Genetics: Published Articles Ahead of Print, published on September 2, 2005 as 10.1534/genetics.105.047449

Compensatory *cis-trans* Evolution and the Dysregulation of Gene Expression in Interspecific Hybrids of *Drosophila*

Christian R. Landry[*], Patricia J. Wittkopp[†], Clifford H. Taubes[‡], Jose M. Ranz[§], Andrew G. Clark[†] and Daniel L. Hartl[*]

* Department of Organismic and Evolutionary Biology, Harvard, University, Cambridge.

† Department of Molecular Biology and Genetics, Cornell University, Ithaca.

‡ Department of Mathematics, Harvard University, Cambridge.

§ Department of Genetics, University of Cambridge, Cambridge, UK.

Running head:

Gene and expression in interspecific hybrids

Key words:

Gene regulation, *Drosophila*, hybrid, dysregulation, allele-specific expression

Corresponding author:

Daniel L. Hartl

Department of Organismic and Evolutionary Biology

16 Divinity Avenue

Cambridge, MA 02138

dhartl@oeb.harvard.edu

Phone: 617-496-3917

Fax: 617-496-5854

ABSTRACT

Hybrids between species are often characterized by novel gene-expression patterns. A recent study on allele-specific gene expression in hybrids between species of *Drosophila* revealed cases in which *cis* and *trans* regulatory elements within species had coevolved in such a way that changes in *cis*-regulatory elements are compensated by changes in *trans*-regulatory elements. We hypothesized that such coevolution should often lead to gene misexpression in the hybrid. To test this hypothesis, we estimated allele-specific expression and overall expression levels for 31 genes in *D. melanogaster*, *D. simulans*, and their F1 hybrid. We found that 13 genes with *cis-trans* compensatory evolution are in fact misexpressed in the hybrid. These represent candidate genes whose dysregulation might be the consequence of coevolution of *cis* and *trans* regulatory elements within species. Using a mathematical model for the regulation of gene expression, we explored the conditions under which *cis-trans* compensatory evolution can lead to misexpression in interspecific hybrids.

INTRODUCTION

Variation in Darwinian fitness results from interactions among genes in the context of environmental variation. Through the process of natural selection, genes within a genome become coadapted (DOBZHANSKY 1937). The extent of genomic coadaptation within species can be measured *ex post facto* by disrupting the harmonious genetic background. A classical example is the production of novel phenotypes in hybrids between species, which is thought to result, at least in part, from combinations of incompatible gene products encoded by the respective genomes now present in the same cells. Although few examples of the molecular basis of such interactions are known, it is likely that regulatory differences are important (ORR and PRESGRAVES 2000; OSBORN *et al.* 2003).

Gene regulation involves numerous molecular interactions. Elements within a given regulatory module are expected to be coadapted (DOVER 1992; PORTER and JOHNSON 2002). So far, there is limited evidence for regulatory coadaptation, owing to the difficulty of addressing this question on a large scale. Recent genome-scale measurements of gene expression in interspecific hybrids of *Drosophila* (MICHALAK and NOOR 2003; RANZ *et al.* 2004), maize (AUGER *et al.* 2004), and *Arabidopsis* (WANG *et al.* 2004; COMAI *et al.* 2003) have revealed numerous genes with expression levels in F1 hybrids that are completely outside the range of that found in the parental species. These observations support the hypothesis of pervasive coadaptation of genetic regulatory elements. The reasons for the hybrid misexpression are unclear. As general possibilities, anomalies in the hybrid transcriptome could be a consequence of irregular

development, generalized stress response associated with the presence of two incompatible genomes in the same cells, or dysregulation of a few major regulatory genes whose effects cascade through the regulatory network (COMAI *et al.* 2003; OSBORN *et al.* 2003). In general, these scenarios are difficult to disentangle. The hypothesis of *cis x trans* interaction at individual genes, however, can be subjected to a rigorous experimental test. We can show whether, for individual genes, the interaction of the *trans* regulatory elements from one species with the *cis* regulatory elements of the other is responsible for the dysregulation in the hybrid. This model might be most appropriate for genes whose *cis* and *trans* regulatory elements have evolved since the time that the parental species became reproductively isolated (e.g., SHAW *et al*. 2002). It is this hypothesis that we set out to examine for a sample of genes.

The rationale for the hypothesis is that the regulation of gene expression requires the harmonious interaction of *trans*-regulatory elements (in the simplest case, transcription factors) with *cis*-regulatory elements (in the simplest case, DNA binding sites for the transcription factors) (e. g., PTASHNE and GANN 2002). If natural selection acts to maintain an optimal level of gene expression through time (stabilizing selection), as it appears to (LEMOS *et al.* 2005; DENVER *et al.* 2005), then genetic changes in *cis* and *trans* regulatory elements that compensate each other may accumulate. Consequently, regulatory elements may diverge genetically between species even though the level of gene expression remains approximately constant. Thermodynamic models that have examined the properties of gene regulatory systems also suggest that similar phenotypes may be covering up genetic variation in the underlying regulatory elements and proteins (GIBSON 1996; VEITIA 2003). Compensated *cis* and *trans* regulatory evolution between species is a manifestation of a process coined developmental-

system drift by TRUE and HAAG (2001), through which the phenotypes are evolutionarily maintained despite a turnover of the underlying developmental networks.

Experimentally, different types of regulatory divergence can be detected by assaying allele-specific gene expression in related species and their hybrids (e.g. OHNO 1969; PARKER *et al.* 1985; WRAY *et al.* 2003; PASTINEN and HUDSON 2004). When the alleles differ in their expression to the same extent in the parental species as in the hybrids, then *cis*-acting genetic differences may be inferred (e. g., YAN *et al.* 2002; WITTKOPP *et al.* 2004). When the alleles differ in their expression to a larger extent in the parental species than in the hybrids, then *trans*-acting genetic differences may be inferred. However, when the *cis* regulatory divergence between alleles detected in the hybrid background is larger than the divergence between species, we can infer that there is also divergence in *trans* that compensates the changes in *cis* to bring the level of expression of the gene to a more similar level between species than expected given the divergence in *cis.* An extreme example of this scenario would be when the allelic divergence in the hybrid is in the opposite direction to the species divergence. A recent survey of allelic expression of ~30 genes in *Drosophila melanogaster, D. simulans,* and their F1 hybrid has revealed that, for roughly one-third of the genes, the alleles differed in their expression to a smaller extent between the parental species than in the hybrids (WITTKOPP *et al.* 2004). This result suggests both *cis*-acting and *trans*-acting regulatory differences between the parental species, whose interactions within species produce more similar expression between species that expected given the regulatory divergence in *cis.*

To test for *cis-trans* regulatory coadaptation that would result in misexpression, it is necessary to assay the total levels of gene expression in hybrids and the parental species, as well as the ratio of expression of the parental alleles in the hybrid background. The reason is that if there is coevolution between *cis*-acting and *trans*-acting factors originating from the parental genomes, it may result in regulatory incompatibilities and in overall misexpression in the hybrids. To test whether misexpression in hybrids is accompanied by patterns of *cis-trans* compensatory regulation, we tested whether genes that are known to be misexpressed in the F1 hybrid between *Drosophila melanogaster* and *D. simulans* (RANZ *et al.* 2004) show differential expression of the parental alleles concordant with coevolution of *cis* and *trans* regulatory elements. We also test for total misexpression in hybrids for genes that exhibited *cis-trans* compensatory regulation in WITTKOPP *et al.* (2004). We combined these results to identify genes whose compensatory *cis-trans* regulation might contribute to their misexpression in hybrids between these two species. Finally, to explore the kinds of molecular changes that could underlie compensatory regulatory changes and hybrid misexpression, we developed a mathematical model of evolutionary divergence of gene regulation that incorporates changes in *cis* and *trans* regulatory elements. The qualitative behavior of this model suggests that *cis-trans* compensatory evolution can lead to misexpression in the hybrid as well as differential allelic expression, a pattern observed for a number of the genes examined experimentally.

