REGULATORY ELEMENTS

Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence

Patricia J. Wittkopp[1,2] and Gizem Kalay[2]

Abstract | *Cis*-regulatory sequences, such as enhancers and promoters, control development and physiology by regulating gene expression. Mutations that affect the function of these sequences contribute to phenotypic diversity within and between species. With many case studies implicating divergent *cis*-regulatory activity in phenotypic evolution, researchers have recently begun to elucidate the genetic and molecular mechanisms that are responsible for *cis*-regulatory divergence. Approaches include detailed functional analysis of individual *cis*-regulatory elements and comparing mechanisms of gene regulation among species using the latest genomic tools. Despite the limited number of mechanistic studies published to date, this work shows how *cis*-regulatory activity can diverge and how studies of *cis*-regulatory divergence can address long-standing questions about the genetic mechanisms of phenotypic evolution.

Cis-regulatory elements (CREs). Collections of transcription factor binding sites and other non-coding DNA that are sufficient to activate transcription in a defined spatial and/or temporal expression domain.

[1]*Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109, USA.*
[2]*Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, Michigan 48109, USA.*
Correspondence to P.J.W.
e-mail: wittkopp@umich.edu
doi:10.1038/nrg3095
Published online
6 December 2011

Over 40 years ago, mutations affecting the regulation of gene expression were predicted to be a common source of evolutionary change[1–3]. Since this time, and most rapidly during the past 5 to 10 years, empirical evidence has accumulated showing this prediction to be true. This evidence includes expression divergence that correlates with phenotypic divergence, manipulations of gene expression that are sufficient for recreating phenotypic differences and genetic mapping identifying regulatory loci as being responsible for divergent phenotypes[4,5]. In fact, data from these types of studies, combined with an improved understanding of the molecular mechanisms regulating gene expression, have prompted a refinement of the original hypothesis: mutations affecting the activity of *cis*-regulatory sequences — as opposed to *trans*-regulatory sequences encoding the transcription factors (TFs) that bind to *cis*-regulatory sequences — are now thought to be the most prevalent cause of phenotypic (especially morphological) divergence[4,5].

Cis-regulatory sequences can be discretized into *cis*-regulatory elements (CREs) that are composed of DNA (typically, non-coding DNA) containing binding sites for TFs and/or other regulatory molecules that are needed to activate and sustain transcription[6]. Promoters and enhancers are the best understood types of CREs[7,8]. Most eukaryotic genes contain a single promoter located close to the transcription start site, although some genes contain alternative promoters that activate transcription at different positions in the genome, often under specific conditions. Promoters are required for transcription in eukaryotes, but alone they only produce basal levels of mRNA. Because promoter sequences bind to a core set of widely used and highly conserved transcriptional regulators, they do not appear to be the primary driver of *cis*-regulatory divergence[9]. Promoter mutations are, however, a frequent cause of human disease[10].

Compared to promoters, enhancers tend to be more variable between species; they are the type of CRE that is most often thought to be responsible for *cis*-regulatory divergence[11], and thus they are the focus of this Review. In metazoans that have many different cell types, gene expression is often regulated by multiple enhancers, each of which controls expression in a limited range of cell or tissue types or during a particular stage in development. Conventionally, it has been thought that each enhancer controls a unique subset of a gene's expression, but pairs of enhancers that have largely overlapping function have recently been identified that contribute to phenotypic stability[12,13]. Functional independence among enhancers allows the effects of a mutation in one enhancer to have limited effects on aspects of the gene's expression that are controlled by other enhancers. Enhancers are typically located upstream (5′), downstream (3′) or in the intron (or introns) of the gene that they regulate, but

Functionally homologous enhancers
Enhancers from different species that drive analogous expression patterns in the same tissue. They may or may not be caused by orthologous DNA sequences.

Divergent sites
Fixed nucleotide differences in orthologous DNA sequences between species.

Pleiotropic
An adjective that is used to describe a mutation or gene that affects multiple (presumably distinct) phenotypes.

Trans-regulatory factors
Proteins, RNAs or other diffusible molecules that affect gene expression.

they can also be located far away[14]. The genomic location of functionally homologous enhancers is often[15,16], but not always[17], conserved between species.

The theory and evidence that support a primary role for *cis*-regulatory mutations in phenotypic evolution have been extensively reviewed in recent years[4,5,11] and are summarized in BOX 1. Here, we take for granted that understanding *cis*-regulatory evolution is essential for understanding phenotypic evolution, and instead we focus on recent investigations into the genetic and molecular changes that underlie *cis*-regulatory differences between species. After briefly reviewing the methods used to study *cis*-regulatory divergence (and highlighting recent advances in this area), we discuss the molecular mechanisms identified to date that are responsible for divergent *cis*-regulatory activity. This discussion includes the types of genetic changes that underlie functional divergence of enhancers, interactions among such divergent sites, the impact of these changes on TF binding and the origin of novel *cis*-regulatory activities. We then consider ways in which understanding the mechanisms of *cis*-regulatory evolution have begun to provide insight into general questions about the evolutionary process, including the relative roles of standing genetic variation and new mutations, the relative contribution of mutations of large and small effect, the genetic basis of convergent phenotypes and the role of selection in phenotypic divergence. Despite the intriguing insights into evolutionary mechanisms that are provided by currently available data, many more case studies should be acquired before quantitative statements are made about the relative frequency of different types of changes. We conclude this Review by identifying experimental approaches that are most likely to enrich our understanding of this fundamental evolutionary process and to provide such data.

Box 1 | **The importance of *cis*-regulatory evolution: evidence and theory**

For most of the twentieth century, evolution of protein-coding sequences was commonly thought to be primarily (if not solely) responsible for phenotypic evolution; it was generally assumed that species-specific traits resulted from species-specific proteins. Although changes in protein function are responsible for phenotypic divergence in some cases[83], many homologous proteins have highly conserved functions among species but display differences in their expression that contribute to phenotypic divergence[4,5,11]. The pervasiveness of regulatory evolution is often rationalized by the fact that changes in protein-coding regions are expected to be more pleiotropic (that is, they affect more phenotypes) than changes in modular, tissue-specific, *cis*-regulatory elements (CREs) (for example, enhancers), and hence they may be more likely to be deleterious. This is because changing the sequence and function of a protein should generally affect all cells in which the protein is active (but see REF. 84), whereas changing a tissue-specific CRE should only have an impact on cells that are affected by the specific expression change.

As described in the main text, gene expression is controlled by both *cis*- and *trans*-regulatory factors, and mutations in either type can alter expression. *Cis*-regulatory DNA sequences contain binding sites that interact with diffusible *trans*-regulatory proteins and RNAs. Between species, *cis*-regulatory divergence accounts for a greater proportion of expression differences than it does within species[20,21,85,86]. This finding is again rationalized by the anticipated lower pleiotropy of *cis*-regulatory sequences relative to *trans*-regulatory factors. This is because the modular structure of *cis*-regulatory sequences allows mutations that affect one module (usually an enhancer controlling expression at a particular time and/or in a particular tissue type) to have little or no impact on expression that is controlled by other modules.

Perhaps most importantly, there is a large and rapidly growing collection of case studies that implicate *cis*-regulatory divergence in phenotypic evolution[5,11]: for every study shown in Supplementary information S1 (table) for which the mutations that are responsible for *cis*-regulatory divergence have been identified, there are about ten more in which clear evidence for *cis*-regulatory divergence is presented[5]. For example, expanded expression of a paired-related homeobox gene (*Prx1*) in bats (*Carollia perspicillata*) that contributes to their elongated forelimbs was shown to be due to differences in a limb-expression enhancer relative to its mouse homologue[87], and differences in red wing colour patterns between *Heliconius* butterfly species were shown to be due to *cis*-regulatory changes causing differential expression of *optix*, a homeobox transcription factor gene[88].

