Gene Regulatory Variation in *Drosophila melanogaster* Renal Tissue

Amanda Glaser-Schmitt, Aleksandra Zečić,¹ and John Parsch²

Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany

ABSTRACT Genetic variation influencing levels of gene expression is abundant in natural populations, and may exert its effects through complex mechanisms that depend on an organism's genetic background and the tissue in which expression is measured. We investigated natural variation in gene expression in the Malpighian tubules of three inbred *Drosophila melanogaster* strains and their F₁ hybrids. One of the strains was from a population in the species' ancestral range (Zambia), while the other two were from a more recently derived population (Sweden). Although closely related, the two Swedish strains differed greatly in terms of their expression inheritance when hybridized with the Zambian strain, with one Swedish strain showing a large excess of genes with recessive expression inheritance, as well as a large number of genes with overdominant inheritance. Although most expression variation could be attributed to *trans*-regulation, there were ~200 genes that showed allele-specific expression differences in each of the between-population hybrids, indicating that *cis*-regulation contributes as well. The *cis*-regulated genes were enriched with cytochrome P450 genes, and the upstream regions of six of these genes were incorporated into transgenic reporter gene constructs to test their effects on expression. Differential expression was observed for five of the six reporter genes in the Malpighian tubule, suggesting that a large proportion of *cis*-regulatory variation lies directly upstream of the affected gene. In most cases, the differential expression was specific to the Malpighian tubule or greater in this tissue than in the rest of the body, highlighting the importance of single-tissue studies of gene expression variation.

KEYWORDS gene expression; *cis*-regulation; *trans*-regulation; cytochrome P450

GENE expression is a complex phenotype that is influenced by multiple genetic and environmental factors. In terms of genetics, differences in DNA sequence can lead to expression variation between species, populations, individuals, or even between two alleles in a diploid individual (Jin *et al.* 2001; Meiklejohn *et al.* 2003; Ranz *et al.* 2003; Wittkopp *et al.* 2004; Hutter *et al.* 2008; Müller *et al.* 2011; Coolon *et al.* 2014). In model systems for which inbred strains are available, such as *Drosophila* species, the comparison of gene expression between two parental strains/species and between the two alleles in F₁ hybrids can be used to infer the

genetic basis of expression variation (Wittkopp *et al.* 2004). In general, there are two types of genetic variation that influence gene expression (Wray 2007; Cheung and Spielman 2009). The first is *cis*-regulatory variation, which affects the expression of a gene linked to the genetic variant. The second is *trans*-regulatory variation, which influences the expression of a gene that is not linked to the genetic variant. *Cis*-regulatory variants often occur within regulatory sequences, such as enhancers or promoters, while *trans*-regulatory variants can occur within a diverse range of functional sequences. Because of their linkage to the gene they regulate, *cis*-regulatory variants are expected to affect the expression of only one of the two alleles in an F₁ hybrid, leading to allele-specific expression (ASE), while *trans*-regulatory variants affect the expression of both alleles equally and do not cause ASE (Wittkopp *et al.* 2004). *Cis*-regulatory variants can be further validated using synthetic gene constructs that directly test the effect of a linked regulatory sequence on gene expression (Wittkopp *et al.* 2002; Gompel *et al.* 2005; Saminadin-Peter *et al.* 2012; Catalán *et al.* 2016; Glaser-Schmitt and Parsch 2018).

Copyright © 2018 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.118.301073>

Manuscript received April 25, 2017; accepted for publication July 3, 2018; published Early Online July 5, 2018.

Supplemental material available at Figshare: <https://doi.org/10.25386/genetics.6741245>.

¹Present address: Department of Biology, Ghent University, 9000 Ghent, Belgium.

²Corresponding author: Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilians-Universität München, LMU BioCenter, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany. E-mail: parsch@bio.lmu.de

Although the genetic basis of expression variation in *Drosophila melanogaster* has been studied previously, most studies have used long-term laboratory strains or single populations (McManus *et al.* 2010; Coolon *et al.* 2014; Chen *et al.* 2015; Osada *et al.* 2017). Furthermore, these studies have measured expression in whole flies or in body segments (e.g., heads). Thus, little is known about expression variation in specific tissues. Moreover, the previous studies have not characterized the sequences responsible for the observed *cis*-regulatory variation experimentally, which can provide insight into the location, scope, and effect size of *cis*-regulatory variation in natural populations. To investigate the genetics of natural variation in gene expression variation in a single tissue, we performed high-throughput RNA-sequencing (RNA-seq) of the Malpighian tubule transcriptomes of three inbred *D. melanogaster* strains, as well as their F₁ hybrids. Two of the strains were collected in Umeå, Sweden, which lies at the northern edge of the species' distribution in Europe (Wilches 2014), while the third strain was collected from the species' inferred ancestral range in Siavonga, Zambia (Pool *et al.* 2012). The Malpighian tubules are excretion organs that are analogous to mammalian kidneys, and are responsible for osmoregulation and the detoxification of xenobiotic compounds (Dow and Davies 2001, 2006; Wang *et al.* 2004; Yang *et al.* 2007; Dow 2009; Chahine and O'Donnell 2011). They are also involved in the innate immune response to bacterial pathogens (Tzou *et al.* 2002; McGettigan *et al.* 2005; Zheng *et al.* 2018). Because flies in Sweden must cope with differences in temperature, humidity, food, xenobiotics, and pathogens relative to flies in the ancestral African species range, physiological changes in the Malpighian tubules are likely to have accompanied the species' range expansion. Furthermore, the Malpighian tubules are a major tissue of expression of several *D. melanogaster* genes that have been implicated in adaptive regulatory evolution, including *Cyp6g1* (Daborn *et al.* 2002) and *fezzik* (Saminadin-Peter *et al.* 2012; Glaser-Schmitt and Parsch 2018), which suggests that functional divergence within this tissue is relevant for environmental adaptation.

Our data allow us to determine both the mode of expression inheritance (e.g., dominant or recessive) and the genetic basis of expression variation (e.g., *cis* or *trans*). We find that the two Swedish strains differ greatly from each other in terms of their expression inheritance in hybrids, with one strain showing a large excess of genes with recessive inheritance, as well as an excess of genes with either over- or underdominant inheritance in hybrids with the Zambian strain. Most of the expression divergence between parental strains can be attributed to *trans*-regulatory variation, although we detect ~200 genes with ASE in each of the interpopulation hybrids. For six cytochrome P450 genes showing ASE, we tested the influence of their upstream regulatory regions on the expression of transgenic reporter genes. In all but one case, we observed an expression difference in the expected direction in the Malpighian tubule. The difference in expression was typically weaker or absent in other tissues, indicating that some of the regulatory variants have both allele- and tissue-specific effects.

Materials and Methods

Fly stocks and maintenance

Two isofemale lines collected in Umeå, Sweden (*SU26N* and *SU58N*; here designated as S26 and S58) were kindly provided by Ricardo Wilches and Wolfgang Stephan (LMU-Munich, Germany). These flies were originally collected in August 2012. One isofemale line derived from Siavonga, Zambia (*ZI418N*; here designated as Z418) as part of the *Drosophila* Population Genomics Project (Pool *et al.* 2012) was kindly provided by John Pool. This line was originally collected in July 2010. The Z418 line has the standard arrangement for all known chromosomal inversion polymorphisms (Corbett-Detig *et al.* 2012; Lack *et al.* 2016). On the basis of diagnostic single-nucleotide polymorphisms (SNPs) (Kapun *et al.* 2014), both Swedish strains were inferred to have the standard chromosomal arrangement for all inversion polymorphisms, with the exception of *In(2L)t*, which was present in S26. To homogenize their genetic backgrounds, all of the above lines were passed through five generations of single brother–sister mating, then subsequently maintained for over 10 generations by *en masse* brother–sister mating. All possible F₁ hybrids of the above parental lines (S26×S58, S26×Z418, and S58×Z418) were generated by crossing two to three virgin females of one line with four to five males of the other line. This was carried out in four to eight replicate vials, half of which used the reciprocal cross (i.e., the genotype of the mother and father were swapped). All flies were reared on standard cornmeal-molasses medium and were maintained at 21° with a light:dark cycle of 14:10 hr.

RNA extraction and sequencing

Two biological replicates were performed for each genotype. For each of these replicates, the Malpighian tubules of 60, 6-day-old female flies were dissected in 1×PBS (phosphate buffered saline), transferred to RNA-later stabilizing solution (Ambion-ThermoFisher Scientific, Waltham, MA), and frozen at –80° until the time of RNA extraction. For the F₁ hybrid genotypes, each replicate consisted of 30 tubules from each of the reciprocal crosses. Total RNA was extracted from each replicate of 60 pooled Malpighian tubules using the MasterPure RNA Purification Kit (Epicentre, Madison, WI) following the manufacturer's protocol, but with the omission of the DNase I digestion step. mRNA isolation, fragmentation, reverse transcription, library construction, and high-throughput sequencing were carried out by GATC Biotech (Konstanz, Germany) using the Illumina HiSeq 2500 platform (Illumina, San Diego, CA) to generate paired reads of length 125 bp.