MATERIAL AND METHODS

Gene expression analysis. Relative allele-specific expression in the hybrid background and relative gene expression between parental species was assayed in *D. melanogaster* and *D. simulans* females and their F1 hybrid. This sample includes 31 genes total (supplementary table 1): A set of 23 genes studied in a cross between *Drosophila melanogaster* Canton S females and *Drosophila simulans* Sim1 males, and a second set of 8 genes studied in a cross between *D. melanogaster zhr* females and *D. simulans* Tsimbazaza males. In both cases, virgin females of 3-5 days old were used to produce whole-body extracts of mRNA. The first set of genes was selected on the basis of their total expression level in F1 hybrids and the parental species as determined with DNA microarrays by RANZ *et al.* (2004), with no prior knowledge on the relative allelic expression of the genes in F1 hybrid. For this set of genes, we measured relative allelic expression level in the hybrid and between species using pyrosequencing (AHMADIAN *et al.* 2000). The total gene expression level in hybrids relative to the parents was obtained from Ranz *et al*. (2004). The second set of eight genes was chosen because previous results had shown patterns of expression consistent with compensatory *cis-trans* regulation in 24-28 hours old hybrids (WITTKOPP *et al.* 2004). Because gene expression is often age-specific, we used 3-5 day old flies for this work to match the sample conditions used for the microarray experiment (RANZ *et al*. 2004). Prior to this work, nothing was known about the total expression of these genes in hybrids. We measured the total level of expression in hybrids relative to the parental strains using quantitative real-time PCR (methods below) and allele-specific expression in hybrids and parental species using pyrosequencing. Only 4 of the 8 selected genes showed

patterns consistent with compensatory *cis-trans* regulation (6/8 still show divergence in *cis*) at this later stage.

RNA/DNA extraction for pyrosequencing assays. In the cross between *D. melanogaster* Canton S and *D. simulans* Sim1, frozen samples of adult flies used in the microarray analysis (RANZ *et al*. 2004) were studied. For the cross between *D. melanogaster zhr* and *D. simulans* Tsimbazaza, the same collection conditions were used (3-5 day old virgin females, frozen between 12-2 PM). Seven *D. melanogaster* and 7 *D. simulans* virgin females were pooled for each extraction (Figure S1). Pools containing 14 interspecific hybrid virgin females were extracted in parallel. Four independent RNA extractions, one per pool of flies, were performed and each used for four independent cDNA syntheses (Omniscript, Qiagen, Valencia CA). Total RNA was extracted using the SV RNA system (Promega, Madison, WI) according to WITTKOPP *et al.* (2004). Extractions with the SV total RNA kit allowed the independent isolation of genomic DNA from the same flies, which is required for normalization of the cDNA samples analyzed with pyrosequencing.

Quantitative real-time PCR. RNA was extracted in duplicate as described above without the DNA extraction step and with a DNAse treatment to eliminate any genomic DNA carry-over (DNAfree, Ambion, Austin, TX). Instead of relying on a housekeeping gene to control for RNA abundance among strains and extractions, we measured the concentration of total RNA in quadruplicate using RiboGreen® following the manufacturer's protocol (Molecular Probes, Eugene, OR). An equal amount of total RNA was used (1 ug) for two reverse-transcription reactions per RNA extraction, for a total of four cDNA synthesis reactions per strain using

Superscript II (Invitrogen, Carlsbad, CA) following standard protocols (using poly-dT and random hexamers). The four cDNA samples were pooled, and 4 real-time PCR reactions on each pooled cDNA were performed for each gene using custom primers and SYBR green 1 (Qiagen, Valencia, CA). Primers for the quantitative real-time PCR (rt-qPCR) were the same used in the pyrosequencing assays (see below) and designed relative to conserved regions between the species and strains. In order to establish the relationship between fold change and the cycle-threshold (CT) measured, we established standard curves (8 dilutions and one no-template control) for each of the genes by purifying PCR products (ExoSAP-IT, USB, Cleveland, OH) generated from primers external to the primers used for the rt-qPCR. PCR reactions were run on a MX3000P real-time PCR machine (Stratagene, La Jolla, CA). Ninety-five percent confidence intervals (CI) were calculated around the mean of the 4 replicates and non-overlapping CI for two samples were considered as significantly differentially expressed.

Allele-specific gene expression level. For all of the 31 genes, pyrosequencing was used to measure allele-specific gene expression as described in WITTKOPP *et al.* (2004). Briefly, a fragment of 500-1000 bp of the coding sequence of each of the genes was sequenced for each of the strains studied. Pyrosequencing uses a single nucleotide difference in the transcribed sequence to measure the relative expression of two alleles in the same sample. A primer is annealed upstream of the divergent site and extended one base at a time, with the number of nucleotides incorporated at each position proportional to the number of transcripts in the sample. The relative expression levels of *D. melanogaster* and *D. simulans* alleles were determined by calculating the ratio of species-specific nucleotide incorporation at the divergent site.

Four parental pools, each composed of 14 flies, and an additional 4 pools each with 14 hybrid flies, were analyzed. RNA from each of these 8 pools was used in 4 separate cDNA synthesis reactions. Pyrosequencing was used to measure the ratio of *D. melanogaster* to *D. simulans* alleles in each genomic DNA extraction in duplicate and each cDNA synthesis, for a total of 16 DNA and 32 cDNA measurements for each gene. For zhr/Tsimbazaza, cDNA samples from the four hybrid pools were measured in duplicate, whereas each hybrid DNA was measured once. Parental cDNA samples were measured in triplicate and DNAs were measured twice. A total of 12 DNA and 20 cDNA samples were collected for each gene examined in the zhr/Tsimbazaza cross. Replication of cDNA synthesis reactions and pyrosequencing of hybrid genomic DNA samples was reduced in the zhr/Tsimbazaza samples because these were found to be small sources of error in the Canton S/Sim1 samples (data not shown). Pyrosequencing reactions that did not meet quality control standards (based on manual examination of the signal for conserved bases and background noise) were excluded from analysis. In all, 4% of measurements were excluded for CantonS/Sim1 and 9% were excluded for zhr/Tsimbazaza. The ratios of the species-specific bases for each sample were then calculated. All ratios were $\log_2$-transformed to make them normally distributed. On this scale, a value of 0 means equal expression of both alleles, with positive and negative values representing more transcripts from *D. melanogaster* and *D. simulans* alleles, respectively.