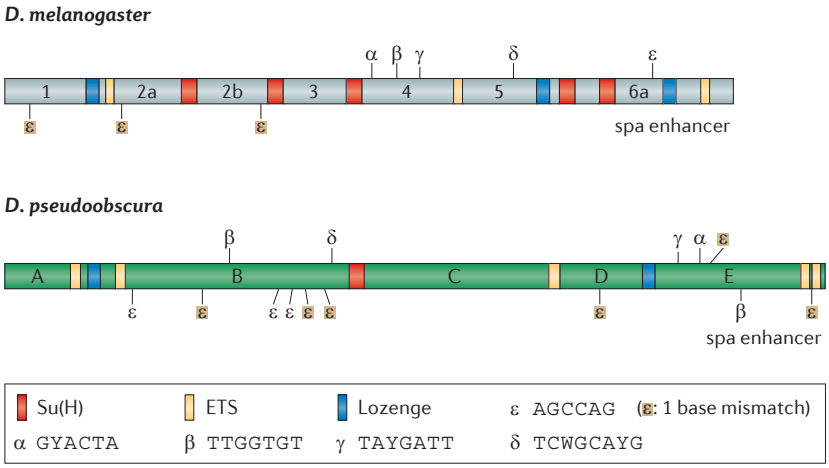
Studying *cis*-regulatory evolution

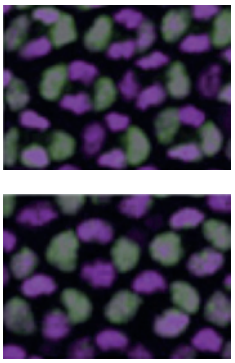
Genomic surveys show that *cis*-regulatory divergence is common between species: using allele-specific expression to detect differences in *cis*-regulatory activity[18,19], researchers have found that >50% of genes show such differences between closely related species of yeast[20] and fruitflies[21]. The genetic basis of this divergent *cis*-regulatory activity remains unknown for the vast majority of genes. This is because elucidating the genetic and molecular mechanisms that are responsible for *cis*-regulatory divergence is a complex empirical task that requires identifying the CRE (or CREs) that controls the expression pattern of interest, finding functionally divergent sites within this region and determining how these changes in *cis*-regulatory sequence alter biochemical interactions with and among *trans*-acting factors that mediate gene expression. Fortunately, technical advances in many areas have begun to make this type of work easier in recent years.

Finding enhancers. As described above, enhancers are the type of CRE that are most often responsible for *cis*-regulatory divergence. Consequently, identifying the genetic changes that are responsible for divergent *cis*-regulatory activity often begins with locating the enhancers that control divergent expression patterns of orthologous genes. Enhancer sequences tend to be more conserved than sequences predicted to be non-functional, and searches for enhancers are often guided by patterns of sequence conservation and/or predicted TF binding sites[22] (BOX 2; FIG. 1a). This approach takes advantage of the many genome sequences that are available, as well as databases describing the sequence specificity of individual TFs[23]. Recently, data from empirical studies determining genome-wide TF binding[24–26] have begun to supplant the use of computationally predicted TF binding sites, improving the accuracy of this approach[27]. Nevertheless, and despite the many cases in which sequence conservation has successfully been used to find orthologous enhancers, it is clear that this approach is insufficient for identifying all enhancers within a genome.

Identifying candidate sites. After a pair of orthologous enhancers with divergent activity has been verified *in vivo* (BOX 3), the next task is to generate a list of candidate sites that might be responsible for the divergent activity. An initial list can be generated by simply

Box 2 | **Conservation and divergence of *cis*-regulatory element sequence and function**





Cis-regulatory activity (green) in cone cells (magenta) of 24-hour-old transgenic *D. melanogaster* pupae

When investigating the genetic basis of divergent *cis*-regulatory activity, researchers often start by using DNA sequence conservation among species to identify regions of the genome that are likely to contain *cis*-regulatory elements (CREs). In other words, sequence similarity is used to search for the genetic basis of functional differences. Because CREs are required for the proper development and function of an organism, there is selection pressure to maintain their activity by eliminating most mutations that disrupt their function. This results in levels of sequence conservation in CREs that are higher than those found in regions of the genome that do not have a known function (for example, pseudogenes); however, more sequence changes are expected in CREs than in coding sequences because of the high degeneracy of transcription factor (TF) binding sites and the flexibility in the number, order and positioning of these sites within a CRE. The conservation signal of a CRE can persist despite functional divergence because, as described in the main text, divergent activity can result from very few sequence changes. Comparisons of non-coding sequences from vertebrate genomes have revealed discrete regions of highly conserved sequences that often contain CREs[89]. Among *Drosophila* species, however, such highly conserved regions are more difficult to recognize[90], apparently because species in this genus do not have enough 'nonfunctional' DNA to provide the neutrally evolving baseline that is needed to recognize more highly conserved regions[91].

Just as divergent *cis*-regulatory activities can result from highly conserved sequences, similar *cis*-regulatory activities can result from highly divergent sequences[92–94]. A seminal demonstration of this resulted from comparing the activity and sequence of the stripe 2 enhancers of *even skipped* (*eve*) from multiple *Drosophila* species[15,95,96]. These experiments showed that enhancer function was conserved despite extensive changes in the sequence that caused gains, losses and major rearrangements of TF binding sites. Functional stasis that occurs despite sequence and TF binding site turnover has also been reported for other genes in *Drosophila* spp., including *Yolk protein* genes[97], *yellow*[17,98] and *dPax2* (also known as *shaven*)[99]. The figure shows an example from REF. 99. The location of binding sites for the Suppressor of Hairless (Su(H)), Lozenge (a Runx protein) and two ETS family TFs, and the locations for five putative regulatory motifs (labelled α, β, γ, δ and ε), are shown for in orthologous sparkling (spa) enhancers in *Drosophila melanogaster* and *Drosophila pseudoobscura*. These enhancers regulate a conserved pattern of expression in cone cells (images in the right-hand panel). Because of their limited sequence similarity among species, CREs evolving in this manner can be overlooked unless functional *in vivo* tests of *cis*-regulatory activity are performed. The figure is modified, with permission, from REF. 99 © (2011) Elsevier.

Convergent phenotypes
Similar phenotypes that independently evolved in different lineages.

Candidate sites
In the context of this article, one or more nucleotide changes that correlate with a change in expression and thus might be causing the change in expression.

Degeneracy
In the context of transcription factor binding sites, this is the ability of a transcription factor to bind to multiple (usually related) DNA sequences.

Outgroup
A related but taxonomically more distant species that can be used to infer the ancestral state of a particular site in DNA.

comparing the two sequences; however, including an orthologous enhancer from one or more outgroup species that has similar activity to one of the enhancers under study can help to refine the list of candidates by identifying sites at which the species that has derived enhancer activity also has a derived allele (FIG. 1b). Such sites are the strongest candidates for contributing to *cis*-regulatory divergence. Availability of genome sequences from diverse organisms[28], combined with the potential for collecting such sequences from a more suitable outgroup, if needed, using next-generation sequencing[29], have greatly accelerated the identification of candidate sites in recent years. Depending on the level of sequence divergence between the species under study, such analysis may identify either a few or many divergent sites as candidates.

Testing candidate sites. Functional tests comparing the activity of enhancer alleles that vary for one or a few of these candidate sites must be performed *in vivo* to conclusively implicate a particular sequence change in *cis*-regulatory divergence. Ideally, both alleles of a candidate site should be individually tested in both of the species from which the divergent enhancers were derived; this test can be done by constructing reciprocal pairs of genetically modified organisms that only differ in a single mutation in an enhancer sequence (BOX 3). Transgenes can be generated that contain either a complete gene or enhancer sequences positioned upstream of a reporter gene and delivered by various methods, depending on the organism: for example, transfection of cultured cells (such as in human cells), transient transformation of live animals (such as in sticklebacks),

introduction of extra-chromosomal genetic material (such as in nematodes), random integration into the genome (such as in plants and zebrafish) and targeted integration into a predetermined site (such as in flies). The ultimate level of resolution is homologous replacement of the endogenous locus, and this can be done in yeast or in mice. Improvements in cloning methods used for constructing transgenes[30], robustness of reporter genes (that is, genes that produce brighter, more stable and faster maturing fluorescent proteins)[31], targeted integration of transgenes[32] and transformation of non-model species[33] have all helped to lower the barrier for performing these important, but technically demanding, experiments substantially during the last decade.

