3 *cis*- and *trans*-Regulation in *Drosophila* Interspecific Hybrids

Joseph D. Coolon and Patricia J. Wittkopp

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, USA

Introduction

The regulation of gene expression is essential for organismal form, function, and fitness. Complex organismal development from zygote to adult requires finely tuned expression of genes in space, time, and abundance. Proper expression of genes is also required for most other aspects of physiology and function. Even minor changes in gene expression have the potential to significantly impact phenotypes (e.g., Oleksiak et al., 2005; Cooley et al., 2012). Despite the central role that gene expression plays in converting genotypes into phenotypes, abundant variation in gene expression exists within and between species. While not all of this variation is expected to impact phenotypes, changes in gene expression have been shown to contribute to the evolution of diverse traits, including bat wings (Cretekos et al., 2008), turtle shells (Moustakas, 2008), stickleback spines (Shapiro et al., 2004; Chan et al., 2010), corn kernels (Clark et al., 2006), beak morphology in Darwin's finches (Abzhanov et al., 2004), and body color differences within and between species of *Drosophila*, butterfly, and mice (Wittkopp et al., 2003; Kronforst et al., 2012). Genetic changes underlying interspecific expression differences may also contribute to speciation by causing hybrid incompatibilities (e.g., Maheshwari & Barbash, 2012).

Gene expression begins with transcription, and the molecular mechanisms controlling when, where, and how frequently a DNA sequence encoding a gene product is transcribed into RNA are much better understood than those responsible for posttranscriptional regulation. Fundamentally, transcription is controlled by interactions between *cis*-regulatory DNA and *trans*-acting RNA and proteins. *cis*-Regulatory sequences include the basal promoter located 5′ of the transcription start site of each gene upon which the multiprotein RNA polymerase holoenzyme complex of *trans*-acting factors assembles. Alone, this complex produces "basal" levels of RNA transcripts. These levels are often modified in multicellular eukaryotes by additional *cis*-regulatory elements located in noncoding sequences, 5′ or 3′ of the transcription start site, which are known as enhancers. These enhancers bind to sets of *trans*-acting transcription factors that act combinatorially to enhance or repress transcription. Interactions between transcription factor complexes assembled on enhancers and on the basal promoter occur through changes in DNA shape (looping) and/or other molecular mechanisms (reviewed in Wray et al., 2003) (Figure 3.1A). Chromatin structure and DNA sequences

Polyploid and Hybrid Genomics, First Edition. Edited by Z. Jeffrey Chen and James A. Birchler.

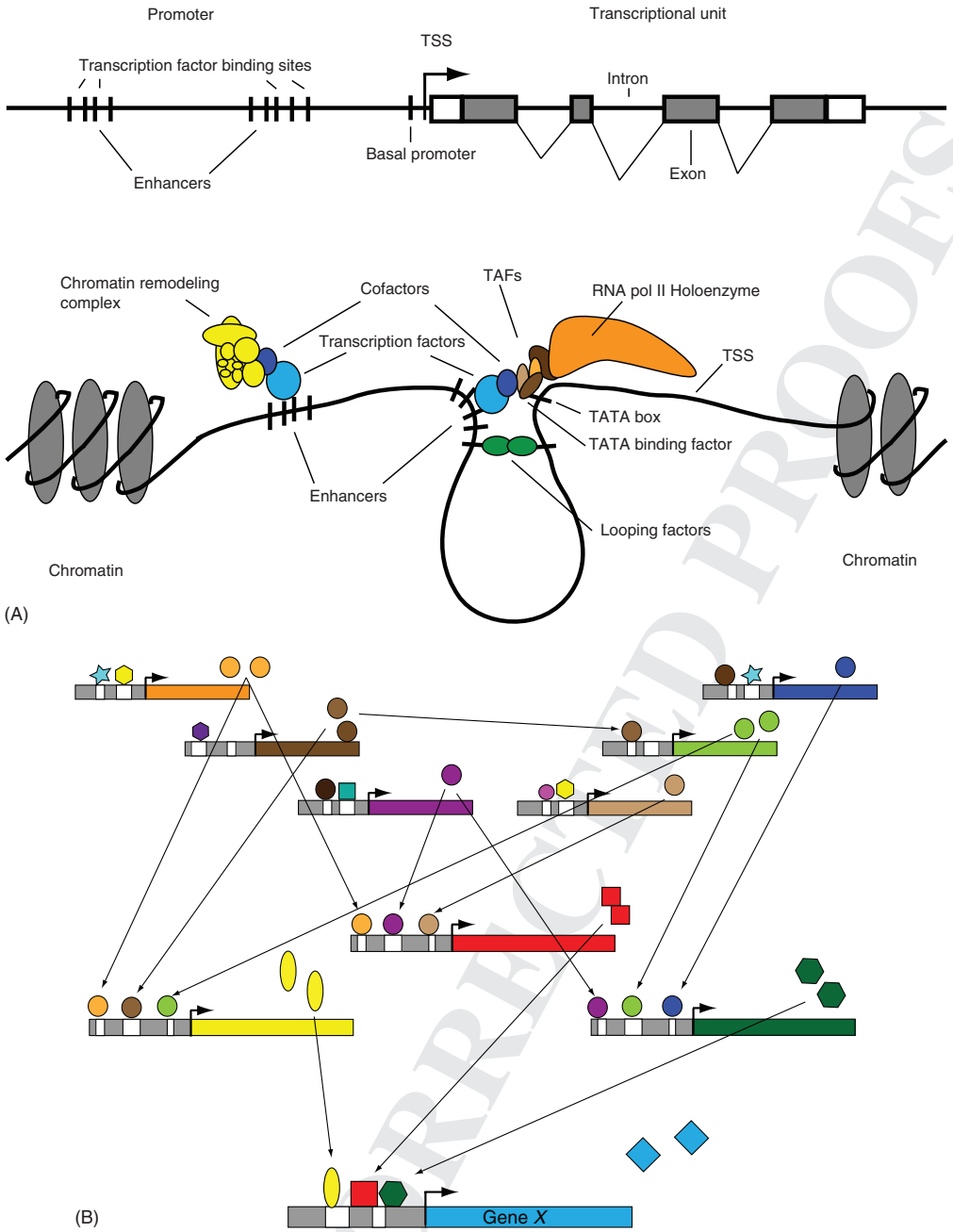

Plate 3.1 Model of transcription in the context of regulatory networks. (A) A gene with key features highlighted including enhancers containing transcription factor binding sites, promoter, exons and introns and below showing how the transcriptional machinery (e.g., chromatin remodeling complex, TATA-associated factors (TAFs), looping factors, and RNA pol II holoenzyme) assembles at a transcription start site (TSS). Modified from Wray et al. (2003). (B) Hypothetical regulatory network involved in the expression of gene *X*. Each gene has enhancer (white) and coding sequence (colored). Arrows indicate *trans*-acting regulation in relation to gene *X*.

Polyploid and Hybrid Genomics, First Edition. Edited by Z. Jeffrey Chen and James A. Birchler.

located in the 5′- and 3′-untranslated regions of a gene can also have *cis*-acting (i.e., allele-specific) effects on RNA transcript abundance. Similarly, other RNAs, proteins, and even the environment can influence the availability or activity of transcription factors and thus have (indirect) *trans*-acting effects on transcription. Together, all these interactions between *cis*- and *trans*-acting factors create complex networks that regulate transcription (Figure 3.1B).

In this chapter, we primarily review studies that have used interspecific F_1 hybrids between two *Drosophila* species to investigate *cis*- and *trans*-regulation of transcription (hereafter referred to simply as gene expression). (Chapters 2, 5, 7, and 8 discuss related topics in *Drosophila*, maize, and yeast.) *Drosophila* is particularly well suited for these studies because of the extensive genomic resources and the ability of many closely related species to form viable hybrids (Lee & Watanabe, 1987; Sawamura et al., 1993). More specifically, we describe the logic and methods used to distinguish between *cis*- and *trans*-regulatory changes based on expression in interspecific hybrids, catalog studies that have examined *cis*- and *trans*-regulation in interspecific hybrids of *Drosophila*, and integrate findings from these studies to address a number of questions about regulatory evolution. We close by describing open questions and future directions for this field.

Distinguishing between *cis*- and *trans*-Regulatory Changes Using eQTL, GWAS, and ASE

As described above, genetic changes affecting gene expression can be classified as either *cis*- or *trans*-regulatory depending on how they exert their effects on the gene of interest. *cis*-Regulatory DNA sequences tend to be located close to the gene they affect, whereas sequences with *trans*-regulatory effects on gene expression can be located anywhere in the genome. More definitively, *cis*-regulatory sequences affect expression of only the allele of a gene that is located on the same chromosome whereas *trans*-regulatory factors have the potential to affect expression of all alleles of a gene within a cell. Both of these differences have been used to distinguish between *cis*- and *trans*-regulatory changes on a genomic scale.

The first high-throughput method developed for disentangling *cis*- and *trans*-regulatory changes was based on the distance between a variable site and the gene whose expression it affects, and is known as expression quantitative trait locus (eQTL) mapping (Brem et al., 2002). Briefly, a difference in expression between two genotypes is treated as a quantitative phenotype and analyzed in a recombinant mapping population; correlations between genotypes and expression level phenotypes are then used to identify regions of the genome that contribute to a change in the target gene's expression. *cis*-/*trans*-Acting eQTL are identified using a physical distance cutoff measured in base pairs: eQTL located further from the affected gene than this distance are assumed to be *trans*-acting, whereas eQTL located closer to the affected gene than this distance are assumed to be *cis*-acting (reviewed in Gibson & Weir, 2005; Li & Burmeister, 2005; Ranz & Machado, 2006). This method has been used to examine the genetic basis of expression differences among strains of yeast, mice, plants, and other species (Brem et al., 2002; Schadt et al., 2003; West et al., 2007), but rarely been used in *Drosophila* (but see Ruden et al., 2009). A related approach, using genome-wide association studies (GWAS) to look for correlations between polymorphic sites and expression levels, is used to study variable gene expression in human populations (Dixon et al., 2007). These distance-based methods can misclassify eQTL with long-range *cis*-regulatory effects and *trans*-acting eQTL that happen to be located close to the affected gene, but are generally reliable (Babak et al., 2010) and have the advantage of identifying the specific regions of the genome responsible for the observed *cis*- and *trans*-regulatory effects as well as the potential to test for additive and epistatic interactions

among them. QTL mapping and GWAS can rarely be used to study expression differences between species, however, because the sterility that often results from hybrid incompatibilities can prevent recovery of recombinant individuals.