Measurements of genomic DNA were used to normalize the cDNA ratios (WITTKOPP *et al.* 2004). No significant difference was observed among DNA measurements from replicate hybrid pools (ANOVA, $P = 0.86$), thus all measurements were combined for normalization. For hybrid cDNA measurements, the median of the log-transformed hybrid DNA measurements was

subtracted from each cDNA value to correct for differences in PCR amplification and/or pyrosequencing. Parental pools have an additional source of error from the potential unequal extraction of *D. melanogaster* and *D. simulans* alleles in the pools of adult flies. For each parental pool, a general linear model (SAS, SAS Institute Inc., Cary, NC, USA) was used to obtain regression fits of parental DNA on median hybrid DNA ($R^2 = 0.94$). The fitted estimate of parental DNA for each gene within each pool was subtracted from the corresponding cDNA measurements to account for differences in PCR amplification, pyrosequencing, and possible extraction differences. This regression model allows DNA measurements of the same pool determined using different genes to be combined to more accurately estimate and incorporate the extraction bias of each pool. Normalized, log-transformed cDNA ratios estimated by pyrosequencing were used for analysis. For each gene, within each cross (CantonS/Sim1 or zhr/Tsimbazaza), data were fitted to the following model using proc Mixed in SAS (SAS Institute Inc., Cary, NC, USA) using REML (Restricted Maximum Likelihood) to estimate parameters:

$$Y_{ij} = G_i + P_{ij} + e,$$

where Y is the log-transformed, normalized cDNA value, G is the genotype (i = parent or hybrid), P is the pool (j = 1, 2, 3, 4) treated as a random effect, and e is a random error term. The variance component of "pool" is determined independently for each generation using the group structure in proc Mixed. A random pool effect (nested within generation: P or H) was included in the model. Separate covariance parameters were fit for the pool effect in the P and H generations because the variance among replicate pools was always much larger for P than H. Student's t tests were computed within the Mixed procedure to test three null hypotheses concerning the relative level of expression of the parental alleles. These tests were for (1) no differential

expression in the parents (*P* ratio = 1), (2) no differential expression in the hybrid (*H* ratio = 1), and (3) ratio of allelic expression in the parental lines equals that in the hybrid (*P* ratio = *H* ratio). These null hypotheses correspond respectively to (1) no expression difference between the parental species, (2) no difference in *cis*-regulatory elements, and (3) no difference in *trans*-regulatory elements.

The DNA measurements provide an empirical reference point for establishing equal expression of the two alleles in cDNA. In all cases, the variance of cDNA measurement was much larger than that of the DNA samples. The range of replicate hybrid DNA measurements was typically1-2%. *Cis*-regulatory differences would only be incorrectly inferred if the difference between the observed and "true" ratio of DNA was greater than the 95% CI on the cDNA. In the mixed model used to assess significance, the observed error among replicate measures was incorporated for each gene individually. No systematic bias should be introduced. The normalization strategy was initially developed to calibrate linear standard curves. Note that a 1.4 fold change corresponds to pyrosequencing measurements of 58% of one allele and 42% of the other. A 2% error (i.e. 48/52) would correspond to only a 1.08 fold change.

Because we are interested in general trends, a *P*-value cutoff of 0.05 was used as the criterion for rejecting the null hypothesis, with no correction for multiple tests. Note that, by treating replicate pools of flies as a random effect, the test is more conservative than merely testing for differences in the means of the observed samples. In our data, 41% of 101 significant tests (150 tests total) have *P* values between 0.05 and the Bonferroni-adjusted cutoff of 0.00033 (31 genes). Pyrosequencing appears to be better able to detect small differences than microarrays, probably

because of the large number of replicates or perhaps because DNA microarrays are inherently noisier. For instance, some genes that were not differentially expressed between species based in the microarray data could be shown to be differentially expressed using pyrosequencing. Because the microarrays were developed for *D. melanogaster*, microarray comparisons of *D. melanogaster* and *D. simulans* may be slightly biased toward *D. melanogaster* depending on the degree of divergence in the coding sequence of the gene studied. Pyrosequencing is unlikely to suffer from this bias, because differential PCR efficiency of the *melanogaster* and *simulans* alleles is controlled by the DNA amplifications. Pyrosequencing and microarray results were nevertheless in good agreement ($r = 0.8$, $P < 1 \times 10^{-5}$). We therefore used microarray measurements from Ranz *et al.* (2004) to compare the expression level of the hybrid relative to the parents, and pyrosequencing to measure the expression difference between species and the differential allelic expression in hybrids for consistency of measurements across the genes studied.

RESULTS

Pattern of gene and allelic expression. Relative allelic expression between *D. melanogaster* and *D. simulans* in their F1 hybrid are shown in Figure 1. We hypothesized that compensatory *cis-trans* regulatory evolution within species should often result in misexpression in the interspecific hybrid owing to incompatibilities in the regulatory systems. Based on this hypothesis, we compared allele-specific expression with total level of expression for 31 genes in *D. melanogaster*, *D. simulans*, and their interspecific hybrid. The goal was to determine what fraction of genes, if any, that exhibited misexpression in F1 hybrids showed a pattern of allelic

expression consistent with *cis-trans* compensatory regulation. This result would be expected for genes that had undergone compensatory *cis-trans* regulatory evolution in at least one of the parental lineages.

Specifically, pyrosequencing was used to examine whether each gene showed (1) differential expression in the parents (*P* ratio ≠ 1), (2) differential expression in the hybrid (*H* ratio ≠ 1), or (3) a difference in the ratio of allelic expression in the parental lines relative to that in the hybrid (*P* ratio ≠ *H* ratio). Different patterns of regulatory evolution predict different results for these three tests. We distinguished five patterns:

- "*conserved*" No change in regulatory functions predicts that *P* ratio = 1, *H* ratio =1, and *P* ratio = *H* ratio.
- "*cis*" A change in only *cis*-regulatory elements predicts that *P* ratio ≠ 1, *H* ratio ≠ 1, and *P* ratio = *H* ratio.
- "*trans*" A change in only *trans*-regulatory elements predicts that *P* ratio ≠ 1, *H* ratio = 1, and *P* ratio ≠ *H* ratio.
- "*cis* + *trans*" A change in both *cis*- and *trans*- regulatory elements predicts that *H* ratio ≠ 1 and *P* ratio ≠ *H* ratio. In a subset of these genes, the ratio of expression levels in the parents is greater than the ratio of allele-specific expression in the hybrids. Changes in *cis* and *trans* contribute to changes in gene expression between species in the same direction.
- "*cis* × *trans*" In another subset of genes with *H* ratio ≠ 1 and *P* ratio ≠ *H* ratio, the ratio of expression levels in the parents is less extreme, or in the opposite direction

from, the ratio of allele-specific expression in the hybrids. These cases are cases of *cis-trans* compensatory regulation.

