Genomic Inferences of the *cis*-Regulatory Nucleotide Polymorphisms Underlying Gene Expression Differences between *Drosophila melanogaster* Mating Races

Naoki Osada,[*][1] *Michael H. Kohn,*[†][1] *and Chung-I Wu*[‡]

*Division of Biomedical Research Resources, National Institute of Biomedical Innovation, Ibaraki, Osaka, Japan; †Department of Ecology and Evolutionary Biology, Rice University; and ‡Department of Ecology and Evolution, University of Chicago

Nucleotide sequence polymorphisms affecting gene expression occur in the regulatory region of genes (in *cis*) and elsewhere in the genome (in *trans*). Further study is required to weigh the relative importance of *cis*- and *trans*-acting mutations in mediating gene expression differences within and between species. Here, microarray hybridization experiments were used to isolate 363 gene expression differences between the female fly head transcriptomes of 2 *Drosophila melanogaster* strains. One strain (French) represented the cosmopolitan M mating race and the other strain (ZS30) represented the Z mating race derived from Zimbabwe, Africa. From chromosomal substitution strains engineered from the 2 strains, we inferred that the expression differences between M and Z alleles largely could be attributed to the genotype of the chromosomes where the differentially expressed genes were located, that is, *cis*-regulatory polymorphisms prominently influence gene expression differences between M and Z. The effects of *trans*-regulatory polymorphisms were apparent yet difficult to quantify. Results have implications for models of gene regulatory evolution as well as experimental studies trying to identify the nucleotide sequence polymorphisms underlying gene expression differences between *Drosophila* strains.

Introduction

Microarray hybridization experiments have begun to unveil variability in gene expression within and among natural populations and species (Primig et al. 2000; Jin et al. 2001; Oleksiak et al. 2002; Schadt et al. 2003; Yvert et al. 2003; Fay et al. 2004; Nuzhdin et al. 2004). Once gene expression differences are identified, a question that poses itself concerns their direct (i.e., functional) link to DNA mutations (e.g., Fay et al. 2004; Wittkopp 2005). DNA polymorphisms occur in the *cis*-regulatory regions of *Drosophila* genes and, thus, could provide a rich source for gene expression variability (e.g., Kohn et al. 2004). Presently, the question whether gene expression differences predominantly are due to *cis*- or *trans*-acting mutations remains to be broadly studied and currently is under investigation (Schadt et al. 2003; Yvert et al. 2003; Wittkopp et al. 2004). Whether this varies between taxonomic groups and is a function of evolutionary distance (e.g., polymorphism vs. divergence) also remains to be examined. Akin to these questions, we wish to know whether *cis*-regulatory mutations play a role in the evolution of the phenotype that is sufficiently important to allow for progress in their more routine molecular genetic dissection. *Cis*-regulatory mutations are more readily studied at the molecular level compared with *trans*-regulatory mutations. The latter, however relevant, are difficult to pursue at the molecular level.

We conduct an analysis of gene expression that is set within the context of the differentiation between Z strains of *Drosophila melanogaster*, from Zimbabwe, Africa, and M strains of cosmopolitan and African descent (Wu et al. 1995; Hollocher et al. 1997; Ting et al. 2001). The racial differentiation between Z and M strains is manifested as a bias in apparent female choice, where Z females mate preferentially with Z males but M females mate with both Z- and M-type males. Other characteristics appear to differ between M and Z as well, including hydrocarbon profile, cold tolerance and starvation resistance, and fecundity (Takahashi et al. 2001; Greenberg et al. 2003). The bias in mate choice could be a by-product of the differentiation at traits and genes unrelated to behavior (Greenberg et al. 2003). The potential importance of *cis*-regulatory mutation in the differentiation between the Z and M strains is beginning to emerge from the detailed study of the *desaturase*2 locus (Takahashi et al. 2001; Fang et al. 2002; Greenberg et al. 2005). There, one specific mutation (a 16-bp deletion) in the *desaturase*2 promoter in the M strains abolishes the expression of the gene, and nucleotide polymorphism data bear the signature of a recent selective sweep. Presumably, numerous gene expression differences could promote the differentiation between Z and M. The differentiation between M and Z is a phenomenon that potentially offers important insights in the process of race formation and speciation.