An alternative way to distinguish between *cis*- and *trans*-regulatory effects utilizes measures of allele-specific expression (ASE) in a pair of inbred genotypes and F_1 hybrids produced by crossing them together. This approach can be used to study the molecular mechanisms underlying expression differences within a species as well as between any species for which F_1 hybrids (either fertile or infertile) can be produced. In essence, relative *cis*-regulatory activity between two alleles is compared by assaying their allele-specific transcription in a shared *trans*-regulatory environment. This shared *trans*-regulatory environment is produced simply by crossing two genotypes with the *cis*-regulatory alleles of interest together, which puts the two *cis*-regulatory alleles into the same cell. Differences in expression between the two alleles indicate differences in relative *cis*-regulatory activity (Cowles et al., 2002), and if these differences are insufficient to account for the expression difference observed between the original inbred genotypes, *trans*-regulatory changes are inferred (Wittkopp et al., 2004). These relationships can be represented as $P = C + T$, where P is the relative expression of a gene of interest between two inbred "parental" genotypes crossed to produce the F_1 hybrid, C is the measure of relative *cis*-regulatory activity provided by relative ASE in the F_1 hybrid, and T is a measure of *trans*-regulatory differences between the parental genotypes affecting expression of the focal gene, inferred as $P - C$ (Figure 3.2). This approach is more accurate than eQTL mapping and GWAS for classifying regulatory changes as *cis*- or *trans*-acting because it relies on the functional differences between *cis*- and *trans*-regulatory sequences rather than an arbitrary distance between the site and the affected gene, but it measures only the net effects of all relevant *cis*- and *trans*-regulatory differences and does not identify the genomic location of any individual change. Measures of ASE in F_1 hybrids have been used to investigate regulatory polymorphisms within species as well as divergence between species of plants (de Meaux et al., 2005; Guo et al., 2008; Zhang & Borevitz, 2009), fungi (Tirosh et al., 2009; Emerson et al., 2010), and animals (Wittkopp et al., 2004; Krishna Pant et al., 2006; Lawniczak et al., 2008; Serre et al., 2008; Wilson

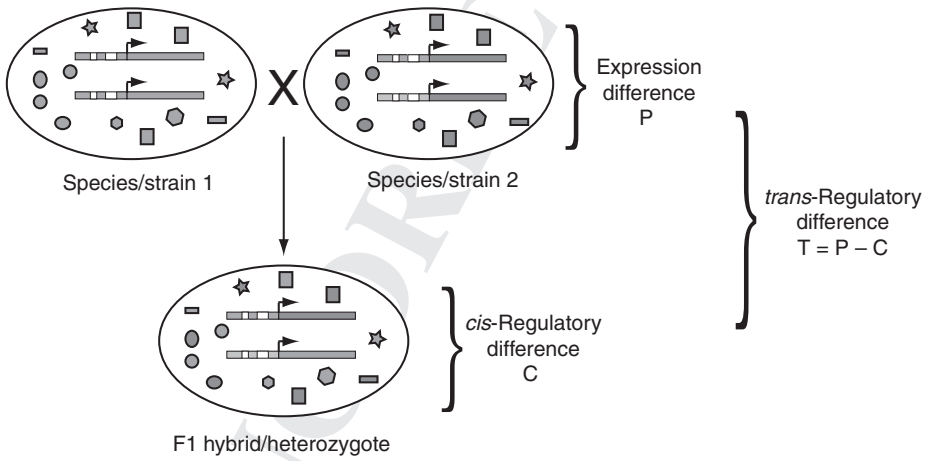

Figure 3.2 Experimental design to use ASE to determine *cis*- and *trans*-regulatory divergence in diploid organisms. Two species or strains are crossed to produce F1 hybrid offspring with a common *trans*-regulatory environment. *trans*-Acting molecules are indicated by colored shapes. Parental expression difference (P), *cis*-regulatory difference (C), and *trans*-regulatory differences are measured as indicated. (For a color version of this figure, see Plate 3.2.)

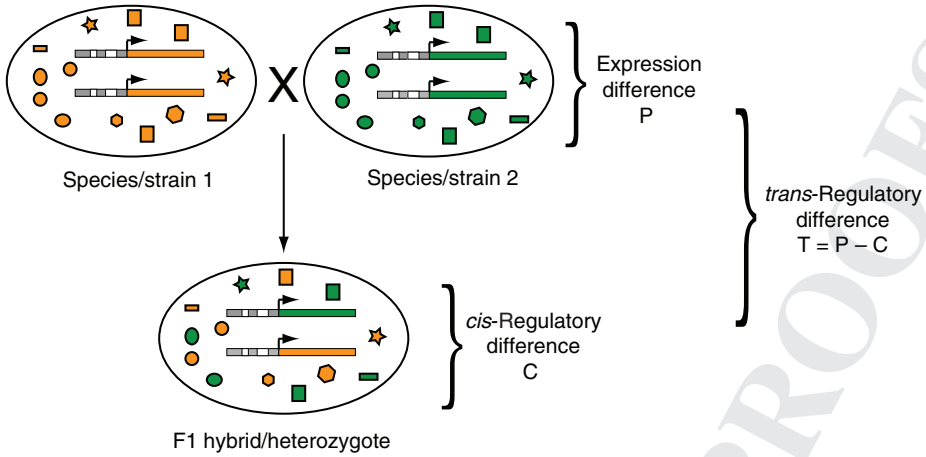

Plate 3.2 Experimental design to use ASE to determine *cis*- and *trans*-regulatory divergence in diploid organisms. Two species or strains are crossed to produce F1 hybrid offspring with a common *trans*-regulatory environment. *trans*-Acting molecules are indicated by colored shapes. Parental expression difference (P), *cis*-regulatory difference (C), and *trans*-regulatory differences are measured as indicated.

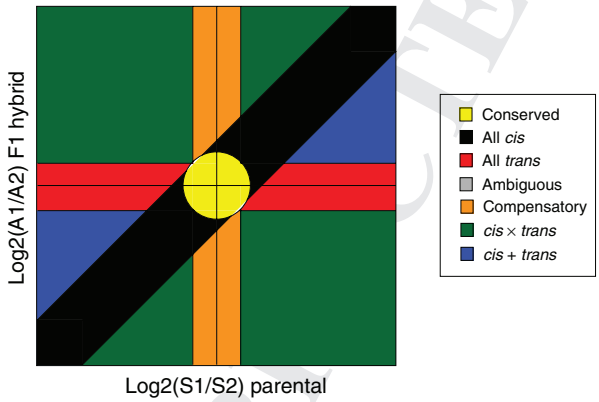

Plate 3.3 Plot of hypothetical results from an ASE study. The results of a hierarchical series of statistical tests of ASE from parental, and F1 hybrid samples. Log2-transformed parental expression ratio is plotted on the *x*-axis and log2-transformed F1 hybrid allelic expression ratio is plotted on the *y*-axis. Regions are color-coded based on classifications that would exist in each location in the plot.

et al., 2008; Main et al., 2009). As described above, this chapter focuses specifically on *cis*- and *trans*-regulatory changes observed using interspecific hybrids of *Drosophila* species.

Methods Used to Quantify ASE

Before the development of high-throughput genotyping techniques, differences in ASE between species were identified in *Drosophila* by creating pairs of transgenes with *cis*-regulatory alleles from different species and transforming them into a common host species to provide a common *trans*-regulatory environment (reviewed by Wittkopp, 2006). Emphasis was on the spatial and temporal expression patterns of divergent alleles rather than the quantitative abundance of RNA transcripts. These experiments were labor-intensive and it often took years to analyze divergent expression for even one gene. In the early 2000s, researchers began to adapt methods developed for genotyping genomic DNA (gDNA) for genotyping cDNA derived from heterozygous RNA samples, allowing levels of ASE to be quantified. These methods work by determining the relative frequency of alternative alleles at one or more polymorphic sites located in the transcribed region of a gene. The sites genotyped are not usually the cause of any *cis*-regulatory differences observed, but rather provide a way to recognize which cDNA molecules were derived from which allele of the gene.

One of the first methods used to detect *cis*-regulatory differences using ASE was single-base extension sequencing. With this technique, a transcribed region surrounding a single-nucleotide polymorphism (SNP) is amplified, a standard primer is annealed to the amplified products that hybridize immediately 5′ of the polymorphic site, fluorescently labeled single nucleotides that terminate the nucleotide chain are added, and the identity of the nucleotide incorporated is determined using a machine designed for Sanger dye-terminator sequencing (Cowles et al., 2002). Relative incorporation of alternative nucleotides (reflecting ASE) is assessed by comparing allele-specific peaks in the chromatogram produced by the sequencing machine.

A related, but more quantitative, technique is pyrosequencing (Ahmadian et al., 2000). This method also starts by amplifying a transcribed region containing a polymorphic site, annealing a standard primer 5′ of the polymorphic site, and extending the primer through the polymorphic site. Differences from the single-base extension method are as follows:

(1) Individual nucleotides (A, G, T, or C) are added one at a time.
(2) Incorporation of specific nucleotides at specific positions in the sequence is determined by the amount of light produced by a cascade of enzymatic reactions.
(3) Primer extension can proceed for ten or more base pairs.
(4) One primer used for PCR amplification must be biotinylated (increasing cost).
(5) A specialized machine (e.g., Qiagen PSQ96) is required to analyze each pyrosequencing reaction.

Pyrosequencing measurements of ASE are highly reproducible and allow expression differences as small as 10% to be detected as significant (Wittkopp, 2011).

