and FEINBERG, A.P. (2000) Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain. *Genome Res.* 10: 1697-1710.

PAIGEN, K. (1989) Experimental approaches to the study of regulatory evolution. *Am. Nat.* 134: 440-458.

PANGANIBAN, G., IRVINE, S.M., LOWE, C., ROEHL, H., CORLEY, L.S., SHERBON, B., GRENIER, J.K., FALLON, J.F., KIMBLE, J., WALKER, M., WRAY, G.A., SWALLA, B.J., MARTINDALE, M.Q. and CARROLL, S.B. (1997) The origin and evolution of animal appendages. *Proc. Natl. Acad. Sci. USA* 94: 5162-5166.

PATEL N.H., BALL, E.E. and GOODMAN, C.S. (1992) Changing role of *even-skipped* during the evolution of insect pattern formation. *Nature* 357: 339-342.

PIANO, F., PARISI, M.J., KARESS, R. and KAMBYSELLIS, M.P. (1999) Evidence for redundancy but not *trans* factor-*cis* element coevolution in the regulation of *Drosophila* Yp genes. *Genetics* 152: 605-616.

RAFF, R.A. (1996) *The Shape of Life: Genes, Development, and the Evolution of Animal Form.* The University of Chicago Press, Chicago.

RIFKIN, S.A., KIM, J. and WHITE. K.P. (2003) Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat Genet.* 33: 138-144.

ROBIN, C., LYMAN, R.F., LONG, A.D., LANGLEY, C.H. and MACKAY, T.F.C. (2002) *hairy*: a quantitative trait locus for *Drosophila* sensory bristle number. *Genetics* 162: 155-164.

ROCKMAN, M.V. (2003) Idiomatic (gene) expressions. *BioEssays* 25: 421-424.

ROCKMAN, M.V. and WRAY, G.A. (2002) Abundant raw material for *cis*-regulatory evolution in humans. *Mol. Biol. Evol.* 19: 1981-1990.

ROMANO, L.A. and WRAY, G.A. (2003) Conservation of *Endo16* expression in sea urchins despite divergence in both *cis* and *trans*-acting components of transcriptional regulation. *Development* 130: 4187-4199.

RONSHAUGEN, M., MCGINNIS, N. and MCGINNIS, W. (2002) Hox protein mutation and macroevolution of the insect body plan. *Nature* 415: 914-917.

SCHADT, E.E., MONKS, S.A., DRAKE, T.A., LUSIS, A.J., CHE, N., COLINAYO, V., RUFF, T.G., MILLIGAN, S.B., LAMB, J.R., CAVET, G., LINSLEY, P.S., MAO, M., STOUGHTON, R.B. and FRIEND, S.H. (2003) Genetics of gene expression surveyed in maize, mouse, and man. *Nature* 422: 297-302.

SCHIFF, N.M., FENG, Y., QUINE, J.A., KRASNEY, P.A. and CAVENER, D.R. (1992) Evolution of the expression of the *Gld* gene in the reproductive tract of *Drosophila*. *Mol. Biol. Evol.* 9: 1029-1049.

SEGAL, J.A., BARNETT, J.L. and CRAWFORD, D.L. (1999) Functional analysis of natural variation in Sp1 binding sites of a TATA-less promoter. *J. Mol. Evol.* 49: 736-749.

SHABALINA, S.A. and KONDRASHOV, A.S. (1999) Pattern of selective constraint in *C. elegans* and *C. briggsae* genomes. *Genet. Res.* 74: 23-30.

SHABALINA, S.A., OGURTSOV, A.Y., KONDRASHOV, V.A. and KONDRASHOV, A.S. (2001) Selective constraint in intergenic regions of human and mouse genomes. *Trends Genet.* 17: 373-376.

SHUBIN, N., TABIN, C. and CARROLL, S.B. (1997) Fossils, genes and the evolution of animal limbs. *Nature* 388: 639-648.

STAUBER, M., JACKLE, H. and SCHMIDT-OTT, U. (1999) The anterior determinant bicoid of *Drosophila* is a derived *Hox* class 3 gene. *Proc. Natl. Acad. Sci. USA* 96: 3786-3789.

STERN, D.L. (2000) Perspective: Evolutionary developmental biology and the problem of variation. *Evolution* 54: 1079-1091.

STERN, D.L. (1998) A role of *Ultrabithorax* in morphological difference between *Drosophila* species. *Nature* 396: 463-466.

STOLLEWERK, A., SCHOPPMEIER, M. and DAMEN, W.G.M. (2003) Involvement of *Notch* and *Delta* genes in spider segmentation. *Nature* 423: 863-865.

STONE, J.R. and WRAY, G.A. (2001) Rapid evolution of *cis*-regulatory sequences via local point mutations. *Mol. Biol. Evol.* 18: 1764-1770.

STREELMAN, J.T. and KOCHER, T.D. (2002) Microsatellite variation associated with prolactin expression and growth of salt-challenged tilapia. *Physiol. Genomics* 9: 1-4.

TAJIMA, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

TAMARINA, N.A., LUDWIG, M.Z. and RICHMOND, R.C. (1997) Divergent and conserved features in the spatial expression of the *Drosophila* pseudoobscura esterase-5B gene and the esterase-6 gene of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 94: 7735-7741.

THURSZ, M. (2001) Genetic susceptibility in chronic viral hepatitis. *Antiviral Res.* 52: 113-116.

TING, C.T., TSAUR, S.C., WU, M.L. and WU, C.I. (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282: 1501-1504.

TREFILOV, A., BERARD, J., KRAWCZAK, M. and SCHMIDTKE, J. (2000) Natal dispersal in rhesus macaques is related to serotonin transporter gene promoter variation. *Behav. Genet.* 30: 295-301.

TÜMPEL, S., MACONOCHIE, M., WIEDEMANN, L.M. and KRUMLAUF, R. (2002) Conservation and diversity in the cis-regulatory networks that integrate information controlling expression of *Hoxa2* in hindbrain and cranial neural crest cells in vertebrates. *Dev. Biol.* 246: 45-56.

WANG, R.L., STEC, A., HEY, J., LUKENS, L. and DOEBLEY, J. (1999) The limits of selection during maize domestication. *Nature* 398: 236-239.

WEST-EBERHARD, M.J. (2003) *Developmental Plasticity and Evolution.* Oxford Univ. Pres, Oxford.

WHITE, R.J. (2001) *Gene Transcription: Mechanisms and Control.* Blackwell Science, Malden, MA.

WILKINS, A.S. (2002) *The Evolution of Developmental Pathways.* Sinauer Associates, Sunderland, MA.

WRAY, G.A. (2002) Resolving the *Hox* Paradox. *Science* 292: 2256-2257.

WRAY, G.A. and LOWE, C.J. (2000) Developmental regulatory genes and echinoderm evolution. *Sys. Biol.* 49: 28-51.

WRAY, G.A. and MCCLAY, D.R. (1989) Molecular heterochronies and heterotopies in early echinoid development. *Evolution* 43: 803-813.

WRAY, G.A., HAHN, M.W., ABOUHEIF, E., BALHOFF, J.P., PIZER, M., ROCKMAN, M.V. and ROMANO, L.A. (2003) The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 29: 1377-1419.

WU, C.Y. and BRENNAN, M.D. (1993) Similar tissue-specific expression of the *Adh* genes from different *Drosophila* species is mediated by distinct arrangements of *cis*-acting sequences. *Mol. Gen. Genet.* 240: 58-64.