To evaluate the relative dependence of transcription on *cis*- and *trans*-regulatory polymorphism, individual chromosomes from one fly strain are shuttled into the genomic background of a second one with opposite genotype, thereby generating chromosome substitution strains. In *D. melanogaster*, 6 chromosome substitution strains can be generated this way from the 2 strains ZS30 (Z strain) and French (M strain), ignoring the small fourth chromosome: MZZ, ZMZ, ZZM, MMZ, MZM, and ZZM, where the first, second, and third letter represents the homozygous genotype of the X, second, and third chromosomes, respectively (Hollocher et al. 1997). Once the engineered strains are assayed for gene expression, it is possible to ask whether any gene expression differences originally observed between the parental strains, M and Z, are recapitulated in the appropriate comparison between substitution strains, indicating that *cis*-regulatory control of the expression changes are quite independent of the genomic background (i.e., are not overwhelmed by *trans*-effects contributed

by the remaining 2 chromosomes). Thus, it should be feasible to infer the significance of *cis*- and *trans*-regulatory polymorphism underlying gene expression changes between M and Z alleles at a genome-wide scale.

Materials and Methods
Fly Strains

One *D. melanogaster* isofemale strain from Zimbabwe (ZS30; Z) and 1 from France (Fr, M) were analyzed, as well as 6 chromosomal substitution strains, MZZ, ZMZ, ZZM, MMZ, MZM, and ZMM engineered from them as described in Hollocher et al. (1997). The genotypes of the parental strains Fr and ZS30 strains are MMM and ZZZ, respectively. M indicates homozygosity for M alleles (derived from the French strain) at loci on the X, second, or third chromosomes. Correspondingly, Z indicates homozygosity for Z alleles (derived from the ZS30 strain).

RNA Isolation and Microarray Hybridization

Two replicate microarray experiments each for the Fr (MMM) and ZS30 (ZZZ) strains were done (array sets 1 and 2). Array set 2 and the chromosome substitution strains (array set 3) were processed together. A total of 60–100 female flies that were 4- to 6-day-old were starved for 1 h prior to sedation on ice. Heads were dislodged from the bodies using sieves after flies were snap frozen in liquid nitrogen. RNA was extracted with the TRIzol Reagent (GIBCO BRL, Gaithersburg, MD), purified with phenol and chloroform and precipitated with ethanol (array set 1). Alternatively, the TRIzol extraction was followed by isopropanol precipitation, after which the resuspended RNA was purified with phenol and chloroform (array sets 2 and 3). cDNA synthesis, labeling, and hybridization to the *Drosophila* Genome Array 1.0 were done according to the manufacturer's instructions (Affymetrix Inc., Santa Clara, CA). Gene expression omnibus (http://www.ncbi.nlm.nih.gov/geo/) IDs for each of the 2 strains and microarray experiments are given in parentheses: Fr (array set 1: GSM29579 and array set 2: GSM29581), ZS30 (GSM29584 and GSM29585), MMZ (GSM29586), MZM (GSM29587), MZZ (GSM29588), ZMM (GSM29589), ZMZ (GSM29590), and ZZM (GSM29591).

Analysis of Gene Expression Data

First, normalization and evaluation of signal intensity were performed by using the perfect match mismatch (PM-MM) analysis implemented in the Affymetrix Software Suite (MAS program). The presence or absence of probe sets on each of the microarrays to which the parental strains were hybridized was evaluated at $\alpha = 0.05$ (α is equivalent to the detection *P* value). The 6189 (44.3%) probe sets that were called absent on these 4 arrays were removed from the analysis, and the 7777 genes that were called present on at least 1 of the 4 arrays were retained. These may be considered the upper estimate of genes expressed in the female fly head in either the Fr or ZS30 strain, or both. Second, the expression differences between Fr and ZS30 on the first set of arrays were evaluated at $\alpha=0.05$ using signed-ranks

tests (Affymetrix Software Suite). The second set of arrays comparing Fr with ZS30 was analyzed in the same fashion. Out of the 7777 probe sets, 1981 (25.5%) were called significantly different on at least 1 of the 2 sets of arrays comparing Fr and ZS30. Of these, 363 (4.7%) were called in the same direction on both sets of arrays. The resulting significance of $\log_2$ ratios between Fr and ZS30 all were below $\alpha \leq 0.004$, and 80% of these were below $\alpha \leq 0.0005$ (Supplementary Table 1, Supplementary Material online). Map location and gene function (Gene Ontology [GO] categories) were based on the NetAffx *Drosophila* annotation (Liu et al. 2003). Third, as outlined below, for the genes that were differentially expressed between Fr and ZS30, inferences were made concerning them being under the control of *cis*- and *trans*-regulatory nucleotide sequence polymorphism.