Quantitative PCR (qPCR) using allele-specific primers or fluorescently labeled probes has also been used to measure ASE for small numbers of genes (e.g., Singer-Sam & Gao, 2001; Ronald et al., 2005), but has not been employed to study *cis*- and *trans*-regulation on a large scale. This is perhaps because it is difficult to generate assays that are truly allele-specific. Another drawback to this approach, as well as to the single-base extension and pyrosequencing methods

described above, is that unique assays must be designed and finely tuned for each gene. This limits the number of genes that can be reasonably analyzed with these types of techniques to tens to hundreds.

To test for differences in *cis*- and *trans*-regulation on a genomic scale, researchers first turned to DNA microarrays. Microarrays measure gene expression by hybridizing fluorescently labeled cDNA samples to DNA sequences arrayed in a grid, and are typically used to measure total (i.e., not allele-specific) expression of a gene (Gibson, 2002). The number of DNA molecules bound to each DNA sequence on the microarray is determined by the intensity of its fluorescent signal. This signal can be influenced by the sequence composition of the DNA spotted as well as hybridization of cDNAs from multiple genes with similar sequences. The potential for cross-hybridization is particularly high for alternative alleles of the same gene. For the most reliable results, custom microarrays should be developed for the particular strains and species used containing allele-specific probes (e.g., Tirosh et al., 2009; Yang et al., 2011), but ASE has also been estimated from standard microarrays using complex statistical analyses (Graze et al., 2009). The time and cost associated with developing allele-specific arrays for each pair of genotypes to be analyzed limits the applications of this approach.

Recently, an alternative to microarrays for quantifying expression genome-wide has emerged that takes advantage of the rapidly falling cost of massively parallel DNA sequencing and has a greater dynamic range than microarrays (Nagalakshmi et al., 2008; Marioni et al., 2008). This approach, known as "RNA-seq," is simply the brute-force, shot-gun sequencing of a cDNA pool. After generating tens of millions of short (30–150 bp) sequences, computational tools are used to map each read back to a specific gene and a specific allele of that gene. Mapping sequencing reads to the correct gene is much easier than mapping them to the correct allele (e.g., Degner et al., 2009), but ASE can be reliably inferred from RNA-seq data when the sequence of the alternative alleles is known and appropriate mapping algorithms are used (e.g., Emerson et al., 2010; McManus et al., 2010; Graze et al., 2012). Because the methods for inferring ASE from RNA-seq data are still being refined, validation using an independent technique such as those listed above remains critical (Deveale et al., 2012). Nevertheless, RNA-seq has the potential to be used to measure ASE on a genomic scale for any gene with one or more transcribed polymorphic sites for nearly any species. Targeted RNA-seq, in which the cDNA sample is limited to transcripts from a few genes, is also possible (Main et al., 2009).

An important consideration for measuring ASE using any of the techniques described above is controlling for the effects of sequence differences between alleles on the measurement of allelic abundance. This is essential for establishing an accurate baseline that corresponds to equal expression of the two alleles. Perhaps the simplest way to do this is to analyze gDNA from F_1 hybrids in addition to the cDNA samples from parental genotypes and hybrid progeny. Assuming that the parental genotypes were highly inbred and homozygous at nearly all loci, F_1 females inherit exactly one allele from each parent at all loci. (The same is true for autosomal genes in F_1 hybrid males.) Deviations from equal representation of the two alleles in F_1 hybrid gDNA indicate an experimental bias that must be taken into account when testing for differential ASE. A variety of statistical tests, either comparing allelic abundance between cDNA and gDNA samples or comparing cDNA measures that have been corrected for bias detected using gDNA to a null expectation, have been used to identify statistically significant differences in total and ASE. These include *t*-tests and their nonparametric equivalent (Wittkopp et al., 2004), mixed linear models (Landry et al., 2005; Wittkopp et al., 2008; Graze et al., 2009), binomial and Fisher's exact tests (Fontanillas et al., 2010; McManus et al., 2010), and Bayesian methods (Graze et al., 2012).

Studies of *cis*- and *trans*-Regulation in Interspecific Hybrids of *Drosophila*

Since 2004, seven studies have been published that analyzed *cis*- and/or *trans*-regulatory changes for at least 10 genes in at least two *Drosophila* species. In each case, *D. melanogaster* was compared to either *D. simulans* or *D. sechellia*, both of which are estimated to have diverged from *D. melanogaster* 1–2 million years ago (Cutter, 2008). Before discussing the insights into regulatory evolution afforded by these studies, we provide a brief overview of the experimental design and biological samples used in each study (Table 3.1):

(1) The first study using ASE to investigate *cis*- and *trans*-regulatory changes in *Drosophila* interspecific hybrids was conducted by Wittkopp et al. (2004) and used pyrosequencing to analyze 34 genes. The genes included in this study were chosen because they were previously shown to have significant expression differences between *D. melanogaster* and *D. simulans* (Rifkin et al., 2003), albeit at different developmental stages and in different strains than those analyzed. ASE was measured in mixed pools of 1-day-old adult virgin female (whole fly) *D. melanogaster* (zygotic hybrid rescue, zhr) and *D. simulans* (tsimbazaza), as well as in F_1 hybrids from reciprocal crosses (i.e., *D. melanogaster* × *D. simulans* and *D. simulans* × *D. melanogaster*). Prior to this work, studies of the regulatory changes underlying expression differences between *Drosophila* species had focused on single genes and used arduous transgenic techniques. This study paved the way for more systematic investigations of *cis*- and *trans*-regulatory changes underlying divergent gene expression in *Drosophila* and other species.

(2) Using similar methods to those described in Wittkopp et al. (2004), Landry et al. (2005) investigated the sources of abnormal gene expression commonly observed in interspecific hybrids of *Drosophila* (e.g., Michalak & Noor, 2003; Ranz et al., 2004). Genes were selected for this analysis because they were previously shown to be misexpressed in F_1 hybrids between *D. melanogaster* and *D. simulans* (Ranz et al., 2004), or because they were found by Wittkopp et al. (2004) to have patterns of *cis*- and *trans*-regulatory changes consistent with those hypothesized to be more likely to lead to misexpression in hybrids. Because these genes were chosen from two different studies, allele-specific and total gene expressions were analyzed in two different sets of parents and F_1 hybrids. The first cross was between *D. melanogaster* Canton-S and *D. simulans* Sim1 (as in Ranz et al., 2004), and 23 genes were tested for *cis*- and *trans*-regulatory changes. The second cross was between *D. melanogaster* zhr and *D. simulans* tsimbazaza (as in Wittkopp et al., 2004), and 8 genes were tested for *cis*- and *trans*-regulatory changes. RNA samples from 3- to 5-day-old adult virgin female (whole fly) *D. melanogaster*, *D. simulans*, and F_1 hybrids produced by crossing *D. melanogaster* females with *D. simulans* males were analyzed in each case. Relative ASE was measured using pyrosequencing and total expression levels were measured using qPCR.

(3) To help infer the evolutionary processes that gave rise to divergent *cis*- and *trans*-regulation between species, Wittkopp et al. (2008) compared the genetic basis of variable gene expression within and between *Drosophila* species. Eight intraspecific crosses (five crosses between five strains of *D. melanogaster* and three crosses between three strains of *D. simulans*) and seven interspecific crosses between *D. melanogaster* and *D. simulans* strains were performed and analyzed. The biological samples assayed in this study were 7- to 10-day-old adult mated females (whole flies), with the exception of one cross (*D. simulans* tsimbazaza females × *D. melanogaster* zhr males) where an additional developmental time point (3- to 5-day-old adult mated female whole flies) was also assayed. ASE was quantified for a total of 75 genes,

Table 3.1 Details of ASE studies in *Drosophila* interspecific hybrids

Study	Method	Species 1	Species 2	Crosses	Age	Tissue	Sex	Mated	Replicates	Reciprocal hybrid	Statistical analyses
Wittkopp et al. (2004)	Pyrosequencing	*D. melanogaster (zhr)*	*D. simulans (tsimbazaza)*	1 + R	1-day-old adult	Whole fly	Female	N	Y	Y	*t*-test, Mann–Whitney *U*
Landry et al. (2005)	Pyrosequencing	*D. melanogaster (zhr, Canton-S)*	*D. simulans (tsimbazaza, Sim1)*	2	3- to 5-day-old adult	Whole fly	Female	N	Y	N	Mixed linear model
Wittkopp et al. (2008)	Pyrosequencing	*D. melanogaster*	*D. simulans*	7 (8 + R)	7- to 10-day-old (3–5) adult	Whole fly	Female	Y	Y	Y	Mixed linear model, Mann–Whitney *U*
Graze et al. (2009)	Microarray	*D. melanogaster (dm3)*	*D. simulans (C167.4, w501)*	2	5- to 7-day-old adult	Head	Female	N	Y	N	Mixed linear model
Fontanillas et al. (2010)	454 RNA-seq	*D. melanogaster (Canton-S)*	*D. simulans (C167.4)*	1	5- to 6-day-old adult	Whole fly	Female	N	N	N	Binomial exact test
McManus et al. (2010)	Illumina RNA-seq	*D. melanogaster (dm3)*	*D. sechellia (droSec1)*	1	2-day-old adult	Whole fly	Female	N	N	N	Binomial and Fisher's exact tests
Graze et al. (2012)	Illumina RNA-seq	*D. melanogaster (Berlin)*	*D. simulans (C167.4)*	1	1- to 1.5-day-old adult	Head	Female	N	Y	N	Bayesian

Details from each study (listed in the first column) are shown including the methods and statistical analyses used, strains and species samples, age, stage, tissue, and sex of flies included.

with 45 genes assayed within a species, 49 assayed between species, and 16 genes analyzed both within and between species. The genes analyzed were selected using prior studies of expression differences within and between *Drosophila* species (Ranz et al., 2003; Gibson et al., 2004; Wittkopp et al., 2004) and the presence of single SNPs suitable for recognizing allele-specific transcripts. As before, pyrosequencing was used to measure ASE in pools of flies containing both parental genotypes as well as in pools of F_1 hybrids made by crossing the two parental genotypes. For all within species comparisons, F_1 hybrids were generated and analyzed from reciprocal crosses.