RNA-seq data analysis

To create a reference genome for each parental strain, we used genome sequence assemblies of S26 and S58 (Wilches 2014), and Z418 (Lack *et al.* 2015, 2016). For each of the parental strains, the new reference genome was created by comparing

the genome sequence of each strain with the *D. melanogaster* reference genome (release 6; Hoskins *et al.* 2015). If there was a nucleotide sequence difference between the parental and reference strains, the parental base was included in the new reference sequence. If there was an uncalled base (“N”) in the parental sequence, which could indicate low coverage, poor quality, or residual heterozygosity, the base in the reference sequence was used. Note, however, that the parental genomes were sequenced from haploid embryos (Langley *et al.* 2011) and heterozygous sites are not expected. The FlyBase annotation version 6.09 (Gramates *et al.* 2017) was then used to extract all transcribed regions (including mRNA, rRNAs, and noncoding RNAs) from each reference genome. For RNA-seq read mapping, the corresponding parental reference genome was used for each parental library. For the F₁ hybrids, mapping was performed to the combined parental reference genomes. This approach was used to prevent mapping bias for genes with greater sequence similarity to the reference genome in one parent than in the other.

RNA-seq reads were mapped to the reference transcriptome sequences using NextGenMap (Sedlazeck *et al.* 2013) in paired-end mode. Read pairs matching more than one transcript of a gene were randomly assigned to one of the transcripts of that gene. For downstream analyses, we used the sum of read counts across all transcripts (*i.e.*, all annotated exons) of a gene and carried out our analyses on a “per gene” basis. Across all libraries, 97% of the reads mapped to annotated transcribed sequences (protein-coding genes, ribosomal RNAs, or other noncoding RNAs). Of the remaining reads, only 1% mapped to intergenic regions. This suggests that our libraries contained, at most, only a very small amount of genomic DNA contamination, although it is also possible that some of these intergenic regions are transcribed to some extent in the Malpighian tubules of these strains.

To standardize statistical power across replicates, we held the total number of mapped reads constant. For this, the maximum number of mapped reads per replicate was set to that of the replicate with the fewest mapped reads (30,834,175). For all other replicates, reads were randomly subsampled (without replacement) until the total number of mapped reads equaled 30,834,175. Differential expression (DE) analysis was carried out with the negative binomial test implemented in DESeq2 (Love *et al.* 2014), using only genes with a minimum of 20 mapped reads in each replicate. This resulted in a total of 7415 genes that could be compared across all genotypes.

Inferring the mode of expression inheritance

To infer the mode of expression inheritance in F₁ hybrids, we compared their expression to that of their parents and classified genes into six categories: “similar,” P1 dominant, P2 dominant, “additive,” “overdominant,” and “underdominant” (Coolon *et al.* 2014). For this, the expression level of each gene in each genotype was calculated as the mean of its two biological replicates. If a gene showed a < 1.25-fold difference in expression between genotypes, those two genotypes were considered to have the same expression level.

Otherwise, the gene was considered to be expressed differently between the two genotypes. A fold-change cutoff was used instead of a statistical test to avoid a bias in detecting differential expression between alleles/genes with higher expression, as the power of statistical tests increases with increasing read counts. Although the use of a 1.25-fold cutoff is arbitrary, we chose it because it has been used in several previous studies with the justification that most of the significant expression differences detected between samples tend to be of this magnitude (Gibson *et al.* 2004; McManus *et al.* 2010; Coolon *et al.* 2014). For our own data, we calculated the GEL₅₀ (Townsend 2004), which is the fold-change difference in gene expression between two samples necessary to have a 50% chance of detecting it as significant at a false discovery rate (FDR) of 5%, to be in the range 1.14–1.37. Thus, a cutoff of 1.25-fold appears to be reasonable for our analysis. Nevertheless, to test the sensitivity of our results to this cutoff, we repeated the analysis using more stringent cutoffs of 1.5-fold and twofold. For these cutoffs, the chances of detecting a difference as significant were 90.4 and 99.6%, respectively. Additionally, we repeated the analysis with a negative binomial test (Love *et al.* 2014) and a FDR of 5%.

ASE analysis

To detect expression differences between the two alleles in hybrids, we first used the reference parental genome sequences to compile lists of SNPs that could distinguish the transcripts of each pairwise combination of parental alleles. This was done by comparing the two parental reference genome sequences over all transcribed regions annotated in the *D. melanogaster* reference genome (release 6.09). As a quality control step to exclude sites with possible sequencing errors or residual heterozygosity in our pooled sample of 60 tubules (120 chromosomes), we required that all SNPs inferred from the genome sequences be confirmed in the parental RNA-seq data. For this, a SNP was required to have coverage of ≥ 20 in each parent and show the expected variant in $\geq 95\%$ of the mapped reads of each parent. In addition, we called a new SNP from the parental RNA-seq data if a site did not differ (or contained an N) in the parental genome sequences, but had ≥ 20 mapped reads in each parent with $\geq 95\%$ having the same base in each parent that differed from the base in the other parent. Of the sites we could test (exonic sites with coverage ≥ 20), 0.99% were excluded from each of the Swedish lines and 1.24% from the Zambian line because the most common base was present in < 95% of the reads. This suggests that there is slightly higher residual polymorphism in the Zambian line. In the end, the total number of high-confidence SNPs was 16,427 for S26×S58, 59,693 for S26×Z418, and 61,064 for S58×Z418, covering the transcripts of 3245, 6411, and 6428 genes, respectively.

To produce ASE counts for the hybrid genotypes, the same mapping data described above in the *RNA-seq data analysis* section were used, but only the mapped reads that covered at least one diagnostic SNP and, thus, could be assigned to a

parental allele were counted. As described above, counts were summed over all transcripts of a gene and the ASE analysis was carried out on a per gene basis. An initial analysis of ASE was performed using all allele counts for a negative binomial test as implemented in DESeq2 (Love *et al.* 2014). However, to standardize statistical power for the S26×Z418 and S58×Z418 comparisons, we repeated the analysis with a fixed number of diagnostic reads over all replicates. For this, the maximum number of diagnostic reads per replicate was set to that of the replicate with the fewest diagnostic reads (6,125,577). For all other replicates, reads were randomly subsampled (without replacement) until the total number of diagnostic reads equaled 6,125,577. Since there were far fewer SNPs between the two Swedish strains, we considered the S26×S58 hybrid separately, but used the same approach to limit the number of diagnostic reads in each replicate to 1,929,623. Differences in allelic expression were tested with the negative binomial test implemented in DESeq2 (Love *et al.* 2014), using only genes with a minimum of 20 diagnostic reads in each replicate. This resulted in a total of 4558 genes that could be analyzed in the S26×Z418 and S58×Z418 hybrids, and 2183 genes that could be analyzed in the S26×S58 hybrid.

Inferring the genetic basis of expression variation

The genetic basis of expression variation for each gene was determined from the outcome of three statistical tests: a negative binomial test for DE between the two parental strains, a negative binomial test for ASE in the F₁ hybrid, and a Cochran–Mantel–Haenszel (CMH) test of the ratio of expression between the two parents and the ratio between the two alleles in the hybrid. The first two tests were carried out using DESeq2 as described above. The CMH test was carried out using the “mantelhaen.test” function in R (R Core Team 2017). In all cases, the *P*-value was adjusted for multiple testing (Benjamini and Hochberg 1995) and an FDR cutoff of 5% was used to define significant differences. To keep the statistical power comparable between the parents and hybrids, a subsampling procedure was performed as described above. Only parental reads covering a diagnostic site in the hybrid were considered and the total number of reads per replicate was held to 6,125,577 for the S26×Z418 and S58×Z418 comparisons and 1,929,623 for the S26×S58 comparison.

Genes were classified into regulatory classes using the following scheme (Coolon *et al.* 2014). Genes that did not show a significant difference in any of the three tests were considered “conserved.” Genes showing significant ASE in hybrids and significant DE between parents, but no significant difference in the ratios of allelic-to-parental expression by the CMH test, were classified as “all *cis*.” Genes that showed significant DE between the parents and a significant difference in the ratios of allelic-to-parental expression, but no ASE, were classified as “all *trans*.” Genes that did not show significant DE between the parents, but showed significant ASE in hybrids and a significant difference in the ratios of

allelic-to-parental expression, were classified as “compensatory.” Genes that gave a significant result for all three tests were classified as “*cis* + *trans*” if the expression difference between the parents was greater than that between the two alleles and “*cis* × *trans*” if the expression difference between the parents was less than that between the two alleles. Genes that gave a significant result for only one of the three tests were classified as “ambiguous.”

To investigate the statistical power of the above classification scheme, we simulated data sets with expression differences between genotypes or alleles ranging from 1.25- to 1.5-fold and with mean count numbers ranging from 125 to 500 per gene/allele. To match the experimental data, two replicates of each parental and hybrid sample were simulated. For each fold-change and count number combination, 1000 genes were simulated by random binomial sampling in R (R Core Team 2017) and the above statistical procedure was used for classification. We then determined the false negative rate as the proportion of simulated genes that were not detected as statistically significant for the expression category under which they were simulated.