Inference of *cis*- and *trans*-Regulation of Gene Expression Differences

Using various combinations of the chromosome substitution strains (fig. 1, left panel), we evaluated the significance of the R^2 values of linear regressions of the $\log_2$ ratios between M and Z alleles in the genomic background of substitution strains on the $\log_2$ ratios between M and Z alleles in the parental strains. If gene expression differences between Z and M were due to *cis*-regulatory nucleotide site polymorphisms, then we would expect these regressions to be significant. Conversely, if *trans*-mutations predominantly would cause the gene expression differences, then such linear regressions should not be significant. Significance was deduced by the fraction of observed R^2 values that were equal or greater than the R^2 values calculated after randomizing the $\log_2$ ratio data 1000 times.

Specifically, let the expression value of a gene in a particular genomic background be E_G. For a gene located on the X chromosome that is upregulated in MMM compared with ZZZ, we write $E_{MMM} > E_{ZZZ}$; for a downregulated gene, we write $E_{MMM} < E_{ZZZ}$. For an upregulated gene on the X chromosome that is due to *cis*-regulatory mutations, we expect that $E_{MMZ} + E_{MZM} > E_{ZMZ} + E_{ZZM}$ (cf., fig. 1, right panel, gray-shaded plots). The "+" indicates that the $\log_2$ ratios were deduced from the combination of array data. Conversely, for an expression difference on the X chromosome that is due to *trans*-regulatory mutations, for example, on the second chromosome, we expect $E_{ZMM} + E_{MMZ} > E_{ZZM} + E_{MZZ}$ and so on (cf., fig. 1, right panel, open plots). The same logic holds for any other comparison between chromosomes. Note how the grouping of substitution strains should statistically remove the expression differences between M and Z alleles on 2 of the 3 chromosomes.