(4) Graze et al. (2009) investigated *cis*- and *trans*-regulatory changes between species on a genomic scale to test the generality of patterns observed in prior allelic-expression studies of *Drosophila* regulatory evolution. Unlike the studies described above, which all examined whole adult flies, only heads from 5- to 7-day-old adult virgin females were used for expression analyses in this study. Microarrays containing probes tiled across the *D. melanogaster* genome (Drosophila Affymetrix GeneChip Tiling 1.0R Arrays) were used to measure allele-specific and total expression levels in *D. melanogaster* (dm3), *D. simulans* (C167.4 and w501), mixed pools of *D. melanogaster* and *D. simulans*, and F_1 hybrids produced by crossing *D. melanogaster* females with *D. simulans* males. Statistical tests showed little difference between the C167.4 and w501 strains of *D. simulans* as well as between the interspecific hybrids produced by crossing *D. melanogaster* with each of these strains; thus, these data sets were treated as single *D. simulans* and F_1 hybrid samples, respectively. A total of 14,027 genes were represented on the tiling microarrays, but only 2,381 genes passed the stringent filters used to identify genes most likely to have ASE measured accurately.

(5) The first study using RNA-seq to quantify ASE in interspecific hybrids of *Drosophila* was carried out by Fontanillas et al. (2010). They used a massively parallel sequencing platform developed by 454 Life Sciences to analyze 5- to –6-day-old adult virgin female (whole fly) F_1 hybrid offspring from a cross between *D. melanogaster* (Canton-S) females and *D. simulans* (C167.4) males. The 454 sequencing generated 36,855 sequencing reads with an average length of 170 bases, allowing 891 genes to be tested for evidence of *cis*-regulatory divergence in F_1 hybrids. (The parental species were not examined, so *trans*-regulatory changes could not be inferred.) At the time of this publication, there were many questions about the parameters involved in designing a good experiment using high-throughput sequencing to quantify ASE, and this study includes mathematical modeling and computer simulations that show how various parameters influence the power of statistical tests for differences in ASE.

(6) Later that same year, another study using RNA-seq to infer ASE in interspecific hybrids of *Drosophila* was published. In it, McManus et al. (2010) used Illumina (formerly Solexa) sequencing to quantify total and ASE in 2-day-old virgin adult female (whole fly) *D. melanogaster* (dm3), *D. sechellia* (droSec1), a mixed parental pool of *D. melanogaster* and *D. sechellia*, and F_1 hybrids made by crossing *D. melanogaster* females with *D. sechellia* males. At least 13 million, 37 bp, paired end-sequencing reads were used to quantify gene expression in each sample. Only reads that mapped to constitutive exons (i.e., exons included in all known RNA isoforms of a gene) and genes with at least 20 allele-specific reads were used for expression analysis. This resulted in 9966 genes being tested for divergence in total expression, *cis*-regulation, and *trans*-regulation.

(7) Graze et al. (2012) used high-throughput sequencing to quantify ASE specifically in heads of *Drosophila* interspecific hybrids. The 1- or 1.5–day-old virgin female F_1 hybrids analyzed were produced by crossing *D. melanogaster* (Berlin) and *D. simulans* (C167.4). Over 128 million, 54 bp, paired-end Illumina sequencing reads from cDNA pools constructed from

RNA extracted from replicate F_1 hybrid heads were obtained. In addition, nearly 142 million, 36 bp, paired-end, Illumina sequencing reads from gDNA were also collected and used as a prior in a Bayesian analysis designed to detect differential ASE. Only genes with at least 100 allele-specific reads in both cDNA and gDNA were included in the final analysis, resulting in 6369 genes being tested for evidence of *cis*-regulatory divergence.

Insights into Regulatory Evolution

Studying gene expression in interspecific *Drosophila* hybrids, and comparing it to expression in the parental species crossed to produce the hybrid, has provided insight into questions such as the following:

(1) What is the relative frequency of *cis*- and *trans*-regulatory changes between species?
(2) How do *cis*- and *trans*-regulatory changes relate to the inheritance of gene expression?
(3) How much do neutral and nonneutral processes contribute to *cis*- and *trans*-regulatory divergence among species?
(4) How is *cis*-regulatory divergence related to local sequence evolution?
(5) What are the phenotypic consequences of *cis*- and *trans*-regulatory divergence?

In what follows, we synthesize data from studies of *Drosophila* interspecific hybrids that speak to each of these questions.

What Is the Relative Frequency of cis- *and* trans-*Regulatory Changes between Species?*

Studies examining *cis*- and *trans*-regulatory divergence using interspecific hybrids of *Drosophila* have thus far been limited to comparisons of *D. melanogaster* with *D. simulans* (six studies) and *D. sechellia* (one study). In all the comparisons between *D. melanogaster* and *D. simulans*, a greater proportion of genes showed significant evidence of *cis*-regulatory divergence than *trans*-regulatory divergence (Table 3.2). This is remarkable given that these studies differed in the number of genes analyzed, the techniques used to measure ASE, and the statistical tests used to identify significant changes in expression (Table 3.1). The comparison of *D. melanogaster* to *D. sechellia*,

Table 3.2 Proportions of genes in different regulatory categories from studies of *Drosophila* interspecific hybrids

Study	Total genes	Parental difference	*cis*	*trans*	Both
Wittkopp et al. (2004)	34	29 (85%)	30 (88%)	18 (53%)	18 (53%)
Landry et al. (2005)	31	NR	26 (84%)	23 (74%)	20 (65%)
Wittkopp et al. (2008)	49	42 (86%)	42 (86%)	38 (78%)	32 (65%)
Graze et al. (2009)	2381	359 (15%)	650 (27%)	390 (16%)	93 (4%)
Fontanillas et al. (2010)	891	NA	107 (12%)	NA	NA
McManus et al. (2010)	9966	7739 (78%)	5042 (51%)	6546 (66%)	3473 (35%)
Graze et al. (2012)	6369	NA	2866 (45%)	NA	NA

Results from each study (listed in the first column) are shown including the number of genes investigated, and number (and percentage) of genes with statistically significant evidence of parental difference, *cis*- and/or *trans*-regulatory difference or both.

however, showed more genes with evidence of *trans*-regulatory changes than *cis*-regulatory changes (66% vs. 51%). It remains to be seen whether this difference is a result of the method used or the species examined. In all cases, comparing the frequency of genes with *cis*- and *trans*-regulatory changes must be done with caution; however, the statistical tests used to identify *cis*-regulatory changes tend to have greater power than the tests used to identify *trans*-regulatory changes because the former compares the measurement of relative allelic expression in F_1 hybrids to a fixed value, whereas the latter compares the measurement of relative allelic expression in F_1 hybrids to a measurement of relative expression between species.

The exact percentage of genes showing evidence of *cis*-regulatory divergence varied among studies, ranging from 12% to 88%. At the high end of this range are Wittkopp et al. (2004), Landry et al. (2005), and Wittkopp et al. (2008) (with 88, 84, and 86%, respectively), all of which examined 30–50 genes that were chosen in part because they were likely to have divergent expression. Inclusion of some genes in more than one of these three studies might have contributed to their consistent findings. In the middle of this range are McManus et al. (2010) and Graze et al. (2012) (with 51 and 45%, respectively), both of which used Illumina sequencing to produce RNA-seq data with sufficient coverage to test over 6000 genes for evidence of *cis*-regulatory divergence. Finally, at the low end of the range are Graze et al. (2009) and Fontanillas et al. (2010), with 27 and 12%, respectively. Both of these studies used methods to measure ASEs that have lower power to detect *cis*-regulatory changes than the other five studies. For example, the microarray data used by Graze et al. (2009) detected only 7% of genes surveyed in heads as differentially expressed between *D. melanogaster* and *D. simulans* compared to 54% of genes that were detected as differentially expressed in a prior microarray analysis of the same species and cell types (Ranz et al., 2004). Fontanillas et al. (2010) inferred ASE from an RNA-seq data set that was shown by theoretical work in the same paper to have low power.

The number of genes showing evidence of *trans*-regulatory divergence also varied among these studies. Landry et al. (2005) and Wittkopp et al. (2008) reported significant changes in *trans*-regulation for 74% and 78% of genes assayed, respectively, compared to only 53% of genes in Wittkopp et al. (2004). This difference might be a consequence of the way genes were selected for analysis in Landry et al. (2005) and Wittkopp et al. (2008); some genes were analyzed in these studies precisely because they showed both *cis*- and *trans*-regulatory divergence in Wittkopp et al. (2004). A more unbiased survey of genes found evidence of *trans*-regulatory changes for 66% of the 9966 genes analyzed (McManus et al., 2010). Graze et al. (2009) showed the lowest percentage of genes with *trans*-regulatory divergence (16%), consistent with the low frequency of total expression differences and *cis*-regulatory changes observed in this study. The remaining two studies (Fontanillas et al., 2010; Graze et al., 2012), analyzed expression only in F_1 hybrids and thus were not able to test for *trans*-regulatory divergence.

cis-Regulatory and *trans*-regulatory changes are not mutually exclusive, and many genes show evidence of both types of changes between species. Specifically, 18 (53%) of 34 genes tested by Wittkopp et al. (2004), 20 (65%) of 31 genes tested by Landry et al. (2005), 32 (65%) of 49 genes tested by Wittkopp et al. (2008), 93 (4%) of 2381 genes tested by Graze et al. (2009), and 3473 (35%) of 9966 genes tested by McManus et al. (2010) showed evidence of both *cis*- and *trans*-regulatory divergence (Table 3.2). For such genes, *cis*- and *trans*-regulatory changes can favor expression of the same allele ("*cis* + *trans*," Landry et al., 2005) or alternate alleles ("*cis* × *trans*," Landry et al., 2005) (Figure 3.3). Wittkopp et al. (2004) and McManus et al. (2010) found that *cis* × *trans* was slightly more common than *cis* + *trans* (10 genes vs. 8 genes and 1770 genes vs. 1703 genes, respectively), whereas Landry et al. (2005) found three times as many *cis* × *trans* changes ($n = 15$) as *cis* + *trans* changes ($n = 5$). The greater frequency of *cis* × *trans* changes