Reporter gene construction and expression assays

The upstream regions of six cytochrome P450 genes showing ASE were tested to determine if at least part of the observed ASE could be explained by upstream *cis*-regulatory divergence. The tested regions varied in size, depending upon the distance to the nearest upstream gene. Tested regions ended as close to the start codon of the target gene as primer design would allow, and for shorter intergenic regions (< 2 kb) began within ~100 bp of the nearest upstream gene. For larger intergenic regions, only the first 2 kb upstream of the gene were considered. The upstream regions of *Cyp12a4*, *Cyp6t1*, *Cyp6a20*, *Cyp12b2*, *Cyp28a5*, and *Cyp12a5*, spanning ~450–1800 bp before their respective start codons (Supplemental Material, Table S1), were PCR-amplified from the S58 and Z418 lines and cloned into the *pCR2.1-TOPO* vector (Invitrogen, Carlsbad, CA). The upstream regions were confirmed via Sanger sequencing using their respective PCR primers, as well as additional sequencing primers for the longer regions (Table S2). A 3.5-kb *NotI* fragment of the *pCMV-SPORT-β-gal* vector (Invitrogen) containing the *Escherichia coli lacZ* gene was inserted into the *NotI* site downstream of the cloned upstream sequences in the *TOPO* vector. A *BamHI/XhoI* fragment containing both the upstream region and *lacZ* coding sequence was then excised and ligated into the *pattB* integration vector (Bischof *et al.* 2007). The *pattB* vectors containing the upstream regions and *lacZ* reporter genes were microinjected into early-stage embryos of the φX-86Fb line (*attP* site at cytological band 86F), which contains a stable source of φC31 integrase on the X chromosome (Bischof *et al.* 2007). Microinjections were performed by Rainbow Transgenic Flies (Camarillo, CA). After microinjection, surviving males were crossed to a *white*[−] strain to remove the integrase and stable, homozygous lines were established for each of the constructs.

For each reporter gene construct, β-galactosidase activity was measured in 4–6-day-old females. β-galactosidase activity

was measured in groups of 15 whole flies, 10 pairs of dissected Malpighian tubules (dissected in $1 \times$ PBS), or 10 carcasses with the Malpighian tubules removed. Soluble proteins were extracted and a β -galactosidase activity assay was performed as described in Hense *et al.* (2007) with the following modifications: whole flies, Malpighian tubules, or carcasses were frozen and homogenized in liquid nitrogen before the addition of 200 μ l (for whole flies and carcasses) or 120 μ l (for Malpighian tubules) of the 0.1 M Tris-HCl, 1 mM EDTA, and 7 mM 2-mercaptoethanol buffer (pH 7.5). β -galactosidase activity was determined spectrophotometrically by following the change in absorbance at 420 nm. A total of four to eight biological replicates were performed for each construct and sample type (whole flies, Malpighian tubules, or carcasses). The significance of differences in β -galactosidase activity between versions of the reporter genes with the S58 and Z418 upstream regions was assessed using a Student's *t*-test.

Analysis of copy number variation in candidate cytochrome P450 genes

Some cytochrome P450 genes are known to show copy number variation in natural populations (Schmidt *et al.* 2010; Najarro *et al.* 2015). To test the possibility that variation in copy number might contribute to the differential expression observed for our six candidate genes, we downloaded raw genomic sequence reads from the NCBI short read archive for strains S26 (SRR2347340), S58 (SR2347343), and Z418 (SR202100) and mapped them to their respective reference genomes with NextGenMap (Sedlazeck *et al.* 2013). The sequence depth at each candidate gene was compared to the genome average to estimate copy number. All of the candidate P450 genes had a depth within 1.2-fold of the average in all lines, with the exception of *Cyp6a20*, which had a depth 1.45-fold higher than average in line Z418 (Table S3). However, this is not unusual given the large variation in depth across the genome and is still well below the expectation for a gene present in two copies (*i.e.*, twice the genome average).

Population genetic and transcription factor-binding analyses of tested upstream regions

To characterize sequence polymorphism and population differentiation within the tested cytochrome P450 upstream regions, we performed local and genome-wide analyses of sequence variation in a European (the Netherlands) and a sub-Saharan African (Zambia) population for which whole-genome sequences are publicly available. The Zambian population is the population from which Z418 was derived (Pool *et al.* 2012) and genome sequences are available from the *Drosophila* Genome Nexus (Lack *et al.* 2015). The Dutch population was used in several previous studies of expression divergence between African and European flies (Hutter *et al.* 2008; Müller *et al.* 2011; Catalán *et al.* 2012), including a previous study of expression divergence within the Malpighian tubule (Huylmans and Parsch 2014). The Dutch genome sequences are described in Voigt *et al.* (2015) and were downloaded from <http://evol.bio.lmu.de/downloads/>.

For the local analysis, we examined the tested cytochrome P450 upstream regions and ~ 50 kb centered around them in the two populations. The following summary and test statistics were calculated using DnaSP version 6 (Rozas *et al.* 2017): mean pairwise nucleotide diversity (π), Watterson's estimate of nucleotide diversity (θ), number of segregating sites, haplotype number, haplotype diversity, F_{ST} , D_{XY} (average pairwise differences between populations and species), number of fixed differences between populations, number of private polymorphisms within populations, and Tajima's *D* (Tajima 1989). To test for potential selective sweeps unique to one of the populations, we calculated Cross Population Extended Haplotype Homozygosity (XPEHH) (Sabeti *et al.* 2007) between Zambia and the Netherlands for SNPs within the ~ 50 -kb regions for each population and upstream region using the rehh package (Gautier *et al.* 2017) in R (R Core Team 2017). Only SNPs present in both populations were included in the analysis. As in Gautier *et al.* (2017), a two-sided *P*-value threshold of 10^{-4} was used to assess significance. To determine if sequence differences in the tested S58 and Z418 cytochrome P450 upstream regions correspond with predicted differential transcription factor binding, the upstream sequences were scanned using the *matrix-scan* program (Turatsinze *et al.* 2008) in the Regulatory Sequence Analysis Tools suite (Medina-Rivera *et al.* 2015) using a first-order Markov *D. melanogaster*-specific background model and position weight matrices of all transcription factors in the JASPAR CORE Insecta database (Mathelier *et al.* 2016).

For the genome-wide analysis, we determine the empirical distribution of F_{ST} between the Dutch and Zambian populations for each of the five major chromosome arms. We then determined the proportion of SNPs within the tested cytochrome P450 upstream regions that fell within the upper 10 or 5% of the F_{ST} distribution, and tested for enrichment of highly differentiated SNPs with a χ^2 test. The analysis was performed separately for each chromosome arm, because the number of sequences and the coverage differed slightly among arms. The F_{ST} values of the upper 10% cutoffs for chromosome arms 2L, 2R, 3L, 3R, and X were 0.194, 0.218, 0.230, 0.196, and 0.195, respectively. The respective F_{ST} cutoffs for the upper 5% were 0.323, 0.382, 0.375, 0.340, and 0.380.

Data availability

The RNA-seq data have been submitted to the National Center for Biotechnology Information Gene Expression Omnibus under the series accession number GSE103645. Sequences of the tested cytochrome P450 upstream regions were deposited in GenBank under the accession numbers MF948224–MF948235. Supplemental material available at Figshare: <https://doi.org/10.25386/genetics.6741245>.

Results

We performed high-throughput sequencing of cDNA derived from the Malpighian tubules of two isofemale strains from

Sweden (S26 and S58) and one from Zambia (Z418), as well as all three possible F_1 hybrids (S26×S58, S26×Z418, and S58×Z418). For each genotype, two biological replicates were performed, leading to a total of 12 RNA-seq libraries. For the F_1 hybrids, progeny of the two reciprocal crosses were pooled equally within each replicate. Thus, we were not able to detect parent-of-origin effects, although these are expected to be very rare or absent in *D. melanogaster* (Coolon *et al.* 2012; Chen *et al.* 2015). The total number of reads per RNA-seq library ranged from 34.5 to 48.9 million, with the number of reads mapping to annotated protein-coding genes ranging from 30.8 to 43.5 million. The number of expressed genes per replicate, as defined by having ≥ 20 mapped reads per gene, ranged from 8015 to 8413.

Differential expression among genotypes

To compare expression among multiple genotypes with consistent statistical power, we randomly subsampled the RNA-seq reads of each replicate so that the total number of mapped reads was 30,834,175 per replicate, which equaled that of the replicate with the fewest mapped reads. Furthermore, we considered only those genes that had at least 20 mapped reads in each replicate. This resulted in a total of 7415 genes that could be compared among all genotypes (Table S4).

Using the negative binomial test implemented in DESeq2 (Love *et al.* 2014) and an FDR of 5%, we found the number of DE genes between genotypes to range from 255 to 1850, with the greatest number of differences being between the parental strains (Figure 1). In terms of DE genes, the two Swedish strains were not more similar to each other than to the Zambian strain. For example, there were more DE genes between S26 and S58 than between S26 and Z418 (Figure 1). This is surprising, given that the two Swedish strains were derived from the same collection at a single location and, across the genome, show much less DNA sequence divergence with each other (0.54%) than either strain does with the Zambian strain (0.97%). However, when comparing all genes and taking the magnitude of expression differences into account, divergence between the populations becomes evident. When expression divergence is quantified by the Spearman rank correlation over all genes (ρ), the two Swedish strains are more similar to each other than either is to the Zambian strain (Figure 1). Similarly, in a principle component (PC) analysis of all counts over all samples, the first PC separates the Swedish strains from the Zambian strain, while the second PC separates the two Swedish strains from each other (Figure 2A). Furthermore, when considering the overlap of DE genes between any pair of strains and the third strain, there is a greater number of genes that differ between both Swedish strains and the Zambian strain than between any other combination (Figure 2B). There was also a highly significant amount of overlap of the DE genes between the Swedish and Zambian lines and those previously reported to show differential expression in Malpighian tubules between samples from the Netherlands and Zimbabwe (Huylmans and Parsch 2014), which suggests that much of the expression

		DE genes					
		S26	S58	Z418	S26x S58	S26x Z418	S58x Z418
ρ	S26		1767	1703	422	938	1232
	S58	0.282		1850	259	964	585
	Z418	0.355	0.378		1019	337	458
	S26x S58	0.124	0.086	0.295		290	301
	S26x Z418	0.325	0.311	0.249	0.227		255
	S58x Z418	0.263	0.107	0.139	0.116	0.194	

Figure 1 Gene expression divergence among genotypes. The numbers above the diagonal indicate the number of DE genes for each pairwise comparison (FDR = 5%). The numbers below the diagonal indicate the expression divergence as determined by subtracting the Spearman correlation coefficient (ρ) over all genes from one. DE, differential expression; FDR, false discovery rate.

divergence between the Swedish and Zambian strains reflects divergence between European and African populations (Table 1).