Quantitative Real-Time PCR

PRIMER 3-Software/PCR-Suite (van Baren and Heutink 2004) was used to design primers for 42 of the genes that were differentially expressed between Z and M on the arrays (Supplement Table 2, Supplementary Material online). Each quantitative real-time polymerase chain reaction (qRT-PCR) included a no template control and were run on an ABI Prism 7000 Sequence Detection System (Applied Biosystems, Foster City, CA) for 40 cycles
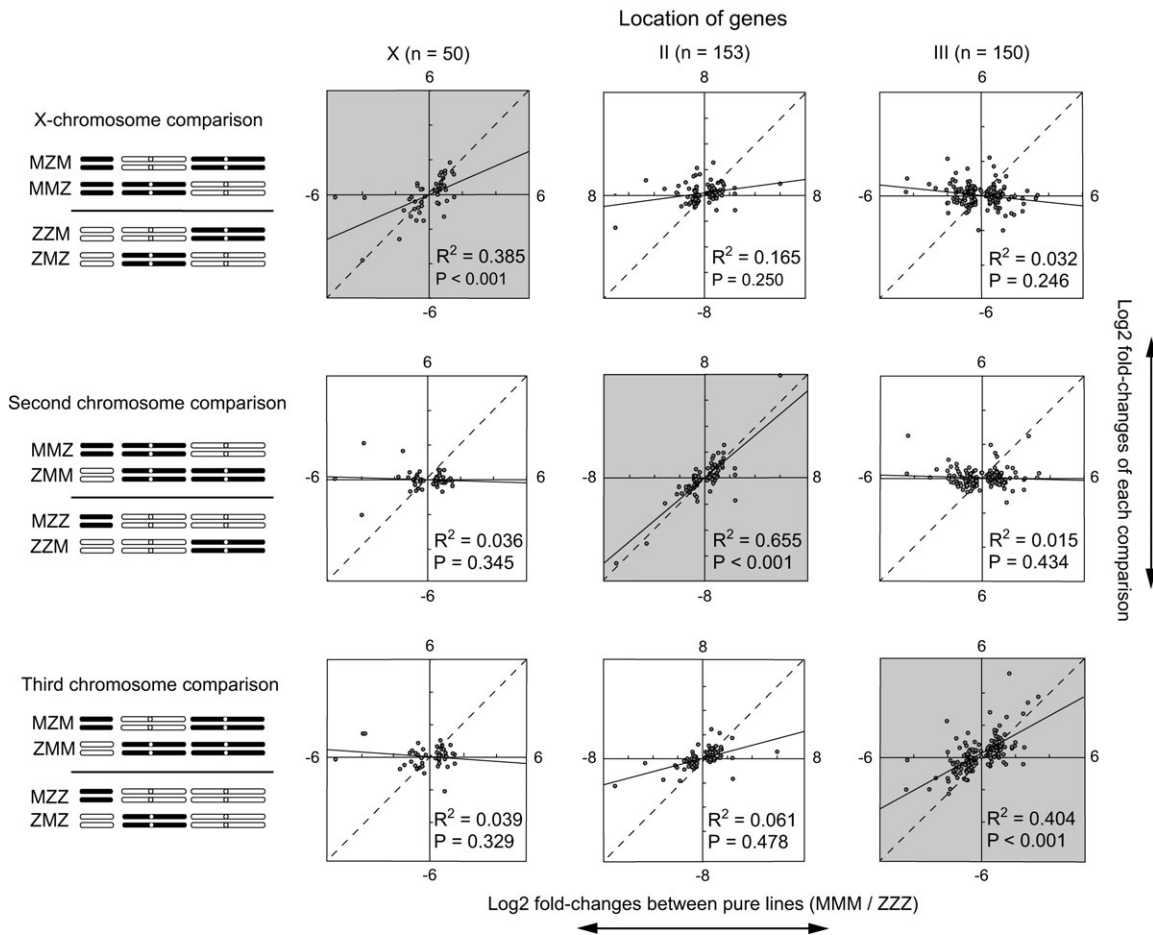
Fig. 1.—Gene expression differentiation between ZS30 and Fr and the inference of *cis*-regulatory nucleotide polymorphism. Left panel: experimental grouping of chromosome substitution strains to study the X (top), second (middle), and third chromosome (bottom). Right panel: for 51 genes on the X chromosome, the relative expression level of M$\underline{M}$Z and M$\underline{Z}$M, over those of Z$\underline{M}$Z and Z$\underline{Z}$M (top), were compared with the relative expression level of MMM, over that of ZZZ. Similarly, for 153 genes on the second chromosome, the relative expression level of M$\underline{M}$Z, Z$\underline{M}$M, over those of M$\underline{Z}$Z, Z$\underline{Z}$M (middle), were compared with the relative expression level of MMM, over ZZZ. Finally, for 150 genes on the third chromosome, the relative expression levels of MZ$\underline{M}$, ZM$\underline{M}$, over those of MZ$\underline{Z}$, ZM$\underline{Z}$ (bottom), were compared with the relative expression level of MMM, over ZZZ. The *x*-axis gives the log$_2$-fold expression difference between the parental lines (i.e., log$_2$ [E_{MMM}/E_{ZZZ}]). The *y*-axis gives the log$_2$-fold expression difference in the engineered genomic backgrounds of the substitution strains (i.e., log$_2$ [(E_{MMZ} + E_{MZM})/(E_{ZMZ} + E_{ZZM})], etc.). R^2 values were deduced from simulations (see text). Regression lines with a slope equal to 1 are depicted as dashed lines. Note that even though each substitution strain was assayed on arrays only once, the grouping of substitution strains for analysis resulted in replicated data.

with the following profile: 94 °C 2 min, 94 °C 10 s, 55 °C 45 s, and 72 °C 45 s. Standard 25-μl PCR reactions contained 0.1 μl cDNA, 1.5 mM MgCl$_2$, and SYBR Green (1.25 μl of a 1:10 000 dilution; Stratagene, La Jolla, CA). Cycle thresholds (CTs) exceeding 35 were discarded because qRT-PCR is saturated thereafter. CT numbers for a control gene (Actin 57B; CG10067) were adjusted such that the CT difference between strains was set as zero. CT values for each gene were then adjusted using the multiplication factor used to adjust Actin, and adjusted CT differences between M and Z were calculated as CT(M) minus CT(Z).

Results
Gene Expression Differences Between the Parental Strains

Based on replicate microarray hybridizations, a minimum set of 363 probe sets differentially expressed in MMM and ZZZ was identified. Roughly 60% of genes had an average (midpoint) log$_2$ ratio <1, 40% had a log$_2$ ratio >1, and about 10% of genes had a log$_2$ ratio >2 (Supplementary Table 1, Supplementary Material online).