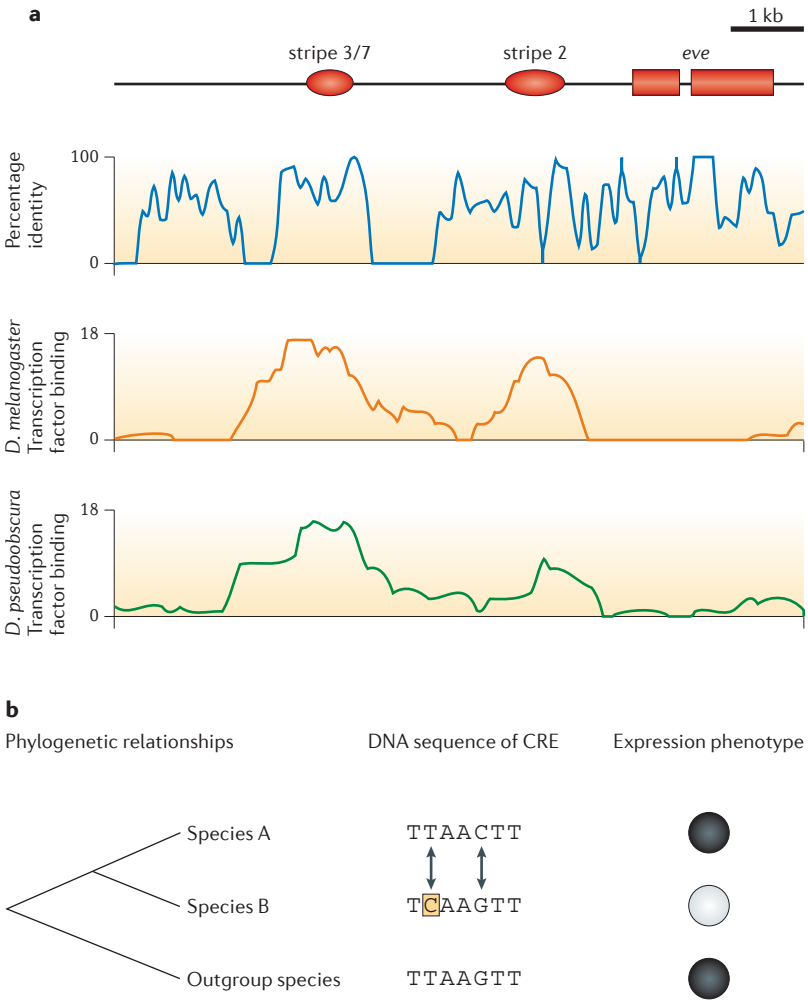


Figure 1 | **Strategies for finding enhancers and divergent sites within them.** **a** | Many DNA sequences that function as enhancers show substantial sequence conservation and clusters of transcription factor (TF) binding sites. For example, in *Drosophila melanogaster* and *Drosophila pseudoobscura*, DNA sequence similarity (percentage identity) and the number of predicted TF binding sites have peaks that overlap with each other, as well as with regions of the genome that have been shown to activate expression of the *D. melanogaster even skipped* (*eve*) gene. Rectangles indicate exons of *eve*, whereas ovals indicate the stripe 3/7 and stripe 2 enhancers of *eve*. The TF binding scores that are plotted on the *y* axis reflect predicted binding sites for five TFs (*bicoid*, *hunchback*, *kruppel*, *knirps* and *caudal*). Data derived from REF. 102. **b** | Including an outgroup when examining divergent *cis*-regulatory activity can help to identify candidate sites that are most likely to be responsible for the functional divergence. In the hypothetical example shown, comparing the *cis*-regulatory sequences and expression phenotypes of species A and B alone suggests that the T/C and/or C/G nucleotide differences (indicated with double-headed arrows) might contribute to the difference in *cis*-regulatory activity. Also, considering the sequence and expression phenotype of the outgroup species shown allows the derived nucleotide changes to be inferred and suggests that the C in species B (boxed) is most likely to be responsible for the divergent expression phenotype. CRE, *cis*-regulatory element.

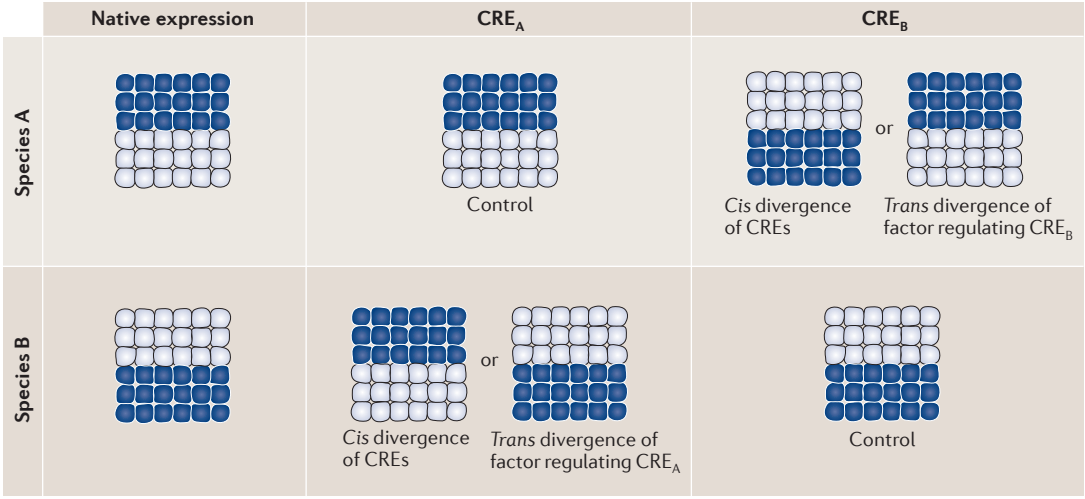
Assessing consequences for transcription factor binding. Sequence changes in CREs are most often assumed to exert their effects on *cis*-regulatory activity by altering TF binding. In most cases, determining the impact of a particular sequence change on TF binding is not straightforward because verified TF binding sites are rarely known for a CRE of interest. Prior knowledge of gene regulation in the tissue type and species of interest can sometimes be used to make educated guesses about which TF binds to a divergent sequence[34–36], but this is only possible for few traits in few species. Methods that systematically test every TF from particular species for binding to a CRE of interest[37] should help to make this type of analysis possible in more cases, and data sets that describe genome-wide binding of tens to hundreds of TFs for model organisms such as *Saccharomyces cerevisiae*[24], *Drosophila melanogaster*[25] and *Caenorhabditis elegans*[26] are already available. A list of candidate TFs can also, in principle, be generated using *cis*-regulatory DNA to purify bound TFs and then sequencing fragments of the extracted proteins by mass spectrometry[38], although current methods would benefit from further refinement.

After the TF that binds to a site of interest has been identified, the impact of *cis*-regulatory sequence divergence on TF binding can be assessed using either a traditional *in vitro* method of measuring protein–DNA interactions — such as electrophoretic mobility shift assays (EMSAs) — or a newer *in vivo* method that involves chromatin immunoprecipitation (ChIP) followed by quantitative PCR, microarray analysis or massively parallel sequencing (reviewed in REF. 39). More precise quantitative estimates of the impact that a mutation has on TF binding can also now be obtained using techniques such as protein binding microarrays[40] and mechanically induced trapping of molecular interactions (MITOMI)[41].

Mechanisms of *cis*-regulatory divergence

With the help of the methodological advances described above, the number of studies identifying mutations that are responsible for differences in *cis*-regulatory activity between species has dramatically increased in recent years, although it remains quite small (Supplementary information S1 (table)). The most mechanistically detailed of these studies examine closely related species that have a limited number of sequence changes between orthologous enhancers, whereas genomic studies of TF binding primarily focus on more distantly related species. Together, these data have begun to answer questions about the genetic mechanisms underlying *cis*-regulatory divergence.

Box 3 | **Testing DNA sequences for *cis*-regulatory activity and divergence**



Activity of a *cis*-regulatory element (CRE) cannot be predicted from its DNA sequence alone; rather, CRE activity must be assessed in living cells. This is most frequently done by using the putative CRE to activate expression of an easy-to-visualize reporter protein with a chimeric transgene[100]. Comparing expression of the reporter gene to expression of the native gene that is thought to be regulated by the CRE is helpful for evaluating whether the reporter gene expression is likely to be an experimental artefact, as well as for determining which portion of the gene's total expression is controlled by that CRE.

To test for *cis*-regulatory divergence, orthologous CREs can be compared in a common *trans*-regulatory environment. This can be accomplished by introducing the pair of reporter genes into the same transgenic host; in the figure, orthologous CREs from species A (CRE$_A$) and from species B (CRE$_B$) are each tested in species A and in species B. Both transgenes are then regulated by the same *trans*-acting factors, making any differences in their activity attributable to *cis*-regulatory divergence. A caveat to this experiment is that one or more *trans*-acting factors that regulate the CREs might also have diverged between species, such that activity of a reporter gene in a heterologous transgenic host species is different than its activity in the species from which it was derived[98,101]. Fortunately, many *trans*-regulatory factors appear to be highly conserved between species, making the use of a single transgenic host species a reasonable starting point for analysis. The most complete understanding of regulatory divergence, however, requires testing for both *cis*- and *trans*-regulatory divergence. Differences in *trans*-regulation between species can be detected by comparing the activity of the same CRE in multiple species. For example, in the figure, the activity is represented by patches of white and blue cells with different spatial arrangements. *Cis*-regulatory divergence can be inferred when the orthologous CREs give different expression patterns in the same transgenic host species. *Trans*-regulatory divergence can be inferred when a CRE results in a different expression pattern when expressed in a different species (for example, when a CRE from species A is transformed into species B, and it does not give the same expression pattern as when it is transformed into species A).