The mode of expression inheritance

To determine the mode of expression inheritance, we compared the expression between the parental strains and their F_1 hybrids following the approach of Coolon *et al.* (2014). Genes that did not differ in expression between the two parental strains, nor between the parental strains and the F_1 hybrid, were considered to have similar expression. Genes that differed in expression between the parental strains, but only between one of the parental strains and the F_1 hybrid, were considered to have dominant inheritance, with the parent that did not differ from the F_1 hybrid being the dominant parent. Cases in which the F_1 hybrid differed in expression from both parents were considered additive if the expression level fell between that of the two parents, overdominant if it was greater than that of both parents, or underdominant if it was less than that of both parents. Following previous studies (Gibson *et al.* 2004; McManus *et al.* 2010; Coolon *et al.* 2014; Chen *et al.* 2015), a fold-change cutoff of 1.25 was used to define differential expression. To assess the robustness of our results to the choice of cutoff, we repeated the analysis using cutoffs of 1.5 and 2.0, or requiring a significant expression difference by a negative binomial test with an FDR of 5% (Table 2).

In hybrids involving the Swedish strain S26, we detected a high proportion of genes showing dominant inheritance of expression, especially those for which the non-S26 parent was dominant (Table 2). Regardless of the criterion used to define a difference in expression, the proportion of dominance cases in which the non-S26 parent was dominant was significantly $> 50\%$ for hybrids with both S58 (sign test, $P < 3 \times 10^{-10}$) and Z418 ($P < 10^{-15}$). While S26 expression

Figure 2 The structure of expression variation among genotypes. (A) PC analysis of gene expression counts for all six genotypes. The biological replicates of each genotype are shown in the same color. (B) Numbers of genes that show a significant expression difference (FDR = 5%) between both of the parental strains given on the top line of the x-axis and the strain shown underneath. DE, differential expression; FDR, false discovery rate; PC, principle component.

was mostly recessive to Z418 expression, S58 expression was generally dominant over that of Z418 at all fold-change cutoffs ($P < 0.03$), but was generally recessive when a FDR cutoff of 5% was employed ($P = 3 \times 10^{-7}$; Table 2).

Hybrids of the S26 and Z418 strains showed a large excess of cases of misexpression, in which the hybrid displayed either over- or underdominance, with overdominance being more common at all fold-change cutoffs, but underdominance being more common with a FDR 5% cutoff (Table 2). In all cases, the proportion of misexpressed genes in this hybrid was significantly greater than that in the other two-hybrid combinations (Fisher's exact test, $P < 3 \times 10^{-11}$). Unlike the results for

dominance described above, this misexpression does not appear to be a common feature of the S26 genetic background, as hybrids between S26 and S58 did not show a large number of misexpressed genes (Table 2).

The excess of genes with overdominant inheritance in the S26×Z418 hybrid under all fold-change cutoffs appears to be a result of these genes having high expression variation between biological replicates. With the fold-change approach, the expression level of the two replicates is averaged and the variance between replicates is not considered. In contrast, the statistical approach accounts for this variation and genes with high variation between replicates are unlikely to show a significant expression difference between genotypes. Hybrids involving S26, especially the S26×Z418 hybrid, showed relatively high variance between replicates (Figure S1). When considering all genes, the median coefficient of variation (CV) in read counts for the two replicates of S26×Z418 was 0.091, which was significantly greater than that of the S58×Z418 (0.064) and S26×S58 (0.079) hybrids (Wilcoxon test, $P < 10^{-16}$ in both cases). Among the genes showing misexpression in the S26×Z418 hybrid, the median CV was much higher (0.32). This can explain why there were many more genes with overdominant expression when a fold-change cutoff was used: these genes typically had very high expression in only one of the two replicates, leading to a high fold-change in average expression that would not be detected as significant with a statistical approach. Note, however, that variation between replicates cannot explain the excess of underdominant genes in the S26×Z418 hybrids, which was also detected with the statistical approach.

The genetic basis of expression variation

When considering the full data set of all mapped reads covering a diagnostic SNP in each F₁ hybrid replicate, we detected 93, 187, and 259 genes with significant ASE (FDR < 5%) in S26×S58, S26×Z418, and S58×Z418, respectively (Table S4). These genes show at least some level of *cis*-regulatory divergence between the parental alleles. To infer the genetic basis of expression variation between the parental strains, we followed the scheme of Coolon *et al.* (2014) to classify genes into six regulatory categories: conserved, “*cis*-only,” “*trans*-only,” *cis* + *trans*, *cis* × *trans*, compensatory, and ambiguous (see *Materials and Methods*). Because only RNA-seq reads covering an informative SNP between the parental strains could be used to detect ASE in hybrids, we randomly subsampled reads to balance statistical power across the parental and hybrid comparisons (see *Materials and Methods*). We also limited our analysis to genes with ≥ 20 mapped, informative reads in each replicate. For the two between-population comparisons (Sweden vs. Zambia), this resulted in 4558 genes with 6,125,577 informative reads per replicate. For the comparison of the Swedish strains, which had a much lower SNP density, we were able to analyze 2183 genes with 1,929,623 informative reads per replicate.

The proportions of genes falling into each regulatory category were remarkably similar across comparisons, although

Table 1 Overlapping DE genes

Test 1	Test 2	DE 1	DE 2	Overlap	<i>P</i> (overlap) ^a	Same ^b	<i>P</i> (same)
S26 vs. Z418	S58 vs. Z418	1394	1519	710	2.4×10^{-124}	664	4.6×10^{-95}
S26 vs. Z418	Eur vs. Afr ^c	1394	627	308	1.3×10^{-49}	240	1.1×10^{-18}
S58 vs. Z418	Eur vs. Afr ^c	1519	627	338	3.5×10^{-58}	260	2.8×10^{-20}
S26&S58 vs. Z418	Eur vs. Afr ^c	710	627	213	1.0×10^{-54}	163	3.5×10^{-9}

DE, differential expression; Eur, European; Afr, African.

^a *P*-value from a hypergeometric test. A total of 5930 genes could be compared across all data sets.

^b Number of overlapping genes with differential expression in the same direction.

^c Comparison of pooled European and African populations from Huylmans and Parsch (2014).

the comparison of S26 and Z418 had a smaller proportion of conserved genes and a larger proportion of ambiguous genes (Table 3). When considering only the genes showing non-ambiguous regulatory divergence, the highest proportion fell into the *trans*-only category and these genes greatly outnumbered *cis*-only genes in all comparisons (Table 3). This is not expected to be the result of a difference in the statistical power to detect *cis* or *trans* effects, as total read counts were held constant and both classifications require two significant test results with an FDR of 5% (see *Materials and Methods*). To further explore the statistical power, we simulated data for genes with different read counts and different fold-changes in their expression difference, and determined the false negative rate for detecting all-*cis* effects or all-*trans* effects (Table 4). As expected, the false negative rate varied greatly depending on the fold-change difference in expression and the number of read counts per replicate. However, our power was good in cases where the fold-change was > 1.25 and/or the number of counts per replicate was ≥ 250 . Importantly, there was no systematic bias toward detecting *cis* or *trans* effects (Table 4). Thus, differences in statistical power cannot explain the large excess of *trans* effects observed in our data.

There was a high proportion of genes classified as ambiguous in the S26×Z418 comparison (Table 3). These genes were highly enriched with those that showed Z418-dominant inheritance (Table 2). Of the 1505 genes with ambiguous expression in the S26×Z418 comparison, 500 (33.2%) showed dominant inheritance of the Z418 expression state. In contrast, of the 946 genes with ambiguous expression in the S58×Z418 comparison, only 73 (7.7%) fell into the Z418-dominant category. Similarly, a higher proportion of ambiguous genes showed misexpression (either over- or underdominance) in the S26×Z418 comparison (10.3%) than in the S58×Z418 comparison (1.6%).

Functional classification of genes with ASE

To identify common functions among genes showing ASE in the F₁ hybrids, we looked for an enrichment of gene ontology (GO) biological process terms and InterPro protein domains using FlyMine (Lyne *et al.* 2007). For both the S26×Z418 and S58×Z418 hybrids, the greatest GO enrichment was for “oxidation-reduction process” (FDR < 0.0005 in both cases). Cytochrome P450 genes comprised about half of these genes (13/28 in the S26×Z418 comparison and 14/25 in the S58×Z418 comparison). In agreement with this, the only

protein domain showing significant enrichment was the cytochrome P450 domain (FDR $< 5 \times 10^{-6}$ in both cases). The overlapping genes with significant ASE in both of the above hybrids were also enriched for oxidation-reduction process (FDR < 0.04) and the cytochrome P450 domain (FDR $< 9 \times 10^{-6}$), and included 10 cytochrome P450 genes. The genes showing ASE in S26×S58 hybrids did not show significant enrichment of any GO terms or protein domains.