A 50, 153, and 150 probe sets were assigned to the X, second, and third chromosomes, respectively. Ten were assigned to the fourth chromosome, or the map location was ambiguous. The ratio of 50:153:150 was compatible with the expectation based on the total number of genes on each chromosome (chi-square test, P = 0.289) and the number of genes on each chromosome that were expressed in the head (P = 0.457; table 1).

Various pathways involved in signal transduction, cell communication, phototransduction, and a range of others that potentially could be affected by transcriptional change. If these were adjusted to the expectations based on the frequency of each GO term in the fly genome, a range of biological processes were overrepresented (Supplementary Table 3, Supplementary Material online).

Table 1
Number of Genes Expressed in the Fly Female Head and the Number of Genes Differentially Expressed between Z and M

	X	II	III	IV and Unlocalized	Total
All probe sets[a]	2331	5086	6035	558	14010
Expressed in the head[b]	1219	2865	3301	392	7777
Differentially expressed between ZS30 and Fr[c]	50	153	150	10	363

[a] *Drosophila* array version 1.0.

[b] Upper limit of probe sets that were detected using the Affymetrix Software Suite on at least 1 of the 4 replicate microarrays using parental Z and M strains.

[c] Lower bound for the number of differentially expressed genes, as determined by overlap between 2 hybridization experiments using parental Z and M strains.

Table 2
Quantitative Real-Time PCR Analysis of Array Results

	qRT-PCR Confirms Array[a]	qRT-PCR Rejects Array[a]	N
Decrease in M	14	7	21
Increase in M	19	2	21
N (genes)	33	9	42

[a] Two independent qRT-PCR runs in accordance or conflict with direction of expression change (up- or downregulation) on array.

[b] Two-tailed Fisher's exact test for independence, $P = 0.13$ (NS).

Regulation of Transcriptional Differences: *cis* or *trans*?

We consider the effect of the X, second, and third chromosomes on the expression of X-linked, second-linked, and third-linked genes separately, resulting in 9 panels in figure 1. Let the expression value of a gene in a particular genomic background be E_G. The *x*-axis of each panel is the expression ratio of the pure M lines (E_{MMM}) over the pure Z line (E_{ZZZ}). The *y*-axis is the ratio of the average expression of 2 genotypes over that of 2 other genotypes shown to the left of the panels. For example, in the first row, the comparisons are between ($E_{MZM} + E_{MMZ}$) and ($E_{ZZM} + E_{ZMZ}$), which have the same autosomal (second and third chromosomes) background but differ in the genotype of the X chromosome (see Materials and Methods).

By using this approach, we inferred that the genotype of the chromosome had a significant effect on the expression changes of its resident genes. Specifically, significance was only seen for linear regressions of the $\log_2$ ratios of genes on the X in substitution strains and on the X in the parental strains, the $\log_2$ ratios of genes on the second chromosome in substitution strains and on the second chromosome in the parental strains, and the $\log_2$ ratios of genes on the third chromosome in substitution strains and on the third chromosome in the parental strains (fig. 1, shaded plots, $P < 0.001$ as deduced from simulations). In contrast, none of the 6 other possible linear regressions were significant (fig. 1, open plots).

These observations suggested that *cis*-regulatory mutations strongly determined the expression differences between any 2 M and Z alleles. However, first, for this statement to be justifiable, we need to assume that *trans*-acting factors do not preferentially map onto the same chromosomes as the genes they regulate, but instead, are uniformly distributed across the genome (see below). Second, the effect of *trans*-acting mutations on expression differences between Fr and ZS30 alleles also were apparent from our analysis, as was indicated by the regression lines in the shaded plots, which had slopes of less than 1 (fig. 1, shaded plots). Hence, *cis*–*trans* interactions were apparent from the data.

To what degree does each expression difference between M and Z alleles depend on another *trans*-acting factor of the same genotype? To obtain an estimate, we evaluated the number of genes whose expression values in the parental strains were recapitulated in the comparison among the substitution strains (fig. 1, shaded plots). We found that the expression of 40, 111, and 121 genes agreed with the expression difference in pure lines when placed in the randomized genomic background of the substitution strains.