Reporter genes can detect changes in *cis*-regulatory activity and can even be used to map the genetic changes that cause any observed divergence, but they are insufficient to prove that this divergence affects an organism's phenotype[34]. To do this, divergent sites in CREs must be swapped in the context of the native gene, and the activity of this modified gene copy should be assessed in mutants that otherwise lack function of the gene[42].

What types of mutations underlie* cis*-regulatory divergence? Substitutions of individual nucleotides that result from point mutations are commonly observed between species, and even a small number of nucleotide substitutions can be sufficient to alter *cis*-regulatory activity (FIG. 2a). For example, 13 or fewer nucleotide substitutions within each divergent enhancer are responsible for differences in *cis*-regulatory activity between *Drosophila* species that are associated with loss of trichomes[42] and with divergent pigmentation[35,43]. Similarly, at least two and not more than 13 single-nucleotide changes are sufficient for a gain of gene expression in the limb in humans relative to chimpanzees[44], and at least 11 substitutions are responsible for novel gene expression in the optic lobe of *Drosophila santomea*[45]. In *Drosophila takahashii*, even a single-nucleotide substitution that disrupts binding of the doublesex (DSX) TF to an enhancer has been shown to be sufficient for the evolution of a sexually monomorphic expression pattern from a sexually dimorphic one[36].

Deletions are also common between species and can change *cis*-regulatory activity. Recurrent deletions that disrupt *cis*-regulatory activity have been shown to create alleles that contribute to the loss of pelvic structures in freshwater populations of threespine sticklebacks[46] as well as to the loss of dark abdominal pigmentation in *D. santomea*[43]. In the case of threespine sticklebacks, a 'fragile' sequence composition appears to make the enhancer especially prone to deletions. Five hundred and nine regions of the genome that are highly conserved between chimpanzees and other mammals are missing from the human genome, which is suggestive of human-specific deletions of putative CREs[47]. Transgenic analysis of two of these regions confirmed that *cis*-regulatory activities that are present in sequences from chimpanzees

Trans-regulatory environment
The collection of proteins, RNAs and other *trans*-acting molecules within a cell.

Host species
In the context of transgenic analysis, the species transformed with foreign DNA.

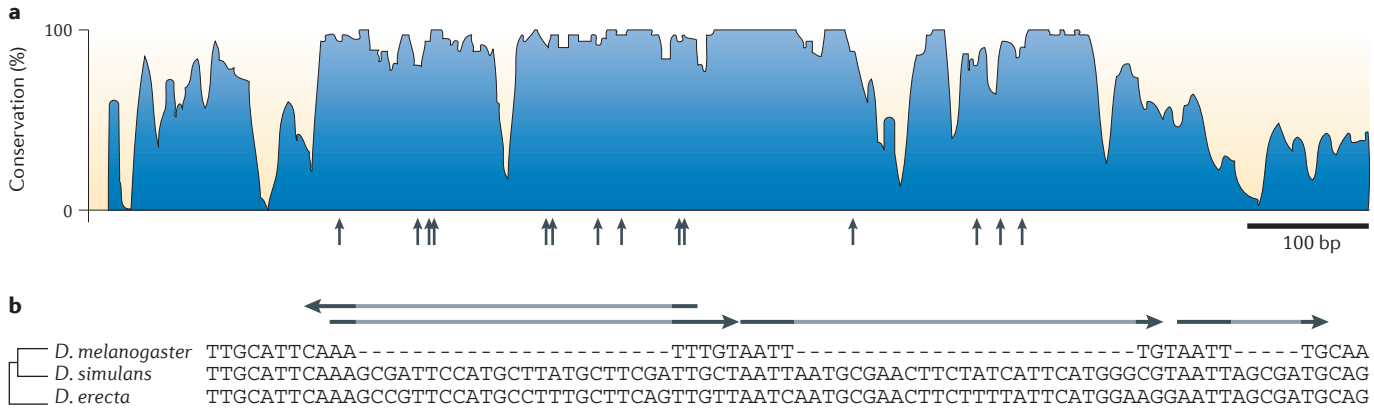


Figure 2 | **Examples of point mutations and deletions causing divergent *cis*-regulatory activity. a** | A 1,007 bp enhancer located 5′ of the *shavenbaby* promoter contains 14 nucleotide differences that are specific to *Drosophila sechellia*. A measure of sequence conservation for this enhancer is plotted, and vertical arrows indicate the location of the 14 *D. sechellia*-specific alleles. **b** | Multiple deletions in an enhancer for the *desatF* gene (also known as *Fad2*) (only part of which is shown) have occurred in *Drosophila melanogaster* and create copies of the hexamer sequence AATTTG — three on one strand and one on the other (indicated with horizontal arrows). The *D. melanogaster* gene sequence is compared with those of *Drosophila simulans* and *Drosophila erecta*. Although no specific TF has yet been shown to bind to these sites, mutating them shows that they are essential for enhancer activity. Part **a** is modified, with permission, from REF. 42 © (2011) Macmillan Publishers Ltd. All rights reserved. Part **b** is modified from REF. 36.

and mice were missing from the orthologous region of the human genome. Loss of these enhancers in humans has been implicated in the loss of sensory vibrissae and penile spines, as well as in the expansion of a specific brain region[47]. In all of these cases, deletions decreased *cis*-regulatory activity (as they are most often expected to do). However, deletions can also increase the activity of a CRE, either by removing binding sites for repressors or — as was recently demonstrated in *D. melanogaster*[36] — by bringing together sequences that create novel binding sites for activators (FIG. 2b).

Insertions are a third potential source of divergent *cis*-regulatory activity, and they can alter a *cis*-regulatory activity by introducing new TF binding sites or disrupting spacing between existing TF binding sites. Indeed, insertions that alter the spacing of binding sites contribute to divergent *cis*-regulatory activity at the *bric a brac 1* gene between *D. melanogaster* and *Drosophila willistoni*[35]. Insertions of transposable elements also have the potential to contribute to *cis*-regulatory divergence by bringing new CREs to existing genes. Such insertions are an abundant source of phenotypic variation within both plant and animal species[48–51], and comparative genomics has identified conserved retrotransposon insertions that appear to function as CREs[52,53].

How does sequence divergence relate to transcription factor binding? TF binding sites are the functional units in *cis*-regulatory sequences; thus, changes that affect *cis*-regulatory activity are expected to alter TF binding sites. Changes in TF binding are commonly described either as gains or losses of binding sites or as changes in the affinity of the site, depending on the sensitivity of the experiment and the magnitude of the change. Changes in the spacing between TF binding sites can also affect enhancer function. To date, only a handful of studies that link a change in *cis*-regulatory activity to specific sequence changes in a CRE have determined the impact of these changes on the binding of a specific TF. These studies include: the work mentioned above regarding DSX binding[36]; a study showing that the acquisition of a novel binding site for the patterning TF Engrailed was necessary for a novel expression pattern in wings of *Drososphila biarmipes* males[34]; and two other studies of *Drosophila* spp. pigmentation genes (*bric a brac 1* (REF. 35) and *yellow*[54]), showing that multiple mutations resulted in the alterations of crucial TF binding sites for Abdominal B (ABDB). Interestingly, these last two studies showed that mutations that change which strand of DNA contains the TF binding site, without altering its affinity for the TF, affected *cis*-regulatory activity.

Studies that link changes in *cis*-regulatory sequence to changes in TF binding without providing evidence of their consequence for *cis*-regulatory activity are more common. For example, He *et al.*[55] examined the relationship between DNA sequence evolution and TF binding by using a set of 645 experimentally identified binding sites for 30 different TFs in *D. melanogaster* to determine whether polymorphisms and substitutions within these sites were expected to increase or decrease affinity for the bound TF (based on published position weight matrices). Polymorphisms that were predicted to decrease affinity were more common than polymorphisms that were predicted to increase affinity in *D. melanogaster*, although the majority of substitutions in *D. melanogaster* relative to *Drosophila simulans* were predicted to increase affinity. ChIP-based methods, such as ChIP followed by microarray (ChIP–chip) and ChIP followed by sequencing (ChIP–seq), have also been used to compare TF binding between species (for a recent comprehensive review of such

Position weight matrices
Quantitative representations of DNA-binding specificity for a transcription factor protein.