In addition to their regulatory evolution being classified as either *conserved*, *cis*, *trans*, *cis* + *trans* or *cis* × *trans*, the genes can be classified orthogonally after GIBSON *et al*. (2004) according to their overall level of expression in the hybrid as assayed either through microarray analysis (RANZ *et al*. 2004) (Figure S2) or rt-qPCR measurements (Figure S3). These categories are as follows:

- "*no change*" The overall level of expression in the hybrid and both parents are equal.
- "*dominant*" The overall level of expression in the hybrid is equal to that of one of the parents.
- "*semidominant*" The overall level of expression in the hybrid is intermediate between that of the parents.
- "*underexpressed"* The overall level of expression in the hybrid is lower than that of both parents.
- "*overexpressed"* The overall level of expression in the hybrid is greater than that of both parents.

Together, these classifications produce the matrix shown in Table 1, where each entry is the number of genes in the corresponding cell. The patterns of allelic expression are also represented in Figure 1. Based on overall level of expression in the hybrids relative to that in the parents, Table 1 includes 3 genes showing no change, 1 showing dominance, 4 showing semidominance,

and 23 showing either underexpression (5 genes) or overexpression (18 genes). Based on allele-specific expression patterns, Table 1 includes 2 genes that are *conserved*, 6 genes showing *cis*-regulatory differences only, 3 genes showing *trans*-regulatory differences only, and a total of 20 genes (almost two-thirds of the total) showing either *cis* + *trans* differences (5 genes) or *cis* × *trans* differences (15 genes). Four genes classified as *cis x trans* showed a significant *cis* regulatory effect but this effect was not statistically significant from the difference between species, which was not different from 0.

In Table 1, the genes that qualify as showing compensatory *cis x trans* regulatory evolution are: *Mth, CG14438, CG11727, CG9273, CG3775, CG5288, CG9390, CG11236, CG14629, CG32444, CG15818, CG8997, CG9338, Cyp6w1*, and *CG17608*. *Mth* shows a semidominant pattern of expression in hybrids, *CG14438, CG11727* and *CG9273* are underexpressed, and *CG17608* does not change in hybrids. The remaining ten genes were overexpressed in hybrids. In all, 10 of the 18 genes that are overexpressed and 3 of the 5 genes underexpressed in hybrids showed *cis* × *trans* regulation based on allele-specific expression and thus are cases of compensatory regulation. This represents a total of 13 out of 23 genes misexpressed in hybrids showing *cis* × *trans* regulation based on allele-specific expression. These results are not only concordant with our hypothesis that some genes would show compensating *cis*- and *trans*-regulatory differences, they indicate that a substantial proportion of regulatory changes fall into this category. Among the 10 *cis* × *trans* genes that are overexpressed in hybrids, in 7 cases (*CG11236, CG15818, CG3775, CG5288, Cyp6w1, CG8997, CG9338*) the differential allelic expression in hybrids is opposite in sign to the difference in expression in the parental species. In other words, the allele originating in the parent with the lower level of expression is expressed to

a greater level than the alternative allele in the hybrid. For example, gene *CG9338* shows a $\log_2$ ratio in expression of –0.37 between species and 0.55 in the hybrid, and the gene is 2.78 times more highly expressed in the hybrid than in the parental species (Figure S3). Another example, *CG15818*, shows $\log_2$ ratio in expression of 0.97 between species and –1.04 in the hybrid, and is overexpressed by 5.15 fold in hybrid relative to the average of the parents. Interestingly, *CG15818* is likely to be associated with spermatogenesis (electronic annotation of gene ontology, www.flybase.org) and is related to rapidly evolving genes (Holloway and Begun 2004). On the other hand, not all cases of overexpression are associated with compensatory regulation between species. To cite one example, gene *CG4847* shows a 5-fold increase in expression level in the hybrid relative to the parental species, yet there is no significant difference in the levels of allele-specific expression. This result suggests that non-additive interaction between *cis*- and *trans*- regulatory elements is unlikely to be the source of misexpression of this particular gene in the hybrid. Since pyrosequencing measures the net effect of *trans*-regulatory changes, it is possible that in the case of *CG4847*, multiple compensatory *trans*-acting changes have evolved with opposing effects on gene expression. Interactions among coevolved *trans*-changes may be just as likely to lead to misexpression in hybrids. Although our results provide strong evidence for frequent compensatory changes in *cis*-acting and *trans*-acting regulatory elements, they do not reveal the molecular mechanisms. Elucidating these mechanisms will require detailed studies of individual genes. Unfortunately, little is known about the mechanisms controlling regulation of the genes examined in this study. Nevertheless, a model describing general aspects of gene regulation suggests that compensatory *cis-trans* changes can lead to similar expression between species and under or overexpression in F1 hybrid (Appendix).

DISCUSSION

Our data provide evidence for rapid compensatory coevolution of genetic regulatory elements. A study of overall gene expression and allele-specific expression in *D. melanogaster*, *D. simulans* and their F1 hybrid revealed that, in 15 of 31 genes, more similar overall levels of expression in *D. melanogaster* and *D. simulans* is observed than expected given their *cis* regulatory divergence, resulting from *cis*-*trans* compensatory regulation. 13 of these the genes were misexpressed in the F1 hybrid. We examined a heuristic model of gene regulation and considered specific examples to explore some of the possible sources of *cis*-*trans* compensation (Appendix). Because little is known about the specific mechanisms that control the expression of the genes studied here, the direct relevance of the heuristic model cannot be evaluated, but the genes that we identified are good candidates for further analysis of regulatory divergence. Direct experimental confirmation that *cis* and *trans* regulatory elements have coevolved by compensatory changes would require (1) characterization of the genetic regulatory elements, and (2) measurements of gene expression of each parental allele in the genetic background of the other species. The key variables of the model, such as binding affinity and number of binding sites in the promoter, can be assayed (e. g., SHAW *et al.* 2002; ODA *et al.* 1998; STORMO and FIELDS 1998). Furthermore, the requisite technology exists to transfer individual Drosophila genes and their promoters between species (e. g., LAURIE *et al*. 1990; WITTKOPP *et al.* 2002). Tissue specificity of expression of these genes in the parental species and their F1 hybrids should also be examined. Because the experiments we describe involve whole-body preparations of mRNA, species-specific differences in a factor that influences organ size may appear like *trans*-acting regulation, even though the interspecific divergence has nothing to do with the kinetics of

transcription factor binding to the target promoter. Epigenetic effects or cytoplasmic-mitochondrial interactions may also act to confound the results. However, we did not examine this possibility here because the direction of the cross on the ratios of allelic expression in hybrids between *D. melanogaster* and *D. simulans* has been shown previously to have a negligible effect by WITTKOPP *et al.* (2004).