These gene expression differences are unlikely to depend on *trans*-acting factors of the same genotype located on chromosomes other than the chromosome where the assayed gene maps. However, these estimates included those genes that are regulated by *trans*-acting factors located on the same chromosome. This is expected to inflate the number of *cis*-regulated genes. Assuming that the *trans*-regulatory factors are uniformly distributed over the chromosomes, *trans*-regulatory factors map onto the same chromosome as the target locus with a probability of 0.17, 0.38, and 0.45 (as was simply determined by the length of each chromosome relative to the *Drosophila* genome) for the X, second, and third chromosomes, respectively. Resulting estimates of *cis*-regulated genes are about 56%, 76%, and 65% for the X, second, and third chromosomes, respectively, and a genome-wide average of ∼68%.

Nucleotide sequence divergence between the Z and M strains might affect the efficiency of hybridization of the cDNA to the GeneChip. Expression differences that arose from such hybridization issues would be mapped in *cis*, thereby leading to an overestimate of the gene expression differences that are caused by *cis*-regulated polymorphism. However, expression differences that arose from hybridization issues would also result in a downward bias of the expression values collected for the strain that is more divergent to the fly strain that was used as template to prepare the GeneChip. Presumably the African strain ZS30 would be expected to be the more divergent strain. Thus, more of the gene expression differences between Z and M should be in the direction of reduced expression in the former compared with the latter. However, for every reduction in gene expression in Z relative to M, we find 1.04 increases in expression of Z relative to M, suggesting the sequence differences between the Z strain, and the GeneChip did not lead to a bias in expression levels measured for Z.

Furthermore, array results for 42 genes (21 upregulated and 21 downregulated) were tested with qRT-PCR. Primers for qRT-PCR were designed independently from oligo probes used for the GeneChip. For 33 genes, the array results were confirmed in both of 2 qRT-PCR runs, and for 9 genes, the qRT-PCR data were in conflict with the array results (table 2). No significant difference was found for the

observed number of confirmed genes and those with conflict with respect to whether they were up- or downregulated on the array (table 2, Fisher's exact test, 2-tailed, $P = 0.13$). Finally, 28 of the 33 (90%) gene expression differences between the pure strains that were confirmed by qRT-PCR were recapitulated in the chromosome substitution strains, indicating that these were due to *cis*-regulatory polymorphisms.

Discussion

Here we inferred that *cis*-regulatory polymorphism is a prominent cause for gene expression differences between the M and Z mating races of *D. melanogaster*, and we present a set of differentially expressed candidate genes associated with the differentiation between M and Z. The experimental design that utilized whole chromosome substitution strains enabled genome-wide inferences of *cis*-regulatory polymorphism underlying gene expression differences between strains. Thus, the approach is somewhat broader in scope than approaches based on pyrosequencing; latter is an elegant and highly sensitive approach that can detect *cis*-regulatory mutations at individual candidate genes (cf., Wittkopp et al. 2004) and, as is, the pyrosequencing approach requires the analysis of F1 hybrids and cannot be applied to the whole genome. In turn, our genome-wide approach is limited to the whole-chromosome resolution (but see below).

Cis-Regulatory Polymorphism Determines Differentiation in Gene Expression between M and Z Alleles

Both *cis*- and *trans*-acting elements and factors commonly control the expression of genes. This is a fundamental result obtained from manifold studies of numerous individual genes over the modern era of molecular biology (e.g., Carroll et al. 2001). The evolution of gene regulation thus can either be due to mutations in the *cis*-regulatory regions of genes or in the *trans*-regulatory factors, or both. Here, we inferred for *Drosophila* that polymorphism in the *cis*-regulatory regions of genes plays a prominent role in the differentiation of gene expression and racial differentiation between the Z and M mating races of *D. melanogaster*. Specifically, gene expression differences between M and Z alleles generally mapped onto the chromosome where the differentially expressed gene resides. This was inferred from the significant correlations of the $\log_2$ ratios between the parental strains and the $\log_2$ ratios between various combinations of chromosome substitution strains (fig. 1).

Moreover, by measuring gene expression differences between M and Z alleles when individual chromosomes were placed in a opposite genomic background, we inferred that a substantial fraction (~68%) of expression differences were due to *cis*-regulatory polymorphism. The interpretation of *cis*-regulation depends on the validity of the assumption that *trans*-factors do not preferentially colocalize on the same chromosome as the genes they regulated in *trans*, but instead, are distributed uniformly (i.e., according to the size of each chromosome) across the genome.