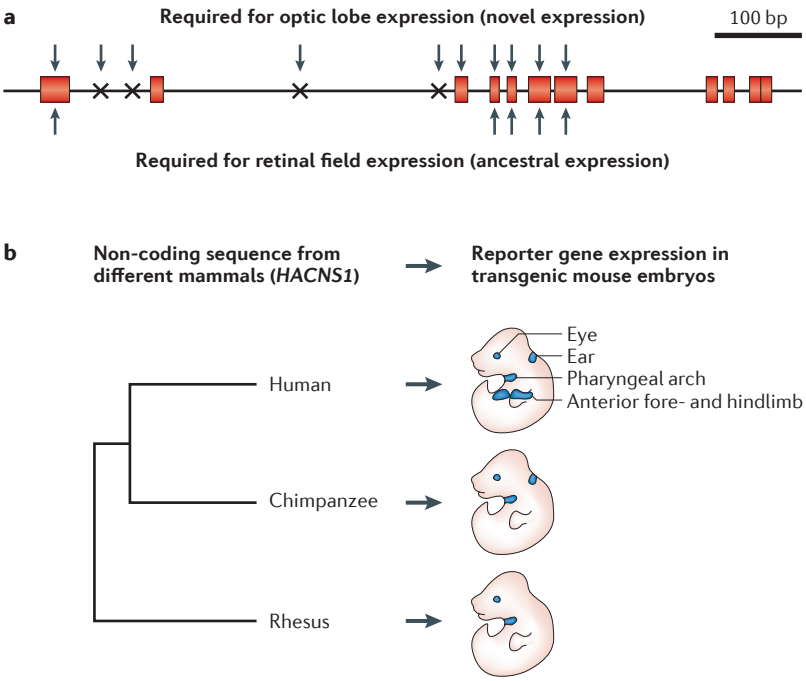


Figure 3 | **Novel enhancer activities have evolved by co-option of existing enhancers. a** | Four mutations that are specific to *Drosophila santomea* (indicated by the crosses) work with conserved sequences nearby (red boxes) to produce expression in the developing optic lobe that is not seen in other species. Some of the conserved regions that are required for this activity are also required for expression in the developing retinal field in *D. santomea* and other members of the melanogaster species subgroup, suggesting that these conserved sequences contributed to enhancer activity before the evolution of the optic lobe enhancer. Down arrows indicate *cis*-regulatory sequences in *D. santomea* that are required for enhancer activity in the optic lobe, and up arrows indicate *cis*-regulatory sequences in *D. santomea* that are required for enhancer activity in the retinal field. Data derived from REF. 45. **b** | A region of non-coding sequence that is highly conserved among mammals, *HACNS1*, has evolved a novel enhancer activity relative to chimpanzees and rhesus macaques. Specifically, the human copy of *HACNS1* activates novel expression of a reporter gene in transgenic mice in the anterior fore- and hindlimb buds during embryonic development. Orthologous sequences from chimpanzees and rhesus macaques lack this enhancer activity when assayed in transgenic mice. *Cis*-regulatory sequences from all three species do, however, activate expression in other tissues. This suggests that the human-specific limb bud enhancer evolved in the context of an existing enhancer. Schematized representations of reporter gene expression (blue) in transgenic mice at embryonic day 11.5 are shown. Data derived from REF. 44.

studies, see REF. 56). All but one of these studies indicates that gains and losses of TF binding are common between species[57–60], even for central components of the transcription machinery such as RNA polymerase II[61], and these changes often appear to be due to the divergence of underlying *cis*-acting sequences[62,63]. The exception is a study comparing binding of six TFs among *Drosophila* species that found conservation of 86–99% of TF binding sites but also found quantitative differences in their affinity[64]. This study differs from the others in that the species compared are more closely related, and binding was assayed earlier in the organisms' development; TF binding (and gene regulation in general) may be more highly constrained early in development.

Where do new expression patterns come from? *Cis*-regulatory divergence that generates a novel expression pattern is particularly interesting to evolutionary biologists because it is expected to contribute to the origin of phenotypic novelty. Such expression patterns can arise from *de novo* evolution of CREs (that is, CREs that result solely from the accumulation of new mutations in sequences formerly lacking *cis*-regulatory activity), duplication and divergence of existing CREs, transpositions that cause existing CREs to control expression of new genes or co-option of existing CREs (that is, new mutations that exploit parts of an existing CRE to evolve novel *cis*-regulatory functions). Examples of each of these mechanisms can be found in the literature[50,65–67]. Despite sparse sampling of CREs controlling novel expression patterns, recent studies suggest that apparently novel *cis*-regulatory activities might often result from co-option of existing enhancers[34,44,45,68]. However, because co-option describes a modification of an existing enhancer, it is debatable whether such activities should be considered to result from the evolution of novel CREs or simply from divergence of an existing CRE.

One example of enhancer co-option comes from Rebeiz *et al.*[45], in which an intronic enhancer controlling a putatively novel expression pattern of *Neprilysin 1* (*Nep1*) in *D. santomea* was identified. Four *D. santomea*-specific mutations were shown to be required for this activity, but the ability of these changes to produce the apparently novel expression pattern was dependent on nearby regions of sequence that also contributed to conserved patterns of *Nep1* expression (FIG. 3a). When these four derived changes were changed back to their ancestral state, 40% of the 'novel' expression pattern was still observed, suggesting that the ancestral enhancer shared with other *Drosophila* species might also have driven expression in this pattern and that the absence of this pattern element in the other surveyed extant species may result from the acquisition of mutations that suppress this activity. Consistent with this hypothesis, a more sensitive detection technique revealed low levels of expression in the '*D. santomea*-specific' expression domain in *D. simulans*. This suggests that these four changes might have simply disrupted repression of a cryptic *cis*-regulatory activity rather than being responsible for the evolution of a completely novel one. A similar situation was observed for two independently evolved novel expression patterns of the *yellow* gene in spots on the wings of various *Drosophila* species[34,68].

In humans, 13 derived changes within an 81 bp region of a highly conserved non-coding regulatory element (*HACNS1*) were shown to confer putatively novel expression in the developing limb buds of transgenic mice; this expression pattern was not seen when orthologous sequences from chimpanzees and rhesus macaques were used to drive reporter gene expression[44] (FIG. 3b). These orthologous human, chimpanzee and macaque *cis*-regulatory sequences did, however, drive similar expression in other developing tissues. Subsequent work, however, has called into question whether these human-specific sequences should be considered to be a part of a novel CRE. Deletion of the entire 81 bp region harbouring

human-specific changes in *HACNS1* resulted in limb bud expression in transgenic mice that was similar to that of the intact human *HACNS1* sequence[69]. This finding strongly suggests that the human-specific changes disrupted repression of an ancestral CRE that was capable of driving expression in the limb buds. Taken together, these case studies from primates and flies suggest that genomic sequences may be primed to evolve new expression patterns by the presence of latent enhancers repressed by neighbouring *cis*-regulatory sequences.

How do sites contributing to CRE divergence interact? When multiple mutations affect a phenotype, the interaction between them can be described as 'additive' or 'epistatic'. Cases in which each mutation has the same effect on the phenotype — regardless of whether another mutation, or mutations, are present — are considered to be additive, whereas cases in which the effect of one mutation depends on the presence or absence of other mutations are considered to be epistatic. Determination of whether divergent sites interact additively or epistatically requires measurement of the effect of each divergent site both individually and in combination with each other. Recently, such an analysis has been completed for two enhancers that have divergent activity between *Drosophila* species[35,42] and one enhancer that has adaptive differences in activity between populations of *D. melanogaster*[70]. In one case, the phenotypic consequences of mutations that affect binding sites for a TF interacted approximately additively with the phenotypic consequences of mutations affecting binding sites for another TF[35]. In the other two cases, evidence of epistasis was observed. Frankel *et al.*[42] found that substituting seven clusters of candidate sites between species at the same time had effects on the phenotype that were stronger than the sum of effects from substituting each cluster individually, whereas Rebeiz *et al.*[70] found that substituting five changes at the same time had less of an effect on the phenotype than the sum of individual effects for the five mutations. Epistatic interactions have also been observed among *cis*-regulatory polymorphisms segregating in human populations[71]. With the limited number of studies comparing the effects of divergent sites individually and in groups, it is premature to speculate whether one type of interaction is more common than another.

Insight into evolutionary processes

Long-standing questions about the evolutionary process can be answered by identifying the genetic basis of divergent phenotypes. Given that changes in *cis*-regulatory activity are a major source of phenotypic divergence, identifying the sites that are responsible for divergent activity of *cis*-regulatory sequences can help to resolve these questions.