Models of gene regulation have made the clear prediction that changes in many parameters can lead to similar gene expression level, and therefore similar levels of gene expression between species can conceal genetic divergence in the regulatory elements (GIBSON 1996; VEITIA 2003). Compensatory evolution has been studied in proteins (KONDRASHOV *et al.* 2002; KULATHINAL *et al.* 2004), RNA (KERN and KONDRASHOV 2004), and in DNA-binding motifs (LUDWIG *et al.* 2000; CARTER and WAGNER 2002). In many cases, quantifying the functional effects of individual nucleotide or amino acid substitutions is difficult, but in the case of regulatory elements the level or pattern of gene expression provides a relevant assay. Only few examples of compensatory evolution within regulatory elements have been documented, and examples of *cis-trans* regulatory evolution are even rarer. LUDWIG *et al.* (2000) showed that *cis*-regulatory elements in the *even-skipped* enhancer of *D. melanogaster* and *D. pseudoobscura* have coevolved in such a way that chimeric enhancers show abnormalities in expression pattern whereas the intact elements functional apparently normally even in the genetic background of the other species. JUAREZ *et al.* (2000) have shown that amino acid substitutions in the DNA-binding domain of the transcription factor *NifA* reduce its binding affinity in *Sinorhizobium meliloti.* However, these changes have been compensated by amino acid replacements in its protein-activation domain, which has a higher affinity for the transcriptional machinery. One of

the few examples of *cis-trans* regulatory incompatibilities that have been identified is between *bicoid* and its binding sites in the *tailless* and *hunchback* promoters in *Drosophila melanogaster* and *Musca domestica*. In most interactions, the *bicoid* protein from one species has a higher affinity for binding sites from the same species, but in the case of the *Musca*, Bicoid protein, incompatibility is manifested through its higher affinity for the *tailless* promoter of both species (SHAW *et al.* 2002). Compensatory evolution can therefore occur at the *cis-cis, trans-trans* and *cis-trans* regulatory levels. The relative importance of each of those levels of interaction cannot be determined using our approach, because we concentrated on cases where there is *cis* regulatory divergence. It is also impossible to determine from the patterns of expression if the changes in *trans* affect the transcription factors themselves, such as changes in the protein sequences, or other *trans* regulatory effects such the concentration of transcription factors or other effects. The contribution of protein sequence variation of transcription factors to genetic variation in gene expression is largely unknown.

Compensatory evolution of *cis* and *trans* regulatory elements may be promoted by the abundant genetic variation in regulatory elements within species that has been reported. In the human population, for example, single-nucleotide polymorphisms and other types of genetic variation occur frequently in the core promoter region of genes and are known to have an impact on the level of expression, with effects as large as a 20-fold difference in gene expression level (reviewed in ROCKMAN and WRAY 2002; TOMSO *et al.* 2005). Genetic variation located in *trans* has also been identified as contributing extensively to gene expression variation within species (e. g., BREM *et al.* 2002; MORLEY *et al.* 2004). A recent comparative genomic study has also shown that transcription factors can evolve rapidly relative to other classes of genes

(CASTILLO-DAVIS *et al.* 2004), which was perhaps unexpected given the potentially large pleiotropic effects of mutations in transcription factors (TAUTZ 2000). Finally, experimental work has shown that amino changes in a transcription factor do not necessarily affect the expression pattern of all of its target genes, but is highly dependent on the particular DNA motifs that are present (INGA *et al.* 2002). A transcription factor can therefore coevolve with a subset of targets while leaving other interactions intact.

Our results suggest that *cis-trans* compensatory evolution occurs rapidly and on a wide scale. In the approximately 2.5 million years since the divergence of *D. melanogaster* and *D. simulans*, more than 40% of our sample of genes show evidence of compensating regulatory changes. The effects appear to be gene-specific, and not due to some generalized malfunction of gene regulation in the interspecific hybrid. Analysis of a much larger sample of genes, and of samples of genes in various gene-ontology categories, would be most welcome to determine whether the results we have observed are pervasive across the genome or are more or less prevalent in particular classes of genes. It would also be of great interest to compare pairs of species that are more closely related in order to examine how these *cis* × *trans* interactions accumulate with time. If they accumulate rapidly, it would support the theoretical prediction that regulatory interactions may play a role in the reproductive isolation between species (JOHNSON and PORTER 2000).

ACKNOWLEDGEMENTS

We would like to thank the members of the Hartl Lab for motivating discussions, and Nadia Aubin-Horth for comments on an earlier draft of the manuscript. We are grateful to the Bauer Center for Genomics Research for the use of their facilities. CRL is a Frank Knox Memorial Fellow at Harvard University and was supported during this work by graduate fellowships from NSERC (Natural Sciences and Engineering Research Council of Canada), FQRNT (Fonds Québecois de Recherche sur la Nature et les Technologies) and Fondation Desjardins du Québec. PJW is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-# 1795-03). This project was funded by NIH grants GM68465 and GM60035 to DLH and R01-AI064950 to AGC.

LITTERATURE CITED

AHMADIAN, A., B. GHARIZADEH, A. C. GUSTAFSSON, F. STERKY, P. NYREN *et al.*, 2000 Single-nucleotide polymorphism analysis by pyrosequencing. Anal Biochem **280:** 103-110.

AUGER, D. L., A. D. GRAY, T. S. REAM, A. KATO, E. H. COE, JR. *et al.*, 2005 Nonadditive gene expression in diploid and triploid hybrids of maize. Genetics **169:** 389-397.

BECKETT, D., D. S. BURZ, G. K. ACKERS and R. T. SAUER, 1993 Isolation of lambda repressor mutants with defects in cooperative operator binding. Biochemistry **32:** 9073-9079.

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

BREM, R. B., J. D. STOREY, J. WHITTLE and L. KRUGLYAK, 2005 Genetic interactions between polymorphisms that affect gene expression in yeast. Nature **436:** 701-703.

BURZ, D. S., and S. D. HANES, 2001 Isolation of mutations that disrupt cooperative DNA binding by the Drosophila bicoid protein. J Mol Biol **305:** 219-230.

CARTER, A. J., and G. P. WAGNER, 2002 Evolution of functionally conserved enhancers can be accelerated in large populations: a population-genetic model. Proc Biol Sci **269:** 953-960.

CASTILLO-DAVIS, C. I., F. A. KONDRASHOV, D. L. HARTL and R. J. KULATHINAL, 2004 The functional genomic distribution of protein divergence in two animal phyla: coevolution, genomic conflict, and constraint. Genome Res **14:** 802-811.

COMAI, L., A. MADLUNG, C. JOSEFSSON and A. TYAGI, 2003 Do the different parental 'heteromes' cause genomic shock in newly formed allopolyploids? Philos Trans R Soc Lond B Biol Sci **358:** 1149-1155.

CORTON, J. C., E. MORENO and S. A. JOHNSTON, 1998 Alterations in the GAL4 DNA-binding domain can affect transcriptional activation independent of DNA binding. J Biol Chem **273:** 13776-13780.

DENVER, D. R., K. MORRIS, J. T. STREELMAN, S. K. KIM, M. LYNCH *et al.*, 2005 The transcriptional consequences of mutation and natural selection in Caenorhabditis elegans. Nat Genet **37:** 544-548.

DOBZHANSKY, T. G., 1937 *Genetics and the origin of species*. Columbia University Press, New York.

DOVER, G. A., 1992 Observing development through evolutionary eyes: a practical approach. Bioessays **14:** 281-287.

GIBSON, G., 1996 Epistasis and pleiotropy as natural properties of transcriptional regulation. Theor Popul Biol **49:** 58-89.

GIBSON, G., R. RILEY-BERGER, L. HARSHMAN, A. KOPP, S. VACHA *et al.*, 2004 Extensive sex-specific nonadditivity of gene expression in Drosophila melanogaster. Genetics **167:** 1791-1799.

GILL, G., I. SADOWSKI and M. PTASHNE, 1990 Mutations that increase the activity of a transcriptional activator in yeast and mammalian cells. Proc Natl Acad Sci U S A **87:** 2127-2131.