Our inferences of the *cis*-regulatory polymorphism do not appear to be a false positive result. First, we would expect to observe positive correlations in some of the plots where the interactions between chromosomes were examined (fig. 1, open plots). Second, the qRT-PCR data indicate that our inferences of *cis*-regulatory polymorphisms unlikely were caused by nucleotide sequence divergence between the Z and M strains that would cause differential hybridization of the cDNA to the GeneChips (table 2).

Establishing the importance of *cis*-regulatory polymorphism does not require diminishing the role of *trans*-regulatory polymorphism. The latter, albeit less prominent in its effects according to our results, nonetheless could account for some of the variance in figure 1. As is evident from figure 1, none of the expression differences between the M and Z alleles in the parental strain comparisons were recapitulated perfectly (the slopes of regression lines were <1). In addition to experimental noise, this is because *cis*-effects in this study in fact are the sum of *cis*- and *trans*-effects on the same chromosome. Moreover, the effects of *trans*-factors and/or *cis*–*trans* interactions are consistent with our data. In that sense, the possibility remains that *trans*-regulation possibly is widespread but in absolute terms is difficult to quantify at the resolution of this study.

Candidate Genes Underlying Z/M Differentiation

A minimum set of 363 gene expression differences between the French and ZS30 strains was isolated. Their value as candidate genes associated with the racial differentiation between M and Z has to be viewed with the following points in mind.

First, emerging premating isolation by apparent female choice is one of the intriguing evolutionary genetic features associated with the racial differentiation between M and Z strains of *D. melanogaster*. To localize our search, we used female fly heads as our material. Evolution in males can elicit evolutionary change of neurological features in females (Coyne and Orr 2004) and potentially affect the recognition of male characters, such as genitalia, song, or courtship display. Moreover, it seems plausible that male reproductive genes from the accessory gland target some of the genes in the female nervous system (MacGraw et al. 2004). However, the M and Z races differ at other characteristics as well, such that candidate gene function not necessarily is related to neurology.

Second, each of the ZS30 and Fr strains was assayed twice on microarrays. Because experiments were independent and notable variances in expression levels were associated with them, we did not merge arrays for statistical analysis. The merging of the 2 sets of arrays for statistical analysis with sequential *t*-tests as implemented in the dCHIP software (Li and Hung Wong 2001) provided similar results however (not shown). Moreover, the recapitulation of expression differences for the majority of genes in the substitution strains, in fact, can be considered as further replicates (fig. 1).

Third, 2 strains were used to isolate candidate genes. Thus, although any 2 African and non-African strains may differ from each other at as many as 363–1981 genes

expressed in the female fly head, none of these expression differences may be diagnostic for the Z and M mating races. We therefore abstain from a deeper discussion of candidate gene functions and their potential link to the M/Z racial differentiation based on the available data. However, inspection of Supplement Table 3 illustrates that any 2 *D. melanogaster* strains may vary at genes with potential effects on a range of biological processes, cellular components, and molecular functions.

Finally, qRT-PCR results suggest that a number of gene expression differences (up to ~21%) between Z and M could be false-positives (table 2), and further detailed expression analyses of individual candidate genes should be done.

Conclusion

Conflicting results regarding the relative importance of *cis*- versus *trans*-regulatory mutation reported by previous studies could in part be due to the criteria applied to assess significance of differently expressed genes (Schadt et al. 2004; Hubner et al. 2005). We applied relatively stringent criteria to obtain the list of genes that differ in expression between ZS30 and Fr. Our criteria excluded weakly expressed genes. Only about 2–3% of total genes remained for further analysis. Moreover, our approach required overlap of 2 microarray data sets, and genes with large variances were excluded. Overall, the minimum set of genes predominantly is expressed at high levels and differed strongly between ZS30 and Fr. However, strong expression differences may be the first to be noted during studies of complex phenotypes and to be followed up by molecular genetic and functional studies and by studies that identify DNA polymorphism associated (either directly or indirectly via linkage) with gene expression differences (Wittkopp et al. 2004, Hubner et al. 2005). In contrast, it is unclear how to effectively pursue potentially large numbers of *trans*-acting mutations of weak effect during subsequent functional studies at this stage.