Standing genetic variation versus new mutations. Population genetic theory often assumes that beneficial mutations sweep to fixation shortly after they arrive, but it is now known that alleles selected from standing genetic variation (that is, alleles that have been segregating within a population for a considerable time) can also contribute to adaptive phenotypes[72]. Studies of *cis*-regulatory divergence provide evidence that both types of changes contribute to phenotypic evolution. For example, *cis*-regulatory sequences of the *ebony* gene that contribute to an adaptive pigmentation difference within *D. melanogaster* show evidence of a selective sweep involving both standing genetic variation and new mutations[70,73]. In other *Drosophila* species, pigmentation differences within and between species are caused (in part) by alleles of the *tan* gene that share putative *cis*-regulatory changes that appear to be derived from standing genetic variation in a common ancestor[74]. In sticklebacks, freshwater populations often lose their bony body armour, which has been shown to result from selection for existing *cis*-regulatory alleles of the *Ectodysplasin* gene from standing genetic variation in the marine population[75]. However, they also tend to lose pelvic structures, and alleles that are responsible for this phenotype appear to have been independently generated many times by recurrent deletions[46].

Mutations of large versus small effect. For most of the twentieth century, phenotypic divergence was assumed to result from the fixation of many mutations of small effect, but this view began to change circa 1980 when quantitative geneticists uncovered 'major' effect loci underlying phenotypic variation[76]. Recently, however, fine-scale genetic dissection of some of these major effect quantitative trait loci (QTLs) has revealed many mutations of small effect[42,77], favouring a model in which effects of substitutions are exponentially distributed — that is, some have large effects, but most have small effects. Based on the data that are currently available, mutations giving rise to *cis*-regulatory divergence appear to fit this model; single mutations of large effect exist[36,46], but multiple mutations of individually small effect that combine to produce divergent *cis*-regulatory activity are more common[35,42–44,68]. However, it is important to note that changing *cis*-regulation of a gene may[42] or may not[34] be sufficient to alter organismal phenotypes, and quantitative genetic studies, as well as population genetic predictions about the effect size of substitutions, are based on organismal phenotypes. Furthermore, much of the data that are currently available on the effect size of fixed *cis*-regulatory mutations come from *Drosophila* spp.; other species — for example, those with different effective population sizes in which the relative effectiveness of natural selection and genetic drift differ[78] — may tend to fix different types of mutations at different rates.

Genetic basis of convergent phenotypes. Cases in which two species have independently evolved similar phenotypes (that is, convergent evolution[79]) are often used to study the genetic basis of adaptation because they allow repeatability of the evolutionary process to be assessed. That is, by studying the genetic basis of convergent phenotypes, it is possible to ask whether the evolution of similar phenotypes multiple times involves the same genes, the same CRE, the same region of a protein, the same sites within these regions or the same type of mutation (or mutations).

Selective sweep
The increase in frequency of an allele (and closely linked chromosomal segments) that is caused by selection for the allele.

Studies of *cis*-regulatory evolution have examined parallel phenotypic changes in a number of organisms that can be used to assess the similarity of molecular mechanisms that underlie convergent evolution. For example, in threespine sticklebacks, multiple freshwater populations have evolved similar morphological phenotypes since diverging from their ocean-dwelling ancestors. Genetic dissection of these phenotypes shows that the same genes, and even the same divergent *cis*-regulatory alleles, are sometimes — but not always — involved[46,75]. Among drosophilids, multiple species have independently lost some of their larval trichomes, and this has been shown to be due to *cis*-regulatory changes affecting the *shavenbaby* (also known as *ovo*) gene in at least two cases[80]. Convergent pigment patterns among *Drosophila* species, such as male-specific wing spots[68] and heavily pigmented head cuticles[17], also appear to be caused (in part) by divergent expression of the same gene (*yellow*), but these expression differences have evolved through the co-option of different CREs in different species. Convergent losses of dimorphic body pigmentation, however, have evolved using different genes in different species[54]. Once again, even this limited number of case studies illustrates a variety of mechanisms.

Neutral versus non-neutral evolution. One of the longest-standing debates in evolutionary biology surrounds the relative contributions of neutral and non-neutral changes to phenotypic evolution. Assuming that the mutations that are responsible for *cis*-regulatory divergence contribute to phenotypic evolution (BOX 1), we can use these types of changes to address this issue. Many of the case studies in which the mutations that are responsible for divergent activity of a CRE have been identified (see Supplementary information S1 (table)) have also examined patterns of polymorphism and divergence to see whether the divergence of the CRE was consistent with a model of neutral evolution. Excess substitutions, suggesting the action of either positive selection or reduced functional constraint, were observed within divergent enhancers for the *Drosophila shavenbaby* gene[42]. A similar pattern of sequence divergence was reported for *HACNS1*, the enhancer with a novel expression domain in developing limbs of humans[44], but this has recently been reinterpreted as the result of biased gene conversion rather than positive selection or relaxed constraint[69,81]. A reduction of heterozygosity (an observation that also suggests positive selection) was observed for the divergent stickleback *Pitx1* CRE[46]. In fact, among the studies listed in Supplementary information S1 (table) that tested for evidence of positive selection, only Jeong *et al.*[43] failed to find any evidence that was consistent with positive selection; rather, these authors found evidence of recent relaxed constraint within the divergent CRE of the *D. santomea tan* gene. Thus, overall, much of the *cis*-regulatory divergence that has been characterized to date appears to be the product of natural selection. Surprisingly, even CREs with activity that does not appear to differ among species show patterns of sequence variation that suggest they may be evolving primarily by natural selection[55].

Concluding remarks

As described in this Review, our understanding of the genetic and molecular mechanisms that are responsible for divergent *cis*-regulatory activity has grown substantially in recent years — a growth that has been facilitated by technological advances in many areas. Despite this growth, the total number of case studies in which sites that are responsible for *cis*-regulatory divergence have been identified remains quite small (Supplementary information S1 (table)), and most of the available data are derived from only a few organisms (for example, *Drosophila* spp. and sticklebacks) and a few traits (for example, pigmentation, trichomes and skeletal structures). Nevertheless, this small collection of case studies includes at least one example of practically every type of evolutionary change predicted. Determining whether some types of change are more common than others under certain conditions requires an expanded collection of case studies. Only then will we be able to answer questions such as: are particular types of mutations more likely to underlie differences in *cis*-regulatory activity within or between species or in large or small populations, and are their phenotypic effects adaptive or neutral? If such generalities exist and can be uncovered, future evolutionary paths might become predictable.

Future studies of *cis*-regulatory divergence should also continue focusing on the relationships between genotypes and phenotypes. This will require elucidating in great detail the molecular mechanisms that link the two for many genotypes and phenotypes. Currently, the literature includes very few studies that connect changes in *cis*-regulatory DNA sequences with changes in TF binding, *cis*-regulatory activity and organismal phenotypes. Comparing species with a range of divergence times using functional genomic tools such as RNA-seq and ChIP–seq will also help to achieve a more complete description of some of these relationships. With high-throughput phenotyping projects now underway in a number of model species (summarized in REF. 82), it may soon be straightforward to link a change in gene expression that is caused by a divergent *cis*-regulatory sequence to a change in a specific organismal phenotype. Ultimately, understanding the molecular mechanisms of *cis*-regulatory evolution is expected to provide novel insight into the origins of biodiversity.

1. Zuckerkandl, E. & Pauling, L. in *Evolving Genes and Proteins* (eds Bryson, V. & Vogel, H.) 97–166 (Academic Press, New York, 1965).
2. Britten, R. J. & Davidson, E. H. Gene regulation for higher cells: a theory. *Science* **165**, 349–357 (1969).
3. King, M.-C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
4. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
5. Stern, D. L. & Orgogozo, V. The loci of evolution: how predictable is genetic evolution? *Evolution* **62**, 2155–2177 (2008).
6. Ong, C.-T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Rev. Genet.* **12**, 283–293 (2011).