HILL, D. E., I. A. HOPE, J. P. MACKE and K. STRUHL, 1986 Saturation mutagenesis of the yeast his3 regulatory site: requirements for transcriptional induction and for binding by GCN4 activator protein. Science **234:** 451-457.

HOLLOWAY, A. K., and D. J. BEGUN, 2004 Molecular evolution and population genetics of duplicated accessory gland protein genes in Drosophila. Mol Biol Evol **21:** 1625-1628.

INGA, A., F. STORICI, T. A. DARDEN and M. A. RESNICK, 2002 Differential transactivation by the p53 transcription factor is highly dependent on p53 level and promoter target sequence. Mol Cell Biol **22:** 8612-8625.

JOHNSON, N. A., and A. H. PORTER, 2000 Rapid speciation via parallel, directional selection on regulatory genetic pathways. J Theor Biol **205:** 527-542.

JUAREZ, K., H. FLORES, S. DAVILA, L. OLVERA, V. GONZALEZ *et al.*, 2000 Reciprocal domain evolution within a transactivator in a restricted sequence space. Proc Natl Acad Sci U S A **97:** 3314-3318.

KERN, A. D., and F. A. KONDRASHOV, 2004 Mechanisms and convergence of compensatory evolution in mammalian mitochondrial tRNAs. Nat Genet **36:** 1207-1212.

KONDRASHOV, A. S., S. SUNYAEV and F. A. KONDRASHOV, 2002 Dobzhansky-Muller incompatibilities in protein evolution. Proc Natl Acad Sci U S A 99: 14878-14883.

KULATHINAL, R. J., B. R. BETTENCOURT and D. L. HARTL, 2004 Compensated deleterious mutations in insect genomes. Science **306:** 1553-1554.

LAURIE, C. C., E. M. HEATH, J. W. JACOBSON and M. S. THOMSON, 1990 Genetic basis of the difference in alcohol dehydrogenase expression between Drosophila melanogaster and Drosophila simulans. Proc Natl Acad Sci U S A **87:** 9674-9678.

LEMOS, B., C. D. MEIKLEJOHN, M. CACERES and D. L. HARTL, 2005 Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. Evolution Int J Org Evolution **59:** 126-137.

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

MICHALAK, P., and M. A. NOOR, 2003 Genome-wide patterns of expression in Drosophila pure species and hybrid males. Mol Biol Evol **20:** 1070-1076.

MONTI, P., P. CAMPOMENOSI, Y. CIRIBILLI, R. IANNONE, A. INGA *et al.*, 2002 Tumour p53 mutations exhibit promoter selective dominance over wild type p53. Oncogene **21:** 1641-1648.

MORLEY, M., C. M. MOLONY, T. M. WEBER, J. L. DEVLIN, K. G. EWENS *et al.*, 2004 Genetic analysis of genome-wide variation in human gene expression. Nature **430:** 743-747.

ODA, M., K. FURUKAWA, K. OGATA, A. SARAI and H. NAKAMURA, 1998 Thermodynamics of specific and non-specific DNA binding by the c-Myb DNA-binding domain. J Mol Biol **276:** 571-590.

OHNO, S., 1969 The preferential actication of maternally derived alleles in development of interspecific hybrids, pp. 137-150 in *Heterospecific genome interaction*, edited by V. DEFENDI. Wistar Institute Press, Philadelphia.

ORR, H. A., and D. C. PRESGRAVES, 2000 Speciation by postzygotic isolation: forces, genes and molecules. Bioessays **22:** 1085-1094.

OSBORN, T. C., J. C. PIRES, J. A. BIRCHLER, D. L. AUGER, Z. J. CHEN *et al.*, 2003 Understanding mechanisms of novel gene expression in polyploids. Trends Genet **19:** 141-147.

PARKER, H. R., D. P. PHILIPP and G. S. WHITT, 1985 Gene regulatory divergence among species estimated by altered developmental patterns in interspecific hybrids. Mol Biol Evol **2:** 217-250.

PASTINEN, T., and T. J. HUDSON, 2004 Cis-acting regulatory variation in the human genome. Science **306:** 647-650.

PORTER, A. H., and N. A. JOHNSON, 2002 Speciation despite gene flow when developmental pathways evolve. Evolution Int J Org Evolution **56:** 2103-2111.

POWELL, J. R., 1997 *Progress and prospects in evolutionary biology: the Drosophila model*. Oxford University Press, New York.

PTASHNE, M., and A. GANN, 2002 *Genes & signals*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

RANZ, J. M., K. NAMGYAL, G. GIBSON and D. L. HARTL, 2004 Anomalies in the expression profile of interspecific hybrids of Drosophila melanogaster and Drosophila simulans. Genome Res **14:** 373-379.

ROCKMAN, M. V., and G. A. WRAY, 2002 Abundant raw material for cis-regulatory evolution in humans. Mol Biol Evol **19:** 1991-2004.

SHAW, P. J., N. S. WRATTEN, A. P. MCGREGOR and G. A. DOVER, 2002 Coevolution in bicoid-dependent promoters and the inception of regulatory incompatibilities among species of higher Diptera. Evol Dev **4:** 265-277.

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

STORMO, G. D., and D. S. FIELDS, 1998 Specificity, free energy and information content in protein-DNA interactions. Trends Biochem Sci **23:** 109-113.

TAUTZ, D., 2000 Evolution of transcriptional regulation. Curr Opin Genet Dev **10:** 575-579.

TOMSO, D. J., A. INGA, D. MENENDEZ, G. S. PITTMAN, M. R. CAMPBELL *et al.*, 2005 Functionally distinct polymorphic sequences in the human genome that are targets for p53 transactivation. Proc Natl Acad Sci U S A **102:** 6431-6436.

TRUE, J. R., and E. S. HAAG, 2001 Developmental system drift and flexibility in evolutionary trajectories. Evol Dev **3:** 109-119.

VEITIA, R. A., 2003 A sigmoidal transcriptional response: cooperativity, synergy and dosage effects. Biol Rev Camb Philos Soc **78:** 149-170.

WANG, J., K. ELLWOOD, A. LEHMAN, M. F. CAREY and Z. S. SHE, 1999 A mathematical model for synergistic eukaryotic gene activation. J Mol Biol **286:** 315-325.

WANG, J., L. TIAN, A. MADLUNG, H. S. LEE, M. CHEN *et al.*, 2004 Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. Genetics **167:** 1961-1973.

WITTKOPP, P. J., B. K. HAERUM and A. G. CLARK, 2004 Evolutionary changes in cis and trans gene regulation. Nature **430:** 85-88.

WITTKOPP, P. J., K. VACCARO and S. B. CARROLL, 2002 Evolution of yellow gene regulation and pigmentation in Drosophila. Curr Biol **12:** 1547-1556.

WRAY, G. A., M. W. HAHN, E. ABOUHEIF, J. P. BALHOFF, M. PIZER *et al.*, 2003 The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol **20:** 1377-1419.

YAN, H., W. YUAN, V. E. VELCULESCU, B. VOGELSTEIN and K. W. KINZLER, 2002 Allelic variation in human gene expression. Science **297:** 1143.