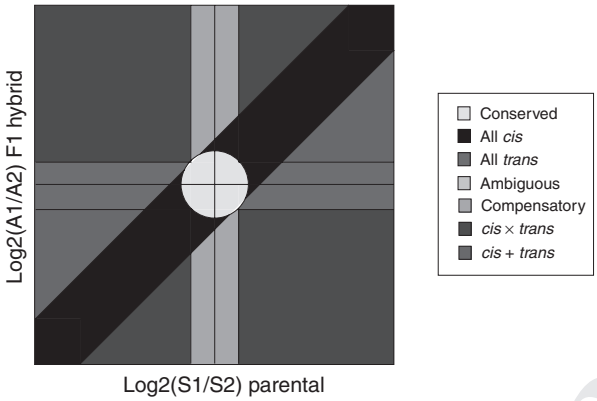

Figure 3.3 Plot of hypothetical results from an ASE study. The results of a hierarchical series of statistical tests of ASE from parental, and F1 hybrid samples. Log2-transformed parental expression ratio is plotted on the *x*-axis and log2-transformed F1 hybrid allelic expression ratio is plotted on the *y*-axis. Regions are color-coded based on classifications that would exist in each location in the plot. (For a color version of this figure, see Plate 3.3.)

observed in Landry et al. (2005) is not surprising given that genes with *cis* × *trans* regulatory changes were overrepresented by design. Stabilizing selection, which maintains phenotypes over evolutionary time, would result in *cis* × *trans* regulatory changes if it caused the fixation of mutations that (partially or fully) compensated at the level of total gene expression for the effects of earlier fixations. Consistent with this idea, differences in total expression between two strains of *D. melanogaster* are similar in magnitude and frequency to those observed between *D. melanogaster* and *D. simulans* despite more *cis*- and *trans*-regulatory changes between species (J.D. Coolon & P.J. Wittkopp, unpublished data).

Surprisingly, changes in total expression and *cis*-regulatory activity do not seem to be equally distributed between species. Sixty-four percent of the 7739 genes found to be differentially expressed between *D. melanogaster* and *D. sechellia* showed lower expression in *D. sechellia* (McManus et al., 2010). These differences might result from the downregulation of expression in *D. sechellia* (as opposed to the upregulation of expression in *D. melanogaster*) given that 90% of 130 genes with significant expression differences between *D. sechellia* and *D. simulans* also showed lower expression in *D. sechellia.* A similar asymmetry can also be seen at the level of *cis*-regulatory activity. Fontanillas et al. (2010), McManus et al. (2010), and Graze et al. (2012) all found that more transcripts were derived from *D. melanogaster* alleles than *D. simulans* or *D. sechellia* alleles in F_1 hybrids, and this does not seem to be a technical artifact. The cause of the observed greater total expression and allelic expression of *D. melanogaster* alleles is currently unknown and warrants further investigation.

A final point to consider when comparing the frequency of *cis*- and *trans*-regulatory changes reported in these studies is that they did not all sample the same cell types and age of adult flies. Gene expression is known to differ among sexes, tissues, and developmental stages (e.g., Churchill & Oliver, 2001; Jin et al., 2001; Ranz et al., 2003; Gibson et al., 2004), but it is less clear how *cis*- and *trans*-regulatory divergence might vary among these conditions. To address this issue, Wittkopp et al. (2008) compared divergence of total gene expression and *cis*-regulation between *D. melanogaster* and *D. simulans* in 7- to 10-day-old adult females (whole flies) to that of 3- to 5-day-old adult females (whole flies). Eight of the 17 genes examined showed significant differences in total gene expression between these samples, whereas 13 of these genes showed

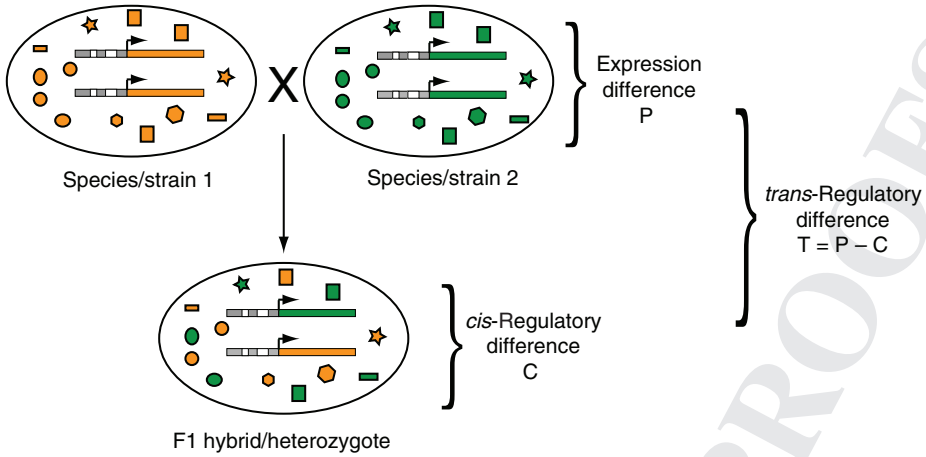

Plate 3.2 Experimental design to use ASE to determine *cis*- and *trans*-regulatory divergence in diploid organisms. Two species or strains are crossed to produce F1 hybrid offspring with a common *trans*-regulatory environment. *trans*-Acting molecules are indicated by colored shapes. Parental expression difference (P), *cis*-regulatory difference (C), and *trans*-regulatory differences are measured as indicated.

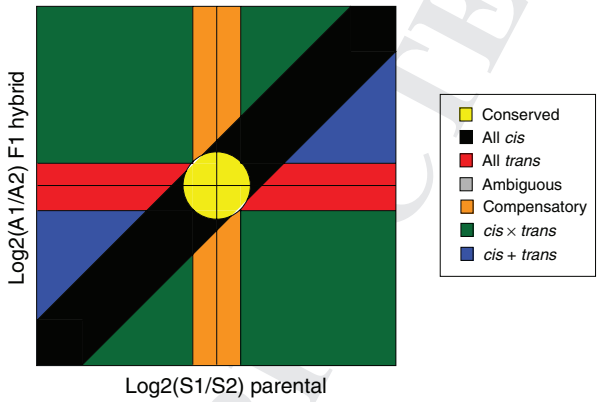

Plate 3.3 Plot of hypothetical results from an ASE study. The results of a hierarchical series of statistical tests of ASE from parental, and F1 hybrid samples. Log2-transformed parental expression ratio is plotted on the *x*-axis and log2-transformed F1 hybrid allelic expression ratio is plotted on the *y*-axis. Regions are color-coded based on classifications that would exist in each location in the plot.

significant differences in *cis*-regulatory divergence. The detection of *cis*-regulatory differences at one age but not another is likely attributable to differences in the availability and/or activity of *trans*-regulatory factors that interact with divergent *cis*-regulatory sequences (e.g., Cooley et al., 2012). *cis*- And *trans*-regulatory divergence can also be compared between studies in which RNA was sampled from *D. melanogaster* × *D. simulans* hybrids using different body parts. The most comparable pairs of such studies are McManus et al. (2010), which examined whole bodies, and Graze et al. (2012), which examined only heads. Heads were used to eliminate effects from atrophied gonads in the bodies of *D. melanogaster* and *D. simulans* F_1 hybrids (Santamaria, 1977; Dickinson et al., 1984). Forty-five percent of genes expressed and tested in the head showed evidence of significant *cis*-regulatory divergence between these species (Graze et al., 2012) compared to 51% of genes expressed and tested in the whole-body samples (McManus et al., 2010) (Table 3.2). This suggests that the additional tissue complexity found in whole bodies had a minimal effect on the overall frequency of *cis*-regulatory divergence, at least in this case.

How Do cis- *and* trans-*Regulatory Changes Relate to Inheritance of Gene Expression?*

Studies of gene expression in *Drosophila* typically use highly inbred strains in which most loci are homozygous for a single allele. When two inbred lines are crossed, the F_1 progeny are heterozygous for genes with different alleles in the two parents. By comparing gene expression (or any other phenotype) in the F_1 hybrids to both parents, the mode of inheritance can be inferred (Figure 3.4). For example, if expression in the F_1 hybrid is indistinguishable from one of the parents, that parental allele is said to be dominant; if the F_1 expression level is equal to the average expression

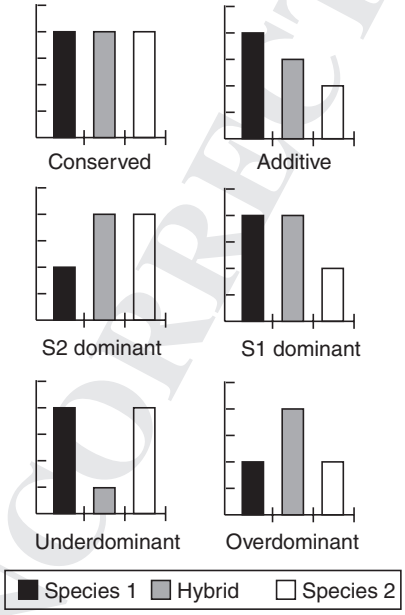

Figure 3.4 Mode of inheritance inferred from total gene expression in parental species and F1 hybrid offspring. Six plots with sample type on the *x*-axis (species 1, hybrid, and species 2) and expression level on the *y*-axis. Six patterns are illustrated. For additive, S1 dominant, and S2 dominant, there are two patterns that would be classified as such; only one is shown.

level of the parents, the alleles are said to be additive (or semidominant); finally, if the F_1 hybrid shows expression that is significantly higher or lower than both parents, the gene is said to be "misexpressed" or dysregulated (Gibson et al., 2004; Landry et al., 2005).