7. Levine, M. Transcriptional enhancers in animal development and evolution. *Curr. Biol.* **20**, R754–R763 (2010).
8. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327–339 (2011).
9. Brown, R. P. & Feder, M. E. Reverse transcriptional profiling: non-correspondence of transcript level variation and proximal promoter polymorphism. *BMC Genomics* **6**, 110 (2005).
10. Savinkova, L. K. *et al.* TATA box polymorphisms in human gene promoters and associated hereditary pathologies. *Biochemistry* **74**, 117–129 (2009).
11. Wray, G. A. The evolutionary significance of *cis*-regulatory mutations. *Nature Rev. Genet.* **8**, 206–216 (2007).
12. Hong, J. W., Hendrix, D. A. & Levine, M. S. Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314 (2008).
13. Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. Shadow enhancers foster robustness of *Drosophila* gastrulation. *Current Biol.* **20**, 1562–1567 (2010).
14. Kleinjan, D. A. & van Heyningen, V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* **76**, 8–32 (2005).
15. Hare, E. E., Peterson, B. K., Iyer, V. N., Meier, R. & Eisen, M. B. Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet.* **4**, e1000106 (2008).
16. Cande, J., Goltsev, Y. & Levine, M. S. Conservation of enhancer location in divergent insects. *Proc. Natl Acad. Sci. USA* **106**, 14414–14419, (2009).
17. Kalay, G. & Wittkopp, P. J. Nomadic enhancers: tissue-specific *cis*-regulatory elements of yellow have divergent genomic positions among *Drosophila* species. *PLoS Genet.* **6**, e1001222 (2010).
18. Cowles, C. R., Hirschhorn, J. N., Altshuler, D. & Lander, E. S. Detection of regulatory variation in mouse genes. *Nature Genet.* **32**, 432–437 (2002).
19. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430**, 85–88 (2004).
20. Tirosh, I., Reikhav, S., Levy, A. A. & Barkai, N. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* **324**, 659–662 (2009).
21. McManus, C. J. *et al.* Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.* **20**, 816–825 (2010).
22. He, X., Ling, X. & Sinha, S. Alignment and prediction of *cis*-regulatory modules based on a probabilistic model of evolution. *PLoS Comput. Biol.* **5**, e1000299 (2009).
23. Matys, V. *et al.* Transfac: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31**, 374–378 (2003).
24. Abdulrehman, D. *et al.* Yeastract: providing a programmatic access to curated transcriptional regulatory associations in *Saccharomyces cerevisiae* through a web services interface. *Nucleic Acids Res.* **39**, D136–D140 (2011).
25. Roy, S. *et al.* Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787–1797 (2010).
26. Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
27. Visel, A. *et al.* ChIP–seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
28. Fujita, P. A. *et al.* The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* **39**, D876–D882 (2011).
29. Hudson, M. E. Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Res.* **8**, 3–17 (2008).
30. Warming, S., Costantino, N., Court, D. L., Jenkins, N. A. & Copeland, N. G. Simple and highly efficient BAC recombineering using GalK selection. *Nucleic Acids Res.* **33**, e36 (2005).
31. Ghim, C. M., Lee, S. K., Takayama, S. & Mitchell, R. J. The art of reporter proteins in science: past, present and future applications. *BMB Reports* **43**, 451–460 (2010).
32. Groth, A. C., Fish., M., Nusse, R. & Calos, M. P. Construction of transgenic *Drosophila* by using the site-specific integrase from phage PhiC31. *Genetics* **166**, 1775–1782 (2004).
33. Holtzman, S. *et al.* Transgenic tools for members of the genus *Drosophila* with sequenced genomes. *Fly* **4**, 349–362 (2010).
34. Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. & Carroll, S. B. Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **433**, 481–487 (2005).
35. Williams, T. M. *et al.* The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* **134**, 610–623 (2008).
36. Shirangi, T. R., Dufour, H. D., Williams, T. M. & Carroll, S. B. Rapid evolution of sex pheromone-producing enzyme expression in *Drosophila*. *PLoS Biol.* **7**, e1000168 (2009).
This study shows how deletions in *cis*-regulatory sequences can lead to the formation of binding sites for transcriptional activators, resulting in divergent expression.
37. Deplancke, B. *et al.* A gene-centered *C. elegans* protein–DNA interaction network. *Cell* **125**, 1193–1205 (2006).
38. Drewett, V. *et al.* DNA-bound transcription factor complexes analysed by mass-spectrometry: binding of novel proteins to the human c-*fos* SRE and related sequences. *Nucleic Acids Res.* **29**, 479–487 (2001).
39. Stormo, G. D. & Zhao, Y. Determining the specificity of protein–DNA interactions. *Nature Rev. Genet.* **11**, 751–760 (2010).
40. Berger, M. F. *et al.* Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266–1276 (2008).
41. Maerkl, S. J. & Quake, S. R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **315**, 233–237 (2007).
42. Frankel, N. *et al.* Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* **474**, 598–603 (2011).
This study is one of the most detailed dissections of divergent enhancer activity to date. Divergent sites were tested individually and in combination for their effects on gene expression as well as on the divergent phenotype (loss of trichomes in *D. sechellia* larvae).
43. Jeong, S. *et al.* The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell* **132**, 783–793 (2008).
44. Prabhakar, S. *et al.* Human-specific gain of function in a developmental enhancer. *Science* **321**, 1346–1350, (2008).
45. Rebeiz, M., Jikomes, N., Kassner, V. A. & Carroll, S. B. Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proc. Natl Acad. Sci.USA* **108**, 10036–10043 (2011).
This study screened tissue-specific gene expression across multiple developmental stages among closely related species of *Drosophila* to identify novel expression patterns. Genetic changes responsible for these putatively novel enhancer activities were then identified.
46. Chan, Y. F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**, 302–305 (2010).
This study implicates recurrent deletions of fragile *cis*-regulatory DNA in the repeated loss of pelvic structures in freshwater threespine stickleback.
47. McLean, C. Y. *et al.* Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**, 216–219 (2011).
This study identifies over 500 sequences that are highly conserved among chimpanzees and other mammals but that are deleted in humans. In two cases, the deletions are confirmed to remove *cis*-regulatory sequences and to cause changes in expression that correlate with human-specific traits.
48. Kidwell, M. G. & Lisch, D. Transposable elements as sources of variation in animals and plants. *Proc. Natl Acad. Sci. USA* **94**, 7704–7711 (1997).
49. Daborn, P. J. *et al.* A single p450 allele associated with insecticide resistance in *Drosophila*. *Science* **297**, 2253–2256 (2002).
50. Chung, H. *et al. Cis*-regulatory elements in the accord retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. *Genetics* **175**, 1071–1077 (2007).
51. Schlenke, T. A. & Begun, D. J. Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proc. Natl Acad. USA* **101**, 1626–1631 (2004).
52. Bejerano, G. *et al.* A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441**, 87–90 (2006).
53. Wang, T. *et al.* Species-specific endogenous retroviruses shape the transcriptional network of the human tumour suppressor protein p53. *Proc. Natl Acad. Sci. USA* **104**, 18613–18618 (2007).
54. Jeong, S., Rokas, A. & Carroll, S. B. Regulation of body pigmentation by the abdominal-b hox protein and its gain and loss in *Drosophila* evolution. *Cell* **125**, 1387–1399 (2006).
55. He, B. Z., Holloway, A. K., Maerkl, S. J. & Kreitman, M. Does positive selection drive transcription factor binding site turnover? A test with *Drosophila cis*-regulatory modules. *PLoS Genet.* **7**, e1002053 (2011).
This study identifies derived changes in either *Drosophila melanogaster* or *D. simulans* in 645 experimentally identified binding sites for 30 different TFs across 118 CREs and infers the effect of the derived change on TF binding affinity.
56. Dowell, R. D. Transcription factor binding variation in the evolution of gene regulation. *Trends Genet.* **26**, 468–475 (2010).
57. Odom, D. T. *et al.* Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nature Genet.* **39**, 730–732 (2007).
58. Borneman, A. R. *et al.* Divergence of transcription factor binding sites across related yeast species. *Science* **317**, 815–819 (2007).
59. Tuch, B. B., Galgoczy, D. J., Hernday, A. D., Li, H. & Johnson, A. D. The evolution of combinatorial gene regulation in fungi. *PLoS Biol.* **6**, e38 (2008).
60. Schmidt, D. *et al.* Five-vertebrate ChIP–seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 (2010).
This study uses ChIP coupled with next-generation sequencing to quantify the binding of two TFs in liver cells from five vertebrate species.
61. Kasowski, M. *et al.* Variation in transcription factor binding among humans. *Science* **328**, 232–235 (2010).
62. Wilson, M. D. *et al.* Species-specific transcription in mice carrying human chromosome 21. *Science* **322**, 434–438 (2008).
63. Zheng, W., Zhao, H., Mancera, E., Steinmetz, L. M. & Snyder, M. Genetic analysis of variation in transcription factor binding in yeast. *Nature* **464**, 1187–1191 (2010).
64. Bradley, R. K. *et al.* Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related *Drosophila* species. *PLoS Biol.* **8**, e1000343 (2010).
65. Conte, C., Dastugue, B. & Vaury, C. Coupling of enhancer and insulator properties identified in two retrotransposons modulates their mutagenic impact on nearby genes. *Mol. Cell. Biol.* **22**, 1767–1777 (2002).
66. Lerman, D. N., Michalak, P., Helin, A. B., Bettencourt, B. R. & Feder, M. E. Modification of heat-shock gene expression in *Drosophila melanogaster* populations via transposable elements. *Mol. Biol. Evol.* **20**, 135–144 (2003).
67. Eichenlaub, M. P. & Ettwiller, L. *De novo* genesis of enhancers in vertebrates. *PLoS Biol.* **9**, e1001188 (2011).
68. Prud'homme, B. *et al.* Repeated morphological evolution through *cis*-regulatory changes in a pleiotropic gene. *Nature* **440**, 1050–1053 (2006).
This study identifies the enhancers as well as localizes the specific mutations within those enhancers responsible for the independent gain of wing-specific pigmentation spots in two *Drosophila* species.
69. Sumiyama, K. & Saitou, N. Loss-of-function mutation in a repressor module of human-specifically activated enhancer *HACNS1*. *Mol. Biol. Evol.* **28**, 3005–3007 (2011).
This study shows that expression appearing to result from the evolution of a novel CRE in humans was more likely caused by human-specific mutations disrupting ancestral repressive sequences masking latent *cis*-regulatory activities.
70. Rebeiz, M., Pool, J. E., Kassner, V. A., Aquadro, C. F. & Carroll, S. B. Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science* **326**, 1663–1667 (2009).
71. Babbitt, C. C. *et al.* Multiple functional variants in *cis* modulate *PDYN* expression. *Mol Biol. Evol.* **27**, 465–479 (2010).