APPENDIX

Model of gene expression regulation. Among the factors that influence the amount of transcript produced by a gene are the concentration of transcription factors, their binding affinity for the *cis*-regulatory motifs, and the strength of the cooperativity among them. Cooperativity here refers to the facilitation of the binding of transcription factors by the previous binding of transcription factors (PTASHNE and GANN 2002). Experimental work has shown that the number of *cis*-regulatory elements, nucleotide variation in these elements, amino acid differences in DNA-binding domains of transcription factors, variation in cooperativity among transcription factors, or variation in the recruitment of the general transcription machinery can independently, but not necessarily additively, lead to changes in expression level (HILL *et al.* 1986; BECKETT *et al.* 1993; GILL *et al.* 1990; ODA *et al.* 1998; BURZ and HANES 2001; INGA *et al.* 2001; MONTI *et al.* 2002; CORTON 1998). Models of gene regulation that have included these variables (GIBSON 1996; VEITIA 2003) have made the clear prediction that changes in many parameters can lead to similar gene expression level, and therefore similar levels of gene expression between species can conceal genetic divergence in the regulatory elements. In this section we examine the qualitative behavior of a heuristic model for the regulation of gene expression in order to determine whether misexpression of a gene in a hybrid can result from simple *cis-trans* compensatory regulation within species.

The regulation of gene expression in eukaryotes can be modeled using laws of statistical thermodynamics and Boltzmann's factor (e.g. GIBSON 1996; WANG *et al.* 1999). Our model

includes variables sufficient to capture the qualitative behavior of gene expression regulation: the concentration of the transcription factor regulating the gene of interest (x, Molar); its binding affinity for the *cis*-regulatory elements of the gene it regulates, expressed as free energy ΔG_{tf} (kcal/M); the strength of cooperativity among the transcription-factor molecules, also expressed in terms of free energy ($\Delta G_{co,}$ kcal/M); and the number of binding sites for the transcription factor upstream of the gene of interest (n). We assume that the transcriptional output (R) depends on the fractional occupancy of the promoter, which refers to the proportion of time (p_σ) that the promoter spends in the particular configuration σ. The term configuration refers here to the number and arrangement of transcription factors on the binding sites in a promoter region. The probability of a given configuration p_σ will depend on the concentration of the transcription factor, the binding affinity of the transcription factor for the binding sites, the number of binding sites, and the cooperativity among the transcription factor molecules. The cooperativity among the transcription factor molecules is assumed to depend on the proximity of the sites to which they bind, and therefore the strength of cooperativity will depend on the number of nearest neighbor pairs p of transcription factors ($p \leq n - 1$). The fractional occupancy of a configuration σ then becomes:

$$p_\sigma = \frac{b^p A^n}{Z} \qquad [1]$$

Here, $Z = \Sigma_\sigma p$ serves to normalize the probabilities to 1, and

$$b = \exp(-\Delta G_{co}/RT) \qquad [2]$$

$$A = [x] \exp(-\Delta G_{tf}/RT), \qquad [3]$$

where R is the ideal gas constant and T the absolute temperature, RT = 0.59 kcal at 25 °C.

Once the fractional occupancy of each configuration is computed, we can assign to each a transcriptional response r_σ and sum over all the configurations to obtain the expression level *R*.

$$R = R_0 \Sigma_\sigma r_\sigma p_\sigma \quad [4]$$

We will assume that the R_0 does not differ between species, and can be observed as the transcriptional output given that the gene is transcribed. The transcriptional response function r_σ may depend on the number of sites bound (e.g. GIBSON 1996) or the binding affinity of the transcription factors for the transcriptional machinery. For simplicity, we will let r_σ be equivalent between species and set $r_\sigma = 1$; this corresponds to transcription taking place whenever the promoter is bound. Accordingly, the transcriptional response R takes the following form:

$$R \sim R_0 \frac{W_N(A, b) - 1}{W_N(A, b)} \quad [5]$$

Where W_N is defined by:

$$W_n(\alpha, \beta) = \sum_n \alpha^n \beta^p \quad [6]$$

We can now imagine two species that can be hybridized to form an F1 hybrid. The species may differ in the values of the parameters *x*, *n*, ΔG_{tf}, or ΔG_{co}. When necessary, we will distinguish the parameters of the two species by unprimed and primed symbols. A difference in *x* would correspond to divergence in the level of expression of the transcription factor (*trans*); *n* to divergence in the number of binding sites upstream of the gene (*cis*); ΔG_{tf} to divergence in the affinity of the transcription factor for the binding sites [which can result from mutation of the transcription factor protein (*trans*) or the DNA binding sites (*cis*)]; and ΔG_{co} to divergence of the transcription factor molecules that changes their cooperativity (*trans*), or a difference in the distance among the binding sites (*cis*) that would produce a similar effect. The total expression

level in the hybrid will be the sum of the expression level of the two alleles. This will depend on the fractional occupancies of the two promoters, which can be deduced on the assumption that the transcription factors are equally available to each promoter.

Species that can hybridize are generally closely related. For example, the *Drosophila* species examined here diverged 2.5 Myr ago (~ 5% nucleotide divergence) (POWELL 1997). It is therefore reasonable to assume mutational changes affecting only a minimal number of parameters necessary for compensatory changes, say, two. Divergence only in *cis*-acting factors, through different numbers of binding sites $n' \neq n$, or different binding affinity of the binding sites ($A' \neq A$) or distance between binding sites resulting in changes in binding cooperativity ($\Delta G_{co} \neq \Delta G_{co}$' and therefore $b' \neq b$), leads to a similar difference in expression level between species and between alleles in the hybrid, and leads to an intermediate gene expression level in the hybrid (R_h):

$$R_h = [1/2 \frac{W_{N'}(A',b')-1}{W_{N'}(A',b')} + 1/2 \frac{W_N(A,b)-1}{W_N(A,b)}] = (1/2)R + (1/2)R' \qquad [7]$$

$$D_{sp} = D_{al} = \frac{W_{N'}(A',b')-1}{W_{N'}(A',b')} / \frac{W_N(A,b)-1}{W_N(A,b)} \qquad [8]$$

However, changes only in *trans*-acting factors ($A' > A$, $\Delta G_{tf}' < \Delta G_{tf}$ or $x' > x$) lead to differential expression between parents but not between alleles in the hybrid. Note that in this case, contrary to the previous case where a *cis*-acting change was examined, change in A represents a change in *trans* so that it is consistent with the case above:

$$R_h = \tfrac{1}{2} R_0 \left(\frac{W_N(\frac{1}{2}(A'+A),b)-1}{W_N(\frac{1}{2}(A'+A),b)} + \frac{W_N(\frac{1}{2}(A'+A),b)-1}{W_N(\frac{1}{2}(A'+A),b)} \right) \quad [9]$$

$$D_{sp} = \frac{W_N(A',b)-1}{W_N(A',b)} / \frac{W_N(A,b)-1}{W_N(A,b)} > 1 \quad [10]$$

Note that, because the promoters are functionally equivalent,

$$D_{al} = \frac{W_N(\frac{1}{2}(A'+A),b)-1}{W_N(\frac{1}{2}(A'+A),b)} / \frac{W_N(\frac{1}{2}(A'+A),b)-1}{W_N(\frac{1}{2}(A'+A),b)} = 1 \quad [11]$$

In this case:

$$R < R_h \leq R' \quad [12]$$

R_h can be as large as R', for example, if $x' > x$, so that one species is closer to saturation, for instance in cases where most TF binding sites are occupied by transcription factors. The ratio of allelic expression in the hybrid is nevertheless expected to be 1 because the promoters are functionally indistinguishable.