The mode of inheritance for each gene is determined by the specific genetic changes underlying divergent gene expression, and patterns of gene expression in interspecific hybrids of *Drosophila* suggest that the way in which these changes affect gene expression (i.e., in *cis* or in *trans*) is correlated with particular modes of inheritance. For example, *cis*-regulatory changes appear to result in additive inheritance of gene expression more often than *trans*-regulatory changes; this relationship has been seen for new regulatory mutations in the baker's yeast *Saccharomyces cerevisiae* (Gruber et al., 2012), for polymorphic expression in *D. melanogaster* (Lemos et al., 2008), and for divergent expression between *D. melanogaster* and *D. sechellia* (McManus et al., 2010). Molecularly, this relationship is expected because transcripts from each allele in a diploid cell are thought to contribute to total gene expression independently.

Another example of how *cis*- and *trans*-regulatory divergence relates to the mode of inheritance for gene expression comes from the observation that *cis*- × *trans*-regulatory divergence (i.e., compensatory changes in both *cis*- and *trans*- that favor expression of opposite alleles) is more common for genes that show misexpression in F_1 hybrids than for genes that show other types of inheritance. This relationship was first reported for *D. melanogaster* and *D. simulans* by Landry et al. (2005), who reasoned that dysregulation could result from the co-evolution of *cis*- and *trans*-acting factors within a species: when such *cis*- and *trans*-acting factors from two different species meet in interspecific hybrids, interactions between them can cause abnormally high or low expression. Consistent with this hypothesis, 13 (57%) of 23 genes that were misexpressed in *D. melanogaster* × *D. simulans* F_1 hybrids showed *cis*- × *trans*-regulatory divergence (Landry et al., 2005). By contrast, only 1 (20%) of 5 genes with dominant or additive inheritance showed evidence of *cis*- × *trans*-regulatory divergence (Landry et al., 2005). At a genomic scale, and comparing *D. melanogaster* with *D. sechellia* instead of *D. simulans*, genes showing misexpression were also significantly enriched for *cis* × *trans* types of changes (McManus et al., 2010): 21% of 2518 genes that were misexpressed in F_1 hybrids showed evidence of *cis*- × *trans*-regulatory divergence compared to only 12% of genes with dominant or additive inheritance. These types of regulatory changes (i.e., *cis* × *trans*) might play an important role in speciation if misexpression of genes commonly results in hybrid incompatibilities. Indeed, *cis*- × *trans*-regulatory divergence affecting expression of the *lethal hybrid rescue* gene has recently been shown to contribute to hybrid lethality in *D. melanogaster* and *D. simulans* (Maheshwari & Barbash, 2012).

How Much Do Neutral and Nonneutral Processes Contribute to cis- *and* trans-*Regulatory Divergence among Species?*

Expression differences between species can result from natural selection and/or genetic drift (Fay & Wittkopp, 2008). The same is true for the *cis*- and *trans*-regulatory changes that give rise to expression divergence. Comparing the relative contributions of *cis*- and *trans*-regulatory changes within and between species is one way to test for evidence of natural selection. Wittkopp et al. (2008) did just that for *D. melanogaster* and *D. simulans* and found that *cis*-regulatory changes explained a greater proportion of expression differences between than within species. A similar trend was observed for yeast species (Tirosh et al., 2009; Emerson et al., 2010). Even within *D. melanogaster*, the relative contribution of *cis*-regulatory changes to expression differences was greater for strains isolated from different geographic locations than for strains collected from the same population

(Lemos et al., 2008). These observations suggest that natural selection influences the molecular mechanisms by which gene expression evolves (i.e., through *cis*- or *trans*-acting changes).

Assuming that most *cis*- and *trans*-regulatory factors have not diverged to the point where they fail to interact at all between species (Takahashi et al., 2011), the greater contribution of *cis*-regulatory changes between than within species suggests that *cis*-regulatory changes accumulate preferentially during evolution (Wittkopp et al., 2008). This could occur if natural selection either favors the fixation of *cis*-regulatory changes or disfavors the fixation of *trans*-regulatory changes. To distinguish between these models, Emerson et al. (2010) used *cis*- and *trans*-regulatory polymorphisms observed in yeast to simulate the expected divergence between yeast species and compared these simulated values to the observed divergence. This analysis showed that *trans*-regulatory divergence was consistent with neutral evolution, whereas *cis*-regulatory divergence exceeded neutral expectations, suggesting that the greater contribution of *cis*-regulatory changes to interspecific expression differences (as compared to intraspecific expression differences) results primarily from selection favoring *cis*-regulatory variants.

The most common argument invoked for why *cis*-regulatory changes might fix preferentially over *trans*-regulatory changes involves differences in pleiotropy. Specifically, *cis*-regulatory mutations are thought to have lower pleiotropy (i.e., affect fewer phenotypes), on average, than *trans*-regulatory mutations (Carroll, 2005; Stern & Orgogozo, 2008). This lower pleiotropy is assumed to result in less deleterious effects on fitness. However, other differences in the properties of *cis*- and *trans*-regulatory mutations also exist that can influence their probability of fixation (Wittkopp, 2005). For example, *trans*-regulatory mutations appear to be recessive more often than *cis*-regulatory mutations (Gruber et al., 2012), suggesting that they are less efficient targets of natural selection. Consistent with this idea, *trans*-regulatory variation segregating within *D. melanogaster* (Lemos et al., 2008) and *trans*-regulatory divergence between *D. melanogaster* and *D. sechellia* (McManus et al., 2010) were found to be recessive more often than *cis*-regulatory variation. The flip side of this is that *cis*-regulatory mutations, polymorphisms, and divergence were found to be more likely to be at least additive/semidominant (Lemos et al., 2008; McManus et al., 2010; Gruber et al., 2012). This should allow *cis*-regulatory changes to be subject to natural selection as soon as they arise. Determining the relative impact of pleiotropy and dominance on the evolution of *cis*- and *trans*-regulation addresses a fundamental question about the evolutionary process that deserves much further investigation.

How Is cis*-Regulatory Divergence Related to Local Sequence Evolution?*

Changes in *cis*-regulatory activity often result from mutations in *cis*-regulatory regions located close (in the genome) to the affected gene; by contrast, *trans*-regulatory divergence is caused by changes that can exist anywhere in the genome. This suggests that local sequence evolution might correlate with divergent *cis*-regulatory activity. To test this idea, McManus et al. (2010) measured the level of sequence divergence (SNPs and indels) in the 1 kb region 5′ of the transcription start site of each gene and found that genes with significant *cis*-regulatory changes between *D. melanogaster* and *D. sechellia* had significantly greater levels of sequence divergence than genes without significant *cis*-regulatory changes. A similar observation has been made for *Saccharomyces* species (Tirosh et al., 2009), variation within *Arabidopsis thaliana* (Zhang & Borevitz, 2009) and for variation within *D. simulans* (Lawniczak et al., 2008).

Divergence of total expression and *cis*-regulatory activity has also been shown to correlate with divergence of coding sequences between *Drosophila* species. This was first reported by Nuzhdin et al. (2004), who found that expression differences between *D. melanogaster* and *D. simulans* detected

using a microarray with probes matching *D. melanogaster* sequence correlated with divergence at nonsynonymous sites. Although it is possible that this correlation was driven by differences in hybridization between *D. simulans* transcripts and *D. melanogaster* probes, the lack of correlation between expression divergence and synonymous sites argues against this explanation. Graze et al. (2012) used McDonald–Kreitman tests (McDonald & Kreitman, 1991) in conjunction with the direction of selection statistic (Stoletzki & Eyre-Walker, 2011) to demonstrate a correlation between *cis*-regulatory divergence and sequence divergence in coding sequences that suggest the action of positive selection. Their analysis includes the observation that groups of genes with larger *cis*-regulatory differences between *D. melanogaster* and *D. simulans* contain a larger proportion of genes that showed evidence for positive selection in the associated coding sequences than groups of genes with smaller *cis*-regulatory differences. Taken together, these observations suggest that natural selection might have correlated effects on coding and *cis*-regulatory sequences for many genes.

What Are the Phenotypic Consequences of cis- *and* trans-*Regulatory Divergence?*

Genomic scans for divergent expression between *Drosophila* species and divergent ASE in interspecific hybrids of *Drosophila* can identify "candidate" genes that might contribute to the evolution of adaptive phenotypes. To search for phenotypes most likely to have diverged because of changes in gene expression, genes with expression differences between species can be tested for an enrichment of genes with related functions using Gene Ontology (GO) terms. Using this approach, Graze et al. (2009) found that genes differentially expressed between *D. melanogaster* and *D. simulans* were enriched for olfaction and defense functions. Specifically, 41 defense genes and 7 odorant-binding genes showed divergent expression and evidence of positive selection (Graze et al., 2009). The functional role of these genes in species differences is not yet known, but defense genes might have co-evolved with species-specific pathogens that they interact with in their natural environment and the chemosensory genes might be involved in pheromone sensation, important for species-specific reproduction behaviors. Genes downstream of the sex-determination pathway were also found to be enriched among genes with divergent expression (Graze et al., 2009) and could also contribute to reproductive isolation between the two species.

More Recently, Graze et al. (2012) examined changes in *cis*-regulation between *D. melanogaster* and *D. simulans* on a larger scale and found that the functional enrichment for particular GO groups was sometimes specific to the direction of the bias of allelic expression. For example, genes with *cis*-regulatory changes favoring the *D. melanogaster* allele were enriched for H3-K4 methyltransferase activity and RNA-induced silencing complex genes, whereas those with *cis*-regulatory changes favoring the *D. simulans* allele showed enrichment of genes involved in the sensory perception of chemical signals in addition to H3-K4 methyltransferase activity. Genes involved in defense were also overrepresented among genes with *cis*-regulatory activity, consistent with the enrichment of defense genes among genes with divergent total expression between *D. melanogaster* and *D. simulans* (Graze et al., 2009). Studies of yeast and mice have identified consistent *cis*-regulatory changes for groups of genes with related functions as well, which suggests that natural selection is responsible for coordinately changing their expression (Bullard et al., 2010; Fraser et al., 2011). These observations suggest hypotheses about the genetic basis of adaptive evolution that should be tested functionally in the future.