Influence of inversion *In(2L)t* on expression variation

The S26 strain differed from the other strains used in our analysis in that it had the chromosomal inversion *In(2L)t* (see *Materials and Methods*). To test if this inversion could contribute to the observed expression variation, we compared the differentially expressed genes between S26 and the other strains with those previously found to have their expression influenced by *In(2L)t* (Lavington and Kern 2017) (Table S5). When comparing expression between the parental strains, there was an enrichment of *In(2L)t*-influenced genes for the S26–S58 comparison (χ^2 test, $P = 0.007$), with the overlapping genes accounting for 2.5% (43/1756) of the differentially expressed genes. Similarly, there was an enrichment of *In(2L)t*-influenced genes among those showing ASE in the S26×Z418 hybrid (χ^2 test, $P < 0.001$), which accounted for 5.6% (10/178) of the genes showing ASE. Thus, it appears that *In(2L)t* makes a minor contribution to our observed expression variation. In contrast, there was no enrichment of *In(2L)t*-influenced genes among those showing S26-recessive inheritance in either the S26×S58 or S26×Z418 hybrids (χ^2 test, $P > 0.5$ in both cases), nor among the genes showing misexpression in the S26×Z418 hybrid, for which there was a significant paucity of overlapping genes (χ^2 test, $P = 0.04$). Thus, the presence of *In(2L)t* cannot explain the unusual patterns of dominance or misexpression seen in the S26 hybrids (Table 2).

Experimental test of putative *cis*-regulatory elements

To determine if at least part of the ASE that we detected between the Swedish and Zambian strains was caused by variation in regulatory sequences upstream of the affected genes, we used transgenic reporter gene constructs to test the influence of the upstream regions of six cytochrome P450 genes on gene expression (Figure 3A). We focused on cytochrome P450 genes because they were highly enriched among the genes showing ASE (see above), and the examined genes

Table 2 Genes with each mode of inheritance at different fold-change or FDR cutoffs

Cutoff	Mode	S26×S58	S26×Z418	S58×Z418
1.25-fold	Similar	4926	3062	5326
	P1 dominant ^a	539	639	753
	P2 dominant ^b	1365	2031	663
	Additive	479	414	547
	Overdominant	52	1061	56
1.5-fold	Underdominant	54	208	70
	Similar	6555	5246	6631
	P1 dominant ^a	190	336	320
	P2 dominant ^b	510	1164	265
	Additive	137	131	177
Twofold	Overdominant	8	451	7
	Underdominant	15	87	15
	Similar	7151	6587	7145
	P1 dominant ^a	67	152	131
	P2 dominant ^b	162	426	93
FDR 5%	Additive	31	29	43
	Overdominant	0	182	1
	Underdominant	4	39	2
	Similar	6839	6290	6594
	P1 dominant ^a	154	187	236
	P2 dominant ^b	317	788	363
	Additive	100	91	215
	Overdominant	1	3	3
	Underdominant	4	56	4

FDR, false discovery rate.

^a Expression level of the first parent is dominant.

^b Expression level of the second parent is dominant.

were chosen as a subset of those showing significant ASE between both Swedish strains and Z418. Furthermore, none of the genes showed evidence of being present in multiple copies within the parental genomes, based on their coverage in the raw genomic sequence reads (Table S3). The upstream regions from both the S58 and Z418 alleles were cloned in front of a *lacZ* reporter gene and integrated into the genome using ϕ C31 integration (Bischof *et al.* 2007). Reporter gene expression was determined by assaying β -galactosidase activity in whole females, Malpighian tubules, and carcasses with the Malpighian tubules removed.

The tested upstream regions drove significantly different expression in the expected direction in the Malpighian tubules for all of the constructs except *Cyp12b2* (Figure 3B). For *Cyp12b2*, the S58 upstream region drove significantly higher expression in whole bodies and carcasses, but not Malpighian tubules (Figure 3B). This suggests that it is responsible for a general increase in *Cyp12b2* expression that is not specific to the Malpighian tubule. The variation responsible for the Malpighian tubule-specific expression divergence of *Cyp12b2* may lie outside of the upstream region included in our reporter gene construct. All of the other upstream regions showed greater expression divergence in the Malpighian tubules than in whole flies or carcasses (Figure 3B). This signal of tissue-specific regulation was especially evident for *Cyp12a4*, *Cyp28a5*, *Cyp6a20*, and *Cyp12a5* (Figure 3B). Interestingly, the upstream region of *Cyp12a5* led to opposite patterns of expression divergence in Malpighian tubules than

Table 3 Numbers of genes showing *cis*- and *trans*-regulatory divergence between strains

Classification	S26×S58 ^a	S26×Z418	S58×Z418
Conserved	1390	2440	2949
All <i>cis</i>	34	39	50
All <i>trans</i>	210	446	483
<i>cis</i> + <i>trans</i>	16	65	63
<i>cis</i> × <i>trans</i>	17	24	33
Compensatory	15	39	34
Ambiguous	502	1505	946

^a Due to the low density of SNPs between the Swedish strains, this analysis covers 2183 genes, with 1,929,623 informative reads per replicate. The other comparisons cover 4558 genes, with 6,125,577 informative reads per replicate.

in whole bodies or carcasses (Figure 3B), further underscoring that this *cis*-regulatory change is tissue-specific. Half of the tested upstream regions (*Cyp12a4*, *Cyp28a5*, and *Cyp12a5*) drove levels of tubule-specific expression divergence that were similar to those seen in either the parental strains or the F₁ hybrid (Figure 3, B and C). Thus, variation in these upstream regions can account for almost all of the *cis*-regulatory expression divergence observed between S58 and Z418 for these genes, while for *Cyp6a20* and *Cyp6t1*, upstream variation accounts for 50–70% of the observed *cis* variation (Figure 3C).

Predicted transcription factor binding and *cis*-expression QTLs in the tested upstream regions

To determine if the expression divergence driven by the tested upstream regions (Figure 3B) was associated with differential transcription factor binding, we scanned the tested upstream regions for predicted transcription factor-binding sites (TFBSs) using the *matrix-scan* program (Turatsinze *et al.* 2008) with the position weight matrices of all transcription factors in the JASPAR CORE Insecta database (Mathelier *et al.* 2016). All of the tested upstream regions, with the exception of *Cyp12b2*, had at least one TFBS with predicted differential binding (absence or lower binding score in one sequence) between the Z418 and S58 sequences (Table 5 and Table S6).

We also determined the frequency of all SNPs segregating between the Z418 and S58 tested upstream regions within the Dutch and Zambian populations (Table S7). All but three SNPs in the *Cyp6t1* upstream region and one SNP in the *Cyp12a4* upstream region were polymorphic in at least one population (Table S7). For each of the five upstream regions with predicted differential transcription factor binding (Table 5 and Table S6), at least one SNP overlapped with a predicted TFBS difference. Most of these SNPs did not differ in frequency by > 10–20% between the two populations, although several showed larger frequency differences (Table 5). To determine if any of the SNPs segregating within the tested upstream regions could be associated with previously identified expression variation, we compared them with the female *cis*-expression QTLs (eQTLs) identified in a North American population by Massouras *et al.* (2012). Nine female *cis*-eQTLs were located within our tested regions, all of which were associated

Table 4 False negative rate for detecting ASE/DE, all-*cis*, and all-*trans* effects

Fold-change ^b	Counts ^c	False negative rate (%) ^a		
		ASE/DE	All <i>cis</i>	All <i>trans</i>
1.25	125	100.0	100.0	100.0
1.25	250	35.1	59.2	52.1
1.25	500	2.1	2.9	6.0
1.33	125	62.2	85.4	74.3
1.33	250	7.1	11.6	11.1
1.33	500	0.1	0.1	0.1
1.50	125	10.5	11.6	18.3
1.50	250	0.2	0.2	1.0
1.50	500	0.0	0.0	0.0

ASE, allele-specific expression; DE, differential expression

^a The percentage of 1000 genes for which the simulated type of regulatory divergence was not detected as significant.

^b The true difference in expression between samples or alleles.

^c Mean number of simulated counts per gene or allele.

with *Cyp6a20* expression. However, only one of these *cis*-eQTLs segregated between Z418 and S58 (Table 5). The lack of congruence between studies may stem from the fact that the *cis*-eQTLs were identified from whole-fly expression data, while most of the expression divergence that we detected was specific to Malpighian tubule (Figure 3B).

Polymorphism and divergence in the tested cytochrome P450 gene regions

To characterize sequence polymorphism and divergence within and surrounding the tested upstream cytochrome P450 gene regions, we calculated several summary statistics in the Dutch and Zambian populations. We did not identify any fixed sequence differences between the two populations for either the tested upstream regions or the 50 kb surrounding them (Table S8 and Table S9). The tested upstream regions had more private polymorphisms in the Zambian population (~10–50 per region) than in the Dutch population (0–2 per region; Table S8), which is expected given that the Zambian population is located in the putative ancestral range and has a larger sample size. Generally, sequence diversity, as measured by π and θ , was similar between the tested upstream regions and their surrounding 50-kb regions, although diversity was slightly higher (median = 1.25-fold) in the tested regions (Tables S8 and S9). Divergence from *D. simulans*, as measured by D_{XY} , also tended to be higher in the tested regions (median = 1.67-fold) than in the surrounding 50-kb regions (Tables S8 and S9).

To determine if the Dutch and Zambian cytochrome P450 upstream regions displayed greater population differentiation than the rest of the genome, we tested for an excess of high- F_{ST} outliers in these regions in comparison to all SNPs on their respective chromosome arms. We found an excess of highly differentiated SNPs in the upstream regions of *Cyp6t1*, *Cyp6a20*, *Cyp12a4*, and *Cyp12b2*, which was significant for SNPs falling within both the upper 10% and the upper 5% of the chromosome-wide F_{ST} distribution (Figure 3D).