Our main interest is cases in which there is *cis*-regulatory divergence between alleles ($D_{al} \neq 1$) but the expression level between species (D_{sp}) is compensated so that the difference in expression between the parental species is either (1) smaller than that between alleles in the hybrid background, or (2) in the opposite direction. This can be achieved in many ways, but we concentrate on cases in which regulatory compensation is brought about by differences in the number of binding sites (n, a *cis* effect) balanced against differences in the expression level or binding affinity of the transcription factor (A, a *trans* effect). Other scenarios can readily be

imagined (e. g., SHAW *et al.* 2002), but we focus on this mechanism because it can evidently lead to a rapid increase in genetic variation (STONE and WRAY 2001).

Consider therefore the consequences of the assumptions that $R = R'$, $n' < n$, and $A' > A$. The expression level in hybrids is then:

$$R_h \sim \tfrac{1}{2} R_0 \left(\frac{W_N(\frac{1}{2}(A+A'),b)-1}{W_N(\frac{1}{2}(A+A'),b)} + \frac{W_{N'}(\frac{1}{2}(A+A'),b)-1}{W_{N'}(\frac{1}{2}(A+A'),b)} \right) \quad [13]$$

Because $(W_N - 1)/W_N > (W_{N'} - 1)/W_{N'}$ follows from $A' > A$,

$$R_h \geq \frac{1}{2} R \quad [14]$$

$$D_{al} = ((W_{N'} - 1)/W_{N'}) \,/\, ((W_N - 1)/W_N) < 1 \quad [15]$$

$$D_{sp} = R/R' = 1 \quad [16]$$

However, if the regulatory systems are near saturation for the transcription factor, then $(W_N - 1)/W_N \approx (W_{N'} - 1)/W_{N'}$ and therefore

$$R/2 \leq R_h \leq R \quad [17]$$

In other words, gene expression in hybrids can be lower than that in the parents despite the conserved level of expression between the parents. On the other hand, if the transcription factor in one species is at a concentration much smaller than saturation, then, for example, $(W_N - 1)/W_N$ will be small and R_h can exceed R, and the ratio can be as large as

$$R_h / R \leq \frac{1}{2(W_N(A,b)-1))} \quad [18]$$

A specific example is presented in Figure S4 where the affinity of the transcription factor for the binding sites in one species compensates for the difference in the number of binding sites. Although the gene is expressed at the same level in both species, D_{sp}= 1, the value of $D_{al} \neq 1$ and the expression level of the gene in the hybrid is greater than that in the parental species. Through *cis-trans* compensatory regulation, therefore, the level of gene expression can be conserved between species but the gene will be overexpressed in the hybrid

TABLE

Table 1. Gene classification according to the overall level of expression in hybrids relative to parents and according to the relative level of allelic expression in the hybrids and in the parents.

	Allelic divergence in hybrid background [1]				
Hybrid relative to parental species [2]	*Cis*	*Trans*	*Cis + Trans*	*Cis × Trans*	*Conserved*
Dominant	1				
Semidominant			3	1	
Underexpressed		1		3	1
Overexpressed	3	2	2	10	1
No change	2			1	

1: As assayed by pyrosequencing in *D.melanogaster*/*D.simulans* pools and in their F1 hybrids.

2: According to microarray and quantitative real-time PCR (RANZ *et al.* 2004 and by this work)

FIGURE LEGENDS

FIGURE 1: Ratios of allelic expression ($\log_2$) between species and in the hybrid background for the 31 genes studied. Error bars indicate 95% confidence intervals. Genes are grouped by the inferred patterns of divergence. *Mbc*, *fbp*, *CG8539*, *CG1558*8 and *CG936*0 show divergence in *cis*. *CG7670*, *CG15582* and *CG18228* show divergence in *trans*. *CG8232*, *Cyt-b5-r*, *CG15814*, *Nmdmc*, *torp4a* and *CG4716* show divergence in *cis* and *trans* (*cis* + *trans*). The remainder, except *mus209* and *CG4847*, show the *cis* × *trans* pattern of divergence. In four of these cases (*CG14629*, *CG9390*, *CG11727*, *CG17608*) there was significant *cis* divergence but this divergence was not significantly different from the species divergence, which itself is not different from 0. *CG4847* shows neither divergence between species nor *cis* divergence. *Mus209* shows divergence between species but no significant *trans* effect as detailed in the model. Genes studied previously by Wittkopp et al. (2004) are indicated with *.

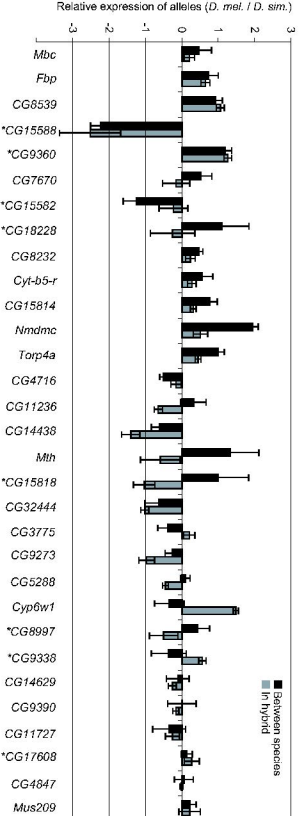
SUPPLEMENTARY FIGURE LEGENDS

FIGURE S1: Overview of the experimental design used to estimate allele specific gene expression using (A) pyrosequencing and (B) total gene expression level using quantitative real-time PCR.

FIGURE S2: Measures of gene expression level in females of *Drosophila melanogaster* Canton S and *Drosophila simulans* Sim 1 and their F1 hybrid as measured with microarrays (data from RANZ et al. 2004). Bars indicate 95% credible intervals.

FIGURE S3: Measures of gene expression level in females of *Drosophila melanogaster* Zhr and *Drosophila simulans* Tsim and their F1 hybrid as measured with quantitative real-time PCR. Bars indicate 95% confidence intervals.

FIGURE S4: Numerical example of compensatory *cis* and *trans* divergence (*cis x trans*). *q* has 3 binding sites, $\Delta G_{tf} = -10$ kcal/M; *q´* has 1 binding site; $\Delta G_{tf'} = -11$ kcal/M; TF concentration 6.52E–9M.


Relative expression of alleles (D. mel. / D. sim.)
-3
-2
-1
0
1
2
3
Mbc
Fbp
CG8539
*CG15588
*CG9360
CG7670
*CG15582
*CG18228
CG8232
Cyt-b5-r
CG15814
Nmdmc
Torp4a
CG4716
CG11236
CG14438
Mth
*CG15818
CG32444
CG3775
CG9273
CG5288
Cyp6w1
*CG8997
*CG9338
CG14629
CG9390
CG11727
*CG17608
CG4847
Mus209
Between species
In hybrid