In *Drosophila*, current evidence points to a prominent role of *cis*-regulatory mutations in the divergence of gene expression between species (Wittkopp et al. 2004). Our results show that *cis*-regulatory mutations are equally relevant at the population level in *Drosophila*, and our observation is in accordance with recent data derived from yeast (Fay et al. 2004). Thus, assuming that phenotypic variation often is caused by gene expression changes, the molecular genetic dissection of complex phenotypic variation within *Drosophila* populations, strains, or between *Drosophila* species should directly benefit from the availability of microarray data.

Supplementary Material

Supplementary Tables 1–3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

Literature Cited

Carroll SB, Grenier JK, Weatherbee SD. 2001. From DNA to diversity: molecular genetics and the evolution of animal design. Malden, MA: Blackwell Science. p 192.

Coyne JA, Orr HA. 2004. Speciation. Sunderland, MA: Sinauer Associates. p 545.

Fang S, Takahashi A, Wu C-I. 2002. A mutation in the promoter of *desaturase*2 is correlated with sexual isolation between *Drosophila* mating races. Genetics 162:781–4.

Fay JC, McCullough HL, Sniegowski PD, Eisen MB. 2004. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. Genome Biol 5:R26.

Greenberg AJ, Moran JR, Coyne JA, Wu C-I. 2003. Ecological adaptation during incipient speciation revealed by precise gene replacement. Science 302:1754–7.

Greenberg AJ, Moran JR, Fang S, Wu C-I. 2005. Adaptive loss of an old duplicated gene during incipient speciation. Mol Biol Evol 23:401–10.

Hollocher H, Ting CT, Wu ML, Wu C-I. 1997. Incipient speciation by sexual isolation in *Drosophila melanogaster*: extensive genetic divergence without reinforcement. Genetics 147: 1191–201.

Hubner N, Wallace CA, Zimdahl H, et al. (11 co-authors). 2005. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. Nat Genet 37:243–53.

Jin W, Riley RM, Wolfinger RD, White KP, Passador Gurgel G, Gibson G. 2001. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. Nat Genet 29:389–95.

Kohn MH, Fang S, Wu C-I. 2004. Inference of positive and negative selection on the 5′ regulatory regions of *Drosophila* genes. Mol Biol Evol 21:374–83.

Li C, Hung Wong W. 2001. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol 2:RESEARCH0032.

Liu G, Loraine AE, Shigeta R, Cline M, Cheng J, Valmeekam V, Sun S, Kulp D, Siani Rose MA. 2003. NetAffx: affymetrix probesets and annotations. Nucleic Acids Res 31:82–6.

McGraw LA, Gibson G, Clark AG, Wolfner MF. 2004. Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster*. Curr Biol 14:1509–14.

Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM. 2004. Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. Mol Biol Evol 21:1308–17.

Oleksiak MF, Churchill GA, Crawford DL. 2002. Variation in gene expression within and among natural populations. Nat Genet 32:261–6.

Primig M, Williams RM, Winzeler EA, Tevzadze GG, Conway AR, Hwang SY, Davis RW, Esposito RE. 2000. The core meiotic transcriptome in budding yeasts. Nat Genet 26: 415–23.

Schadt EE, Monks SA, Drake TA, et al. (14 co-authors). 2003. Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297–302.

Takahashi A, Tsaur SC, Coyne JA, Wu C-I. 2001. The nucleotide changes governing cuticular hydrocarbon variation and their

evolution in *Drosophila melanogaster*. Proc Natl Acad Sci USA 98:3920–5.

Ting CT, Takahashi A, Wu C-I. 2001. Incipient speciation by sexual isolation in *Drosophila*: concurrent evolution at multiple loci. Proc Natl Acad Sci USA 98:6709–13.

Van Baren MJ, Heutink P. 2004. The PCR suite. Bioinformatics 20:591–3.

Wittkopp PJ. 2005. Genomic sources of regulatory variation in *cis* and in *trans*. Cell Mol Life Sci 62:1779–83.

Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in *cis* and *trans* gene regulation. Nature 430:85–8.

Wu C-I, Hollocher H, Begun DJ, Aquadro CF, Xu Y, Wu ML. 1995. Sexual isolation in *Drosophila melanogaster*: a possible case of incipient speciation. Proc Natl Acad Sci USA 92: 2519–23.

Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L. 2003. *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. Nat Genet 35:57–64.