To test if the SNP differentiation in the cytochrome P450 upstream regions was associated with extended haplotypes

that might be a signal of selective sweeps, we calculated XPEHH (Sabeti *et al.* 2007) between Zambia and the Netherlands for SNPs in the ~50-kb regions surrounding each cytochrome P450 gene of interest. Three of the six genes (*Cyp12a4*, *Cyp12a5*, and *Cyp28a5*) showed peaks of significant XPEHH upstream of the genes of interest (Figure S2); however, these peaks did not correspond to the regions used in our reporter gene experiments. Thus, it is not clear if these extended haplotypes are associated with the observed cytochrome P450 expression differences.

Discussion

Our comparison of gene expression among multiple inbred lines and their F₁ hybrids shed light on several aspects of the genetic contribution to expression variation within species. For example, our results demonstrate that inferences about the mode of expression inheritance are highly dependent on genotype, as even strains from the same population can show extreme and opposite patterns. Although the Swedish strains S26 and S58 are very similar genetically, they differed greatly in terms of their expression inheritance in hybrids, with S26 having more genes with recessive inheritance and more genes that were misexpressed (either over- or underdominant) in hybrids with the Zambian strain Z418 (Table 2). Previous work has shown that the environment (temperature) also has a major impact on the mode of gene expression inheritance in *D. melanogaster* (Chen *et al.* 2015). Given the strong influence of both the genetic background and the environment, one should use caution when drawing general conclusions about the mode of gene expression inheritance.

It is not clear why so many S26 genes show recessive inheritance of expression, but it appears to be related to the overall expression level of the gene. Of the 2031 genes that showed S26-recessive inheritance in S26×Z418 hybrids, 1589 (78%) had lower expression in S26 than in Z418 (sign test, $P < 10^{-16}$). Of the 663 genes that showed S58-recessive inheritance in S58×Z418 hybrids, 399 (60%) had lower expression in S58 than in Z418, which was also a significant excess (sign test, $P < 10^{-16}$), but not as extreme as the S26 case (Fisher exact test, $P < 10^{-16}$). In the S26×S58 hybrid, 962/1365 (70%) of the genes with S26-recessive inheritance had lower expression in S26 than in S58 (sign test, $P < 10^{-16}$). In contrast, only 226/539 (42%) of the genes with S58-recessive inheritance had lower expression in S58 than in S26 (sign test, $P < 0.0002$). Thus, low-expressed alleles are not disproportionately recessive in all hybrid combinations. Instead, this appears to be a feature of the S26 genome. A previous study examining the mode of expression inheritance in chromosomal introgression lines that used four strains from North America and one strain from sub-Saharan Africa found that expression differences between homozygous strains tend to be masked in heterozygotes, with the African strain being particularly efficient at masking recessive differences between two of the North American strains (Lemos *et al.* 2008). Our results are consistent with this, in that the

Figure 3 Genetic differentiation in upstream regions and the effect of upstream variation on expression divergence. (A) Schematic of the reporter gene constructs. The upstream regions of six cytochrome P450 genes from the S58 and Z418 strains were cloned in front of the *lacZ* gene and integrated into the *D. melanogaster* genome. Lines indicate the polymorphic sites between S58 and Z418, with different colors indicating which strain has the ancestral variant (red = Z418, blue = S58, and gray = unknown). The total number of polymorphisms in each category is shown in the same color scheme to the right of the upstream region. (B) Relative expression of each endogenous gene in Malpighian tubules (MTs) of the S58 and Z418 parental strains (blue) and their alleles in the S58×Z418 hybrid (red) as determined by RNA-sequencing. Also shown is the relative expression of the respective reporter genes in MTs (orange), whole flies (white), and carcasses with the MTs removed (gray) as determined by β -galactosidase activity. Error bars represent SEM. For reporter genes, the significance of expression differences between the two upstream regions was assessed using a Student's *t*-test. (C) Expression divergence in the MT between the Z418 and S58 strains partitioned according to type of regulatory variation. The contributions of *cis* variation within the tested upstream region (red), *cis* variation in other regions (gray), and *trans* variation (blue) were estimated from the RNA-sequencing and reporter gene data. For genes showing a nonadditive *cis* × *trans* interaction (Cyp12a5 and Cyp6a20), the total *cis* variance was assumed to be the maximum possible, *i.e.*, the observed expression divergence between the parental strains. (D) The proportion of highly differentiated SNPs within the tested upstream regions relative to that of their respective chromosome arms. Values greater than one indicate an enrichment of SNPs with F_{ST} falling within the top 10% (red) or top 5% (blue) of the chromosomal F_{ST} distribution. Significance was assessed using a χ^2 -test. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.005$.

expression of an African strain (Z418) is overwhelmingly dominant over that of a non-African strain (S26). However, as in the previous study, we find variation among the non-African strains with regard to the mode of expression inheritance. This could be explained if alleles affecting expression typically have recessive, deleterious effects (Lemos *et al.* 2008). If so, such alleles are more likely to be segregating in non-African populations that have passed through a bottleneck and, thus, have a smaller effective population size.

The high proportion of misexpressed genes in the S26×Z418 hybrid, in particular the cases of overdominance, can be attributed to the high expression variance between biological replicates in this background. The high variance can also explain why the number of misexpressed genes is much higher when using a fold-change cutoff than when

using a statistical test with an FDR (Table 2), as the fold-change approach considers the mean expression between replicates, but not the variation between them. Despite this shortcoming, the fold-change approach has the advantage that its power to detect dominance or misexpression does not depend on the expression level of the gene (Coolon *et al.* 2014). With the statistical approach, the power to detect differential expression between the hybrids and the parents increases with the gene's expression level (*i.e.*, the number of read counts in RNA-seq data). A gene in the hybrid with expression that is intermediate to the two parents is more likely to differ significantly from the parent with higher expression, resulting in an overall bias toward detecting dominant inheritance of the parent with lower expression. Thus, it can be informative to apply both fold-change and statistical

Table 5 Location and frequency of SNPs with predicted differential transcription factor binding between Z418 and S58 upstream regions

Gene	Pos ^c	Variant ^a		Frequency Ancestral Variant ^b		TFBS affected ^d
		Z418	S58	ZI	NL	
<i>Cyp6a20</i>	251	G	T	0.510	0.000	<i>B-1H</i>
<i>Cyp6a20</i>	558	A	G	0.138	0.091	<i>Ets21C</i>
<i>Cyp6a20</i>	699*	G	C	0.918	0.800	<i>schlank</i>
<i>Cyp6t1</i>	319	C	T	0.893	0.727	<i>schlank</i>
<i>Cyp6t1</i>	373	T	C	0.893	0.727	<i>Cf2, zen</i>
<i>Cyp6t1</i>	710	A	T	0.036	0.455	<i>cad</i>
<i>Cyp6t1</i>	1588	C	T	0.990	0.091	<i>cad</i>
<i>Cyp12a4</i>	472	G	A	0.069	0.636	<i>Vsx2</i>
<i>Cyp12a5</i>	414	A	G	0.653	0.091	<i>Cf2</i>
<i>Cyp12a5</i>	774	A	T	0.626	0.091	<i>BEAF-32, br(var.3), dl, pnr</i>
<i>Cyp12a5</i>	775	C	T	0.839	0.091	<i>BEAF-32, br(var.3), dl, pnr</i>
<i>Cyp12a5</i>	777	A	C	0.795	0.091	<i>BEAF-32, br(var.3), dl, pnr</i>
<i>Cyp12a5</i>	781	C	A	0.190	0.909	<i>BEAF-32, br(var.3), pnr</i>
<i>Cyp28a5</i>	147	A	A	0.883	0.700	<i>br, slp1, twi</i>
<i>Cyp28a5</i>	482	C	T	0.822	0.800	<i>z</i>

Pos, position; ZI, Zambian population; NL, Dutch population.

^a SNP variant in tested upstream regions. Ancestral variant is shown in bold and was inferred from an alignment with *D. simulans*, *D. sechelia*, and *D. yakuba*.

^b Frequency of ancestral variant in Zambian and Dutch populations. When the ancestral variant could not be inferred, the frequency of Zambian variant is shown.

^c Position in alignment of Zambian and Dutch populations of tested upstream region. Asterisk indicates that site corresponds to female *cis*-eQTL identified by Massouras *et al.* (2012).

^d Matrix models of predicted TFBS in the JASPAR CORE Insecta database (Mathelier *et al.* 2016) affected by SNP.

cutoffs when assessing the mode of expression inheritance. The high variance among replicates may also explain why the S26×Z418 hybrid showed a large number of genes for which the genetic basis of expression divergence was classified as ambiguous (Table 3), as the high variance makes it less likely to detect significant differences with the CMH test used to compare the ratios of parental-to-allele expression. At present it is not possible to determine if the increased expression variance among replicates of S26 hybrids is a technical artifact, or if it represents an increase in expression noise in the hybrid genetic background.