72. Przeworski, M. The signature of positive selection at randomly chosen loci. *Genetics* **160**, 1179–1189 (2002).
73. Pool, J. E. & Aquadro, C. F. The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*. *Mol. Ecol.* **16**, 2844–2851 (2007).
74. Wittkopp, P. J. *et al.* Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in *Drosophila*. *Science* **326**, 540–544 (2009).
75. Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).
76. Orr, H. A. The genetic theory of adaptation: a brief history. *Nature Rev. Genet.* **6**, 119–127 (2005).
77. Mackay, T. F. Mutations and quantitative genetic variation: lessons from *Drosophila*. *Phil. Trans. R. Soc. B* **365**, 1229–1239 (2010).
78. Charlesworth, B. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature Rev. Genet.* **10**, 195–205 (2009).
79. Arendt, J. & Reznick, D. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.* **23**, 26–32 (2008).
80. Sucena, E., Delon, I., Jones, I., Payre, F. & Stern, D. L. Regulatory evolution of *shavenbaby*/*ovo* underlies multiple cases of morphological parallelism. *Nature* **424**, 935–938 (2003).
81. Duret, L. & Galtier, N. Comment on "human-specific gain of function in a developmental enhancer". *Science* **323**, 714; author reply 714 (2009).
82. Houle, D., Govindaraju, D. R. & Omholt, S. Phenomics: the next challenge. *Nature Rev. Genet.* **11**, 855–866 (2010).
83. Hoekstra, H. E. & Coyne, J. A. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995–1016 (2007).
84. Lynch, V. J. & Wagner, G. P. Resurrecting the role of transcription factor change in developmental evolution. *Evolution* **62**, 2131–2154 (2008).
85. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nature Genet.* **40**, 346–350 (2008).
86. Emerson, J. J. *et al.* Natural selection on *cis* and *trans* regulation in yeasts. *Genome Res.* **20**, 826–836 (2010).
87. Cretekos, C. J. *et al.* Regulatory divergence modifies limb length between mammals. *Genes Dev.* **22**, 141–151 (2008).
88. Reed, R. D. *et al.* Optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* **333**, 1137–1141 (2011).
89. Gomez-Skarmeta, J. L., Lenhard, B. & Becker, T. S. New technologies, new findings, and new concepts in the study of vertebrate *cis*-regulatory sequences. *Dev. Dyn.* **235**, 870–885 (2006).
90. Richards, S. *et al.* Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and *cis*-element evolution. *Genome Res.* **15**, 1–18 (2005).
91. Peterson, B. K. *et al.* Big genomes facilitate the comparative identification of regulatory elements. *PLoS ONE* **4**, e4688 (2009).
92. Wittkopp, P. J. Evolution of *cis*-regulatory sequence and function in Diptera. *Heredity* **97**, 139–147 (2006).
93. Weirauch, M. T. & Hughes, T. R. Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. *Trends Genet.* **26**, 66–74 (2010).
94. Meireles-Filho, A. C. & Stark, A. Comparative genomics of gene regulation-conservation and divergence of *cis*-regulatory information. *Curr. Opin. Genet. Dev.* **19**, 565–570 (2009).
95. Ludwig, M. Z., Bergman, C., Patel, N. H. & Kreitman, M. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**, 564–567 (2000).
96. Ludwig, M. Z. *et al.* Functional evolution of a cis-regulatory module. *PLoS Biol.* **3**, e93 (2005).
97. Piano, F., Parisi, M. J., Karess, R. & Kambysellis, M. P. Evidence for redundancy but not *trans* factor—*cis* element co-evolution in the regulation of *Drosophila Yp* genes. *Genetics* **152**, 605–616, (1999).
98. Wittkopp, P. J., Vaccaro, K. & Carroll, S. B. Evolution of yellow gene regulation and pigmentation in *Drosophila*. *Curr. Biol.* **12**, 1547–1556 (2002).
99. Swanson, C. I., Schwimmer, D. B. & Barolo, S. Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr. Biol.* **21**, 1186–1196 (2011).
100. Arnone, M. I., Dmochowski, I. J. & Gache, C. Using reporter genes to study *cis*-regulatory elements. *Methods Cell Biol.* **74**, 621–652 (2004).
101. O'Doherty, A. *et al.* An aneuploid mouse strain carrying human chromosome 21 with down syndrome phenotypes. *Science* **309**, 2033–2037 (2005).
102. Berman, B. P. *et al.* Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Genome Biol.* **5**, R61 (2004).

Acknowledgements
This research was supported by National Science Foundation grant MCB-1021398 to P.J.W., who is an Alfred P. Sloan Research Fellow.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Patricia J. Wittkopp's homepage: http://www.umich.edu/~pwlab

SUPPLEMENTARY INFORMATION
See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Online summary

- *Cis*-regulatory elements (CREs) have essential roles in development, and their divergence is a common cause of evolutionary change.
- Recent technological advances facilitate finding and studying orthologous CREs.
- Functional divergence of CREs has been shown to result from small numbers of single-nucleotide substitutions, insertions and/or deletions that disrupt transcription factor binding and often interact epistatically.
- Novel enhancer activities appear frequently to result from the co-option of transcription factor binding sites that are already present in ancestral enhancers.
- Elucidating the specific genetic changes that are responsible for *cis*-regulatory divergence has provided insight into the relative contributions of new mutations and standing genetic variation, large and small effect mutations and neutral and non-neutral changes.
- Studies of *cis*-regulatory divergence that underlie convergent phenotypes address questions about the repeatability of evolution.
- Future work should focus on elucidating the relationships among sequence changes in CREs, transcription factor binding, gene expression and organismal phenotypes.

Author biographies

Patricia J. Wittkopp is an associate professor in the Departments of Ecology and Evolutionary Biology and Molecular, Cellular, and Developmental Biology at the University of Michigan, Ann Arbor, USA. After earning her B.S. at the University of Michigan, where she conducted research with Greg Gibson, she earned a Ph.D. in genetics at the University of Wisconsin, Madison, USA, working with Sean B. Carroll. She then worked with Andrew G. Clark at Cornell University, Ithaca, New York, USA, as a postdoctoral fellow of the Damon Runyon Cancer Research Foundation. Her laboratory is broadly interested in evolutionary genetics and is currently investigating the evolution of pigmentation among *Drosophila* species and the evolution of gene regulation using both fruit flies and yeast.

Gizem Kalay is a Ph.D. student in the department of Molecular, Cellular, and Developmental Biology at the University of Michigan. Before starting the Ph.D. program, she studied evolution and development with Matthias Gerberding at the Max Planck Institute in Tübingen, Germany. Her dissertation focuses on the evolution of *cis*-regulatory sequences controlling expression of the *Drosophila yellow* gene.

ToC Blurb

SERIES ON REGULATORY ELEMENTS

000 **Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence**

Patricia J. Wittkopp and Gizem Kalay

Changes in *cis*-regulatory sequences, especially enhancers, make a substantial contribution to phenotypic diversity among and within species. Recent studies have begun to uncover the mechanisms underlying *cis*-regulatory divergence, such as the types of mutations that occur and their influence on transcription factor binding.

Subject categories

gene regulation; evolution; gene expression; developmental biology