In contrast to *D. melanogaster* and *D. simulans*, both of which are cosmopolitan species with large population sizes and similar lifestyles, *D. melanogaster* and *D. sechellia* differ substantially in population size and habitat (Kliman et al., 2000; Legrand et al., 2009). *D. sechellia* appears to have gone through a severe bottleneck while colonizing the Seychelles islands and has evolved a

novel feeding specialization for the fruit of *Morinda citrifolia* (Jones, 2005). Interestingly, the fruits of this plant produce toxic compounds (especially octanoic and hexanoic acids) that make it toxic to other *Drosophila* species but serve as attractants for *D. sechellia* (R'Kha et al., 1991). Genes involved in acetylcholine signaling and fatty acid metabolism, both of which are important for the detoxification of plant volatile compounds, have previously been implicated in the adaptation to *M. citrifolia* (Jones, 2005; Dworkin & Jones, 2009;). McManus et al. (2010) found that many of these genes, as well as genes with related functions, have evolved changes in *cis*- and/or *trans*-regulation that contribute to their divergent expression between *D. sechellia* and *D. melanogaster*. In addition, greater overall *trans*-regulatory divergence was observed between *D. melanogaster* and *D. sechellia* than between *D. melanogaster* and *D. simulans*, which might result from the fixation of segregating *trans*-regulatory variants in *D. sechellia* by genetic drift during the bottleneck.

cis- and *trans*-Regulatory Evolution in *Drosophila*: A Look Ahead

As described in this chapter, studies of *cis*- and *trans*-regulation in *Drosophila* interspecific hybrids have provided great insight into regulatory evolution, yet many questions remain:

How are* cis-** ***and* trans-*regulatory effects distributed among isoforms?

Over 60% of genes in *D. melanogaster* appear to be alternatively spliced, and the abundance of specific isoforms differs among tissues and developmental stages (Graveley et al., 2011; Smibert et al., 2012). This suggests that some *cis*- and *trans*-regulatory changes between species might be isoform-specific. With one exception (Graze et al., 2012), existing studies of ASE in interspecific hybrids of *Drosophila* are insufficient to address this issue because they measured ASE using information from only a single exon (Wittkopp et al., 2004, 2008; Landry et al., 2005), from only constitutive exons (Graze et al., 2009; McManus et al., 2010), or by combining data across all exons (Fontanillas et al., 2010). By contrast, Graze et al. (2012) compared ASE among exons for the 623 genes identified as having divergent *cis*-regulatory sequences, and found evidence of isoform-specific *cis*-regulatory divergence for 232 (37.2%) of these genes. This suggests that such changes are common and warrant more in-depth investigation on a larger scale.

***How often do* cis-** ***and* trans-*regulatory changes have sex-specific effects*?**

Gene expression differs between the sexes, and many regulatory changes between species affect expression of sex-biased genes (Ranz et al., 2003). This suggests the presence of sex-specific *cis*- and/or *trans*-regulatory changes, but the frequency of such changes remains unknown because all studies of ASE in *Drosophila* interspecific hybrids to date have analyzed RNA extracted from F_1 hybrid females. This is primarily because F_1 hybrid females allow ASE to be assessed for both autosomal and X-linked genes and because parent-of-origin effects and imprinting can be investigated by comparing daughters from reciprocal crosses with identical nuclear genomes. Regulatory divergence that is sex-specific might play an important role in hybrid incompatibilities and speciation, given that many sex-biased genes are involved in spermatogenesis, oogenesis, and phenotypes important for reproduction such as the sensation of mating pheromones.

***Which nucleotide changes underlie* cis-** ***and* trans-*regulatory divergence*?**

As described in the introduction, using ASE in interspecific hybrids to study regulatory evolution allows the net effects of *cis*- and *trans*-acting changes to be inferred, but provides no information about the specific genomic location of individual genetic changes responsible for these effects. Identifying specific mutations responsible for expression divergence will provide a much better

mechanistic understanding of the evolutionary process. This task will likely be much easier for *cis*-regulatory changes than for *trans*-regulatory changes because *cis*-acting mutations are expected to be located in sequences nearby the affected coding region, allowing candidate sites to be readily identified and tested. Identifying genetic changes responsible for *trans*-regulatory divergence, however, requires linkage mapping and/or GWAS prior to functional tests. Only when individual regulatory mutations are identified and their phenotypic effects confirmed can we ask questions about the evolution of specific traits, the frequency and interactions among mutations that alter gene expression, how regulatory variants are distributed within the genome and within regulatory networks, and how network architecture influences the genomic sources of regulatory evolution.

How generalizable are patterns of* cis- *and* trans-*regulatory divergence observed between D. melanogaster and its close relatives?

For over a century, *D. melanogaster* has proven to be an invaluable model system for molecular, cellular, developmental, and evolutionary biology. Given the conserved structure of regulatory networks among eukaryotes, many patterns of regulatory evolution, observed between *D. melanogaster* and its close relatives, might also hold for other species—both within and outside of the genus *Drosophila*. Consistent with this idea, the greater contribution of *cis*-regulatory changes to expression differences between than within species described above has also been observed for the baker's yeast *Saccharomyces* and its close relative *S. paradoxus* (Emerson et al., 2010). For more information about the generality of conclusions presented here, see Chapters 2 and 7.

Despite many advances in the last 3 years, further refinements in methods for detecting ASE and total expression on a genomic scale that allow analysis of not only whole animals and specific tissues but also individual cell types—combined with advances in the computational and statistical methods needed to analyze these data—will help resolve some of these issues. Other questions will require in-depth genetic, biochemical, and molecular tests of function. *Drosophila* species have been (and are expected to remain) at the forefront of both of these types of research, and we expect comparative studies of *Drosophila* species and their interspecific hybrids to continue, leading the way toward answering questions about *cis*- and *trans*-regulatory divergence, specifically, and the genetic basis of evolutionary change in general.

References

Abzhanov, A., Protas, M., Grant, B.R., Grant, P.R., & Tabin, C.J. (2004) *Bmp4* and morphological variation of beaks in Darwin's finches. *Science* **305**, 1462–1465.

Ahmadian, A., Gharizadeh, B., Gustafsson, A., et al. (2000) Single-nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem* **280**, 103–110.

Babak, T., Garrett-Engele, P., Armour, C.D., et al. (2010) Genetic validation of whole-transcriptome sequencing for mapping expression affected by *cis*-regulatory variation. *BMC Genomics* **11**, 473.

Brem, R., Yvert, G., Clinton, R., & Kruglyak, L. (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755.

Bullard, J.H., Mostovoy, Y., Duboit, S., & Brem, R.B. (2010) Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proc Natl Acad Sci USA* **107**, 5058–5063.

Carroll, S.B. (2005) Evolution at two levels: on genes and form. *PLoS Biol* **3**, e245.

Chan, Y.F., Marks, M.E., Jones, F.C., et al. (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **5963**, 302–305.

Churchill, G.A. & Oliver, B. (2001) Sex, flies and microarrays. *Nat Genet* **29**, 355–356.
Clark, R.M., Wagler, T.N., Quijada, P., & Doebley, J. (2006) A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat Genet* **38**, 594–597.
Cooley, A.M., Shefner, L., Mclaughlin, W.N., Stewart, E.E., & Wittkopp, P.J. (2012) The ontogeny of color: Developmental origins of pigment divergence in *Drosophila americana* and *D. novamexicana*. *Evol Dev* **14** (4), 317–325.
Cowles, C., Hirschhorn, J., Altshuler, D., & Lander, E. (2002) Detection of regulatory variation in mouse genes. *Nat Genet* **32**, 432–437.
Cretekos, C.J., Wang, Y., Green, E.D., et al. (2008) Regulatory divergence modifies limb length between mammals. *Genes Dev* **22**, 141–151.
Cutter, A. (2008) Divergence times in caenorhabditis elegans and *Drosophila* inferred from direct estimates of te neutral mutation rate. *Mol Biol Evol* **25**, 778–786.
Degner, J.F., Marioni, J.C., Pai, A.A., et al. (2009) Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics* **25**, 3207–3212.
Deveale, B., van der Kooy, D., & Babak, T. (2012) Critical evaluation of imprinted gene expression by RNA-seq: a new perspective. *PLoS Genet* **8**, e1002600.
Dickinson, W.J., Rowan, R.G., & Brennan, M.D. (1984) Regulatory gene evolution: adaptive differences in expression of alcohol dehydrogenase in *Drosophila melanogaster* and *Drosophila simulans*. *Heredity* **52**, 215–225.
Dixon, A.L., Liang, L., Moffant, M.F., et al. (2007) A genome-wide association study of global gene expression. *Nat Genet* **39**, 1202–1207.
Dworkin, I. & Jones, C.D. (2009) Genetic changes accompanying the evolution of host specialization in *Drosophila sechellia*. *Genetics* **181**, 721–736.
Emerson, J.J., Hsieh, L., Sung, H., et al. (2010) Natural selection on *cis* and *trans* regulation in yeasts. *Genome Res* **20**, 826–836.
Fay, J.C. & Wittkopp, P.J. (2008) Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* **100**, 191–199.
Fontanillas, P., Landry, C.R., Wittkopp, P.J., et al. (2010) Key considerations for measuring allelic expression on a genomic scale using high-throughput sequencing. *Mol Ecol* **19**, 212–227.
Fraser, H.B., Babak, T., Tsang, J., et al. (2011) Systematic detection of polygenic cis-regulatory evolution. *PLoS Genet* **7**, e1002023.
Gibson, G. (2002) Microarrays in ecology and evolution: a preview. *Mol Ecol* **11**, 17–24.
Gibson, G. & Weir, B. (2005) The quantitative genetics of transcription. *Trends Genet* **21**, 616–623.
Gibson, G., Riley-Berger, R., Harshman, L.G., et al. (2004) Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. *Genetics* **167**, 1791–1799.
Graveley, B.R., Brooks, A.N., Carlson, J.W., et al. (2011) The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479.
Graze, R.M., McIntyre, L.M., Main, B.J., Wayne, M.L., & Nuzhdin, S.V. (2009) Regulatory divergence in *Drosophila melanogaster* and *D. simulans*, a genome-wide analysis of allele-specific expression. *Genetics* **183**, 547–561.
Graze, R.M., Novelo, L.L., Fear, J.M., Casella, G., Nuzhdin, S.V., & McIntyre, L.M. (2012) Allelic imbalance in *Drosophila* hybrid heads: exons, isoforms and evolution. *Mol Biol Evol* **29**, 1521–1532.
Gruber, J.D., Vogel, K., Kalay, G., & Wittkopp, P.J. (2012) Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *Saccharomyces cerevisiae*: frequency, effects, and dominance. *PLoS Genet* **8**, e1002497.
Guo, M., Yang, S., Rupe, M., et al. (2008) Genome-wide allele-specific expression analysis using Massively Parallel Signature Sequencing (MPSS) reveals cis- and trans-effects on gene expression in maize hybrid meristem tissue. *Plant Mol Biol* **66**, 551–563.
Jin, W., Riley, R.M., Wolfinger, R.D., White, K.P., Passador-Gurgel, G., & Gibson, G. (2001) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat Genet* **29**, 389–395.
Jones, C.D. (2005) The genetics of adaptations in *Drosophila sechellia*. *Genetica* **123**, 137–145.
Kliman, R.M., Andolfatto, P., Coyne, J.A., et al. (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**, 1913–1931.
Krishna Pant, P.V., Tao, H., Beilharz, E.J., Ballinger, D.G., Cox, D.R., & Frazer, K.A. (2006) Analysis of allelic differential expression in human white blood cells. *Genome Res* **16**, 331–339.
Kronforst, M., Barsh, G.S., Kopp, A., et al. (2012) Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. *Plant Cell Melanoma Res*. doi:10.1111/j.1755-1148X.2012.01014.x.
Landry, C.R., Wittkopp, P.J., Taubes, C.H., Ranz, J.M., Clark, A.G., & Hartl, D.L. (2005) Compensatory *cis-trans* evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics* **171**, 1813–1822.
Lawniczak, M.K., Holloway, A.K., Begun, D.J., & Jones, C.D. (2008) Genomic analysis of the relationship between gene expression variation and DNA polymorphism in *Drosophila simulans*. *Genome Biol* **9**, R125.
Lee, W.H. & Watanabe, T. (1987) Evolutionary genetics of the *Drosophila melanogaster* subgroup I. Phlyogenetic relationships based on matings, hybrids and proteins. *Jpn J Genet* **62**, 225–239.