In general, we found that most expression divergence between strains could be attributed to *trans*-regulatory variation (Table 3). This is consistent with previous studies that examined whole-fly gene expression in the F1 hybrids of inbred *D. melanogaster* lines and found that *trans*-regulatory changes were more prevalent than *cis*-regulatory changes (Coolon *et al.* 2014; Chen *et al.* 2015). Note, however, that the numbers of genes showing *cis*- or *trans*-regulatory divergence may be influenced greatly by environmental factors, such as temperature (Chen *et al.* 2015). In contrast to these studies, which analyzed gene regulatory divergence in only a small number of inbred lines, studies employing larger numbers of inbred lines (18–39) from a North American population reported a greater number of *cis*-regulatory variants (Massouras *et al.* 2012) and an excess of *cis*-regulatory variants relative to *trans*-regulatory variants (Osada *et al.* 2017). There could be several reasons for this discrepancy among studies. First, the differences in samples size and in statistical analysis methods likely result in differences in statistical power to detect significant regulatory divergence. Second, Massouras *et al.* (2012) found that most *cis*-regulatory variants had sex-specific effects, with over twice as many being

observed in males as in females. Because the current study and the previous studies of ASE used only females (to have the X chromosome of both parents present in the hybrid), they would not be able to detect such sex-specific *cis*-regulatory effects. Third, with the approach of Osada *et al.* (2017), a regulatory effect would only be detected as significant if it showed a consistent pattern in multiple hybrid backgrounds. It is likely that *cis*-regulatory variants more often have additive effects than *trans*-regulatory variants because, by definition, they should affect the expression of only one of the two alleles in hybrids. *Trans*-regulatory variants, on the other hand, may show more epistatic interactions with other variants in the genomic background and, thus, not have consistent additive effects across hybrid genotypes. This could, at least in part, explain why a smaller proportion of genes with *trans*-regulatory effects were detected by Osada *et al.* (2017). Finally, all of the above studies measured expression in whole flies, heads, or headless bodies, while the current study measured expression specifically in Malpighian tubules. It is thus possible that the relative proportions of *cis*- and *trans*-regulatory variants differ among tissues, or when comparing individual tissues to whole flies or body segments.

Genes showing ASE in F₁ hybrids were enriched for cytochrome P450s and, using transgenic reporter gene constructs, we were able to verify the contribution of *cis*-regulatory changes to the observed expression variation in the Malpighian tubules for five of these genes (Figure 3B). Most of the tested genes displayed a combination of *cis* and *trans* effects between S58 and Z418, with two, *Cyp6a20* and *Cyp12a5*, displaying *cis* × *trans* interactions (Table S1). Our reporter gene analysis focused specifically on the genes' immediate upstream regions. For three of the genes (*Cyp12a4*, *Cyp12a5*, and *Cyp28a5*), the upstream region could account

for almost all of the *cis*-regulatory expression divergence observed between S58 and Z418 in the Malpighian tubule, while for two others (*Cyp6t1* and *Cyp6a20*), the upstream region could account for > 50% of the expression variation (Figure 3, B and C).

Two of the tested genes, *Cyp12a4* and *Cyp12a5*, are located in tandem and in the same orientation within a ~5-kb region on chromosome 3R. Despite this proximity, the expression change of these genes is in the opposite direction (Figure 3B). Thus, the two genes do not appear to be coregulated, but instead regulated independently by their individual upstream regions. For *Cyp6t1* and *Cyp28a5*, the differences in the observed fold-change between the native genes and the reporter genes might be caused by more distant *cis*-regulatory variants that were not included in our reporter gene constructs. However, there may also be some degree of ascertainment bias, as we chose candidate genes that showed a large difference in allelic expression for further analysis by reporter gene experiments. Thus, it would not be possible to detect cases in which the reporter gene showed a large fold-change difference, but the RNA-seq data indicated only a small fold-change difference. Taken together, our findings suggest that (i) most *cis*-regulatory variants are located immediately upstream (within 450–1800 bp) of their target genes and (ii) on average, *cis*-regulatory variants located immediately upstream of a target gene have the largest effects on expression.

The majority of the tested upstream regions drove Malpighian tubule-specific expression divergence, with lower or, in one case, opposite expression changes in whole flies and dissected carcasses (Figure 3B). None of the tested genes show tissue-specific expression (Malpighian tubule or otherwise) and all are expressed across a broad range of tissues (Chintapalli *et al.* 2007). They also span a large range of expression levels, from low to high expression in both the Malpighian tubule and other tissues (Chintapalli *et al.* 2007). Thus, the Malpighian tubule-specific regulation that we detected does not appear to be related to a gene's native expression level or tissue specificity, which highlights that *cis*-regulatory changes are able to alter expression in a single tissue. *Cis*-regulatory changes are thought to play a particularly important role during adaptation, in part, precisely because they are able to drive such tissue-specific expression changes (Wray 2007; Carroll 2008).

An enrichment of cytochrome P450 genes was also reported among differentially expressed genes in a study comparing Malpighian tubule expression between pooled isofemale lines from the Netherlands and Zimbabwe (Huylmans and Parsch 2014), and three of the cytochrome P450 genes included in our functional analysis (Figure 3) showed concordant differential expression between the Dutch and the Zimbabwean populations (Figure S3). These results suggest that regulatory evolution of cytochrome P450s may have accompanied the out-of-Africa expansion of the species. Indeed, cytochrome P450 genes represent excellent candidates for adaptation to the new environmental conditions *D. melanogaster*

encountered as it expanded its species range, as (i) a large subset of these genes, especially those expressed in the Malpighian tubules, are predicted to play a role in the detoxification of xenobiotic compounds (Chung *et al.* 2009; The UniProt Consortium 2017), (ii) *D. melanogaster* P450 genes predicted to be involved in detoxification are rapidly evolving (Chung *et al.* 2009), and (iii) the Malpighian tubules are known to play an important role in the detoxification and metabolism of xenobiotic compounds as well as immune defense (Tzou *et al.* 2002; McGettigan *et al.* 2005; Dow and Davies 2006; Yang *et al.* 2007; Zheng *et al.* 2018). The population genetic analysis of the cytochrome P450 upstream regions used in our reporter gene experiments suggests that there have not been classical, hard selective sweeps of *cis*-regulatory variants in derived European populations. Although we have not determined the specific SNP(s) responsible for the observed expression divergence between populations, we find that none of the SNPs in the tested cytochrome P450 upstream regions show a fixed difference between the African and European populations. However, there is an enrichment of highly differentiated SNPs within the cytochrome P450 upstream regions (Figure 3D), which suggests that their population frequencies are shaped by local adaptation. Most of these SNPs segregate within the ancestral population (Table S7), suggesting that selection has acted mainly on standing genetic variation and not new mutations that arose within the derived population.

Acknowledgments

We thank Ricardo Wilches, Stephan Hutter, Wolfgang Stephan, and John Pool for sharing fly strains and genome sequences; Andrew Kern for sharing *In(2L)t* expression data; and Hilde Lainer, Ann Kathrin Huylmans, and Dorothy John Robbert for assistance in the laboratory. The manuscript was improved thanks to constructive feedback from the associate editor and two anonymous reviewers. This work was supported by Deutsche Forschungsgemeinschaft grant PA 903/8-1.

Literature Cited

- Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57: 289–300.
- Bischof, J., R. K. Maeda, M. Hediger, F. Karch, and K. Basler, 2007 An optimized transgenesis system for *Drosophila* using germ-line-specific ϕ C31 integrases. *Proc. Natl. Acad. Sci. USA* 104: 3312–3317. <https://doi.org/10.1073/pnas.0611511104>
- Carroll, S. B., 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134: 25–36. <https://doi.org/10.1016/j.cell.2008.06.030>
- Catalán, A., S. Hutter, and J. Parsch, 2012 Population and sex differences in *Drosophila melanogaster* brain gene expression. *BMC Genomics* 13: 654. <https://doi.org/10.1186/1471-2164-13-654>
- Catalán, A., A. Glaser-Schmitt, E. Argyridou, P. Duchon, and J. Parsch, 2016 An indel polymorphism in the *MtnA* 3' untranslated region is associated with gene expression variation and