Legrand, D., Tenaillon, M., Matyot, P., Gerlach, J., Lachaise, D., & Cariou, M.L. (2009) Species-wide genetic variation and demographic history of *Drosophila sechellia*, a species lacking population structure. *Genetics* **182**, 1197–1206.

Lemos, B., Araripe, L.O., Fontanillas, P., & Hartl, D.L. (2008) Dominance and the evolutionary accumulation of *cis*- and *trans*-effects on gene expression. *Proc Natl Acad Sci USA* **105**, 14471–14476.

Li, J. & Burmeister, M. (2005) Genetical genomics: combining genetics with gene expression analysis. *Hum Mol Genet* **14**, R163–R169.

Maheshwari, S. & Barbash, D.A. (2012) Cis-by-trans regulatory divergence causes the asymmetric lethal effects of an ancestral hybrid incompatibility gene. *PLoS Genet* **8**, e1002597.

Main, B.J., Bickel, R.D., McIntyre, L.M., Graze, R.M., Calabrese, P.P., & Nuzhdin, S.V. (2009) Allele-specific expression assays using Solexa. *BMC Genomics* **10**, 422.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., & Gilad, Y. (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* **9**, 1509–1517.

McDonald, J.H. & Kreitman, M. (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**, 652–654.

McManus, C.J., Coolon, J.D., Duff, M.O., Eipper-Mains, J., Graveley, B.R., & Wittkopp, P.J. (2010) Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res* **20**, 816–825.

de Meaux, J., Goebel, U., Pop, A., & Mitchell-Olds, T. (2005) Allele-specific assay reveals functional variation in the chalcone synthase promoter of *Arabidopsis thaliana* that is compatible with neutral evolution. *Plant Cell* **17**, 676–690.

Michalak, P. & Noor, M.A.F. (2003) Genome-wide patterns of expression in *Drosophila* pure—species and hybrid males. *Mol Biol Evol* **20**, 1070–1076.

Moustakas, J.E. (2008) Development of the carapacial ridge: implications for the evolution of genetic networks in turtle shell development. *Evol Dev* **10**, 29–36.

Nagalakshmi, U., Wang, Z., Waern, K., et al. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349.

Nuzhdin, S.V., Wayne, M.L., Harmon, K.L., & McIntyre, L.M. (2004) Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Mol Biol Evol* **21**, 1308–1317.

Oleksiak, M.F., Roach, J.L., & Crawford, D.L. (2005) Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nat Genet* **37**, 67–72.

R'Kha, S., Capy, P., & David, J.R. (1991) Host-plant specialization in the *Drosophila melanogaster* species complex: a physiological, behavioral, and genetical analysis. *Proc Natl Acad Sci USA* **88**, 1835–1839.

Ranz, J.M., Castillo-Davis, C., Meiklejohn, C.D., & Hartl, D.L. (2003) Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300**, 1742–1745.

Ranz, J.M. & Machado, C.A. (2006) Uncovering evolutionary patterns of gene expression using microarrays. *Trends Ecol Evol* **21**, 29–37.

Ranz, J.M., Namgyal, K., Gibson, G., & Hartl, D.L. (2004) Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. *Genome Res* **3**, 373–379.

Rifkin, S.A., Kim, J., & White, K.P. (2003) Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat Genet* **33**, 138–144.

Ronald, J., Akey, J.M., Whittle, J., Smith, E.N., Yvert, G., & Kruglyak, L. (2005) Simultaneous genotyping, gene-expression measurement, and detection of allele-specific expression with oligonucleotide arrays. *Genome Res* **15**, 284–291.

Ruden, D.M., Chen, L., Possidente, D., et al. (2009) Genetical toxicogenomics in *Drosophila* identifies master-modulatory loci that are regulated by developmental exposure to lead. *Neurotoxicology* **30**, 898–914.

Santamaria, P. (1977) On the causes of sterility in some interspecific hybrids from the *melanogaster* subgroup of *Drosophila*. *Roux Arch* **82**, 305–310.

Sawamura, K., Watanabe, T.K., & Yamamoto, M. (1993) Hybrid lethal systems in the *Drosophila melanogaster* species complex. *Genetica* **88**, 175–185.

Schadt, E.E., Monks, S.A., Drake, T.A., et al. (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302.

Serre, D., Gurd, S., Ge, B., et al. (2008) Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. *PLoS Genetics* **4** (2), e1000006.

Shapiro, M.D., Marks, M.E., Peichel, C.L., et al. (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723.

Singer-Sam, J. & Gao, C. (2001) Quantitative RT-PCR-based analysis of allele-specific gene expression. *Methods Mol Biol* **181**, 145–152.

Smibert, P., Miura, P., Westholm, J.O., et al. (2012) Global patterns of tissue-specific alternative polyadenylation in *Drosophila*. *Cell Rep* **1** (3), 277–289.

Stern, D.L. & Orgogozo, V. (2008) The loci of evolution: how predictable is genetic evolution? *Evolution* **62**, 2155–2177.

Stoletzki, N. & Eyre-Walker, A. (2011) Estimation of the neutrality index. *Mol Biol Evol* **28**, 63–70.

Takahashi, K.R., Matsuo, T., & Takano-Shimizu-Kouno, T. (2011) Two types of cis-trans compensation in the evolution of transcriptional regulation. *Proc Natl Acad Sci USA* **108**, 15276–15281.

Tirosh, I., Reikhav, S., Levy, A.A., & Barkari, N. (2009) A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* **1**, 559–662.

West, M.A., Kim, K., Kliebenstein, D.J., et al. (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. *Genetics* **175**, 1441–1450.

Wilson, M.D., Barbosa-Morais, N.L., Schmidt, D., et al. (2008) Species-specific transcription in mice carrying human chromosome 21. *Science* **322**, 434–438.

Wittkopp, P.J. (2005) Genomic sources of regulatory variation in cis and in trans. *Cell Mol Life Sci* **62**, 1779–1783.

Wittkopp, P.J. (2006) Evolution of cis-regulatory sequence and function in Diptera. *Heredity* **97**, 139–147.

Wittkopp, P.J. (2011) Using pyrosequencing to measure allele-specific mRNA abundance and infer the effects of cis- and trans-regulatory differences. In: *Molecular Methods for Evolutionary Genetics* (eds V. Orgogozo & M. Rockman), pp. 297–317. Humana Press, New York.

Wittkopp, P.J., Carroll, S.B., & Kopp, A. (2003) Evolution in black and white: genetic control of pigment patterns in *Drosophila*. *Trends Genet* **19**, 495–504.

Wittkopp, P.J., Haerum, B.K., & Clark, A.G. (2004) Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430**, 85–88.

Wittkopp, P.J., Haerum, B.K., & Clark, A.G. (2008) Regulatory changes underlying expression differences within and between species. *Nat Genet* **40**, 346–350.

Wray, G.A., Hahn, M.W., Abouheif, E., Pizer, M., Rockman, M.V., & Romano, L.A. (2003) The evolution of transcription in eukaryotes. *Mol Biol Evol* **20**, 1377–1419.

Yang, Y., Graze, R.M., Walts, B.M., et al. (2011) Partitioning transcript variation in *Drosophila*: abundance, isoforms, and alleles. *Genes Genomes Genetics* **1**, 427–436.

Zhang, X. & Borevitz, J.O. (2009) Global analysis of allele specific expression in *Arabidopsis thaliana*. *Genetics* **182**, 943–954.