- local adaptation in *Drosophila melanogaster*. PLoS Genet. 12: e1005987. <https://doi.org/10.1371/journal.pgen.1005987>
- Chahine, S., and M. J. O'Donnell, 2011 Interactions between detoxification mechanisms and excretion in Malpighian tubules of *Drosophila melanogaster*. J. Exp. Biol. 214: 462–468. <https://doi.org/10.1242/jeb.048884>
- Chen, J., V. Nolte, and C. Schlötterer, 2015 Temperature stress mediates decanalization and dominance of gene expression in *Drosophila melanogaster*. PLoS Genet. 11: e1004883 (erratum: PLoS Genet. 11: e1005315); (erratum: PLoS Genet. 12: e1006079). <https://doi.org/10.1371/journal.pgen.1004883>
- Cheung, V. G., and R. S. Spielman, 2009 Genetics of human gene expression: mapping DNA variants that influence gene expression. Nat. Rev. Genet. 10: 595–604. <https://doi.org/10.1038/nrg2630>
- Chintapalli, V. R., J. Wang, and J. A. Dow, 2007 Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat. Genet. 39: 715–720. <https://doi.org/10.1038/ng2049>
- Chung, H., T. Sztal, S. Pasricha, M. Sridhar, P. Batterham *et al.*, 2009 Characterization of *Drosophila melanogaster* cytochrome P450 genes. Proc. Natl. Acad. Sci. USA 106: 5731–5736. <https://doi.org/10.1073/pnas.0812141106>
- Coolon, J. D., K. R. Stevenson, C. J. McManus, B. R. Graveley, and P. J. Wittkopp, 2012 Genomic imprinting absent in *Drosophila melanogaster* adult females. Cell Rep. 2: 69–75. <https://doi.org/10.1016/j.celrep.2012.06.013>
- Coolon, J. D., C. J. McManus, K. R. Stevenson, B. R. Graveley, and P. J. Wittkopp, 2014 Tempo and mode of regulatory evolution in *Drosophila*. Genome Res. 24: 797–808. <https://doi.org/10.1101/gr.163014.113>
- Corbett-Detig, R. B., C. Cardeno, and C. H. Langley, 2012 Sequence-based detection and breakpoint assembly of polymorphic inversions. Genetics 192: 131–137. <https://doi.org/10.1534/genetics.112.141622>
- Daborn, P. J., J. L. Yen, M. R. Bogwitz, G. Le Goff, E. Feil *et al.*, 2002 A single P450 allele associated with insecticide resistance in *Drosophila*. Science 297: 2253–2256. <https://doi.org/10.1126/science.1074170>
- Dow, J. A., 2009 Insights into the Malpighian tubule from functional genomics. J. Exp. Biol. 212: 435–445. <https://doi.org/10.1242/jeb.024224>
- Dow, J. A., and S. A. Davies, 2006 The Malpighian tubule: rapid insights from post-genomic biology. J. Insect Physiol. 52: 365–378. <https://doi.org/10.1016/j.jinsphys.2005.10.007>
- Dow, J. A. T., and S. A. Davies, 2001 The *Drosophila melanogaster* Malpighian tubule. Adv. Insect Phys. 28: 1–83.
- Gautier, M., A. Klassmann, and R. Vitalis, 2017 rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. Mol. Ecol. Resour. 17: 78–90. <https://doi.org/10.1111/1755-0998.12634>
- 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. <https://doi.org/10.1534/genetics.104.026583>
- Glaser-Schmitt, A., and J. Parsch, 2018 Functional characterization of adaptive variation within a *cis*-regulatory element influencing *Drosophila melanogaster* growth. PLoS Biol. 16: e2004538. <https://doi.org/10.1371/journal.pbio.2004538>
- Gompel, N., B. Prud'homme, P. J. Wittkopp, V. A. Kassner, and S. B. Carroll, 2005 Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. Nature 433: 481–487. <https://doi.org/10.1038/nature03235>
- Gramates, L. S., S. J. Marygold, G. D. Santos, J. M. Urbano, G. Antonazzo *et al.*, 2017 FlyBase at 25: looking to the future. Nucleic Acids Res. 45: D663–D671. <https://doi.org/10.1093/nar/gkw1016>
- Hense, W., J. F. Baines, and J. Parsch, 2007 X chromosome inactivation during *Drosophila* spermatogenesis. PLoS Biol. 5: e273. <https://doi.org/10.1371/journal.pbio.0050273>
- Hoskins, R. A., J. W. Carlson, K. H. Wan, S. Park, I. Mendez *et al.*, 2015 The Release 6 reference sequence of the *Drosophila melanogaster* genome. Genome Res. 25: 445–458. <https://doi.org/10.1101/gr.185579.114>
- Hutter, S., S. S. Saminadin-Peter, W. Stephan, and J. Parsch, 2008 Gene expression variation in African and European populations of *Drosophila melanogaster*. Genome Biol. 9: R12. <https://doi.org/10.1186/gb-2008-9-1-r12>
- Huylmans, A. K., and J. Parsch, 2014 Population- and sex-biased gene expression in the excretion organs of *Drosophila melanogaster*. G3 (Bethesda) 4: 2307–2315. <https://doi.org/10.1534/g3.114.013417>
- Jin, W., R. M. Riley, R. D. Wolfinger, K. P. White, G. Passador-Gurgel *et al.*, 2001 The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. Nat. Genet. 29: 389–395. <https://doi.org/10.1038/ng766>
- Kapun, M., H. van Schalkwyk, B. Mc, T. Allister, T. Flatt *et al.*, 2014 Inference of chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations of *Drosophila melanogaster*. Mol. Ecol. 23: 1813–1827. <https://doi.org/10.1111/mec.12594>
- Lack, J. B., C. M. Cardeno, M. W. Crepeau, W. Taylor, R. B. Corbett-Detig *et al.*, 2015 The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. Genetics 199: 1229–1241. <https://doi.org/10.1534/genetics.115.174664>
- Lack, J. B., J. D. Lange, A. D. Tang, R. B. Corbett-Detig, and J. E. Pool, 2016 A thousand fly genomes: an expanded *Drosophila* genome nexus. Mol. Biol. Evol. 33: 3308–3313. <https://doi.org/10.1093/molbev/msw195>
- Langley, C. H., M. Crepeau, C. Cardeno, R. Corbett-Detig, and K. Stevens, 2011 Circumventing heterozygosity: sequencing the amplified genome of a single haploid *Drosophila melanogaster* embryo. Genetics 188: 239–246. <https://doi.org/10.1534/genetics.111.127530>
- Lavington, E., and A. D. Kern, 2017 The effect of common inversion polymorphisms *In(2L)t* and *In(3R)Mo* on patterns of transcriptional variation in *Drosophila melanogaster*. G3 (Bethesda) 7: 3659–3668. <https://doi.org/10.1534/g3.117.1133>
- Lemos, B., L. O. Araripe, P. Fontanillas, and D. L. Hartl, 2008 Dominance and the evolutionary accumulation of *cis*- and *trans*-effects on gene expression. Proc. Natl. Acad. Sci. USA 105: 14471–14476. <https://doi.org/10.1073/pnas.0805160105>
- Love, M. I., W. Huber, and S. Anders, 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Lyne, R., R. Smith, K. Rutherford, M. Wakeling, A. Varley *et al.*, 2007 FlyMine: an integrated database for *Drosophila* and *Anopheles* genomics. Genome Biol. 8: R129. <https://doi.org/10.1186/gb-2007-8-7-r129>
- Massouras, A., S. M. Waszak, M. Albarca-Aguilera, K. Hens, W. Holcombe *et al.*, 2012 Genomic variation and its impact on gene expression in *Drosophila melanogaster*. PLoS Genet. 8: e1003055. <https://doi.org/10.1371/journal.pgen.1003055>
- Mathelier, A., O. Fornes, D. J. Arenillas, C. Y. Chen, G. Denay *et al.*, 2016 JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 44: D110–D115. <https://doi.org/10.1093/nar/gkv1176>
- McGettigan, J., R. K. McLennan, K. E. Broderick, L. Kean, A. K. Allan *et al.*, 2005 Insect renal tubules constitute a cell-autonomous immune system that protects the organism against bacterial infection. Insect Biochem. Mol. Biol. 35: 741–754. <https://doi.org/10.1016/j.ibmb.2005.02.017>
- McManus, C. J., J. D. Coolon, M. O. Duff, J. Eipper-Mains, B. R. Graveley *et al.*, 2010 Regulatory divergence in *Drosophila* revealed by mRNA-seq. Genome Res. 20: 816–825. <https://doi.org/10.1101/gr.102491.109>

- Medina-Rivera, A., M. Defrance, O. Sand, C. Herrmann, J. A. Castro-Mondragon *et al.*, 2015 RSAT 2015: Regulatory Sequence Analysis Tools. *Nucleic Acids Res.* 43: W50–W56. <https://doi.org/10.1093/nar/gkv362>
- Meiklejohn, C. D., J. Parsch, J. M. Ranz, and D. L. Hartl, 2003 Rapid evolution of male-biased gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 100: 9894–9899. <https://doi.org/10.1073/pnas.1630690100>
- Müller, L., S. Hutter, R. Stamboliyska, S. S. Saminadin-Peter, W. Stephan *et al.*, 2011 Population transcriptomics of *Drosophila melanogaster* females. *BMC Genomics* 12: 81. <https://doi.org/10.1186/1471-2164-12-81>
- Najarro, M. A., J. L. Hackett, B. R. Smith, C. A. Highfill, E. G. King *et al.*, 2015 Identifying loci contributing to natural variation in xenobiotic resistance in *Drosophila*. *PLoS Genet.* 11: e1005663 (erratum: *PLoS Genet.* 12: e1005830). <https://doi.org/10.1371/journal.pgen.1005663>
- Osada, N., R. Miyagi, and A. Takahashi, 2017 *Cis*- and *trans*-regulatory effects on gene expression in a natural population of *Drosophila melanogaster*. *Genetics* 206: 2139–2148. <https://doi.org/10.1534/genetics.117.201459>
- Pool, J. E., R. B. Corbett-Detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno *et al.*, 2012 Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS Genet.* 8: e1003080. <https://doi.org/10.1371/journal.pgen.1003080>
- R Core Team, 2017 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Ranz, J. M., C. I. Castillo-Davis, C. D. Meiklejohn, and D. L. Hartl, 2003 Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* 300: 1742–1745. <https://doi.org/10.1126/science.1085881>
- Rozas, J., A. Ferrer-Mata, J. C. Sánchez-DelBarrio, S. Guirao-Rico, P. Librado *et al.*, 2017 DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34: 3299–3302. <https://doi.org/10.1093/molbev/msx248>
- Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter *et al.*, 2007 Genome-wide detection and characterization of positive selection in human populations. *Nature* 449: 913–918. <https://doi.org/10.1038/nature06250>
- Saminadin-Peter, S. S., C. Kemkemer, P. Pavlidis, and J. Parsch, 2012 Selective sweep of a *cis*-regulatory sequence in a non-African population of *Drosophila melanogaster*. *Mol. Biol. Evol.* 29: 1167–1174. <https://doi.org/10.1093/molbev/msr284>
- Schmidt, J. M., R. T. Good, B. Appleton, J. Sherrard, G. C. Raymant *et al.*, 2010 Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus. *PLoS Genet.* 6: e1000998. <https://doi.org/10.1371/journal.pgen.1000998>
- Sedlazeck, F. J., P. Rescheneder, and A. von Haeseler, 2013 NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* 29: 2790–2791. <https://doi.org/10.1093/bioinformatics/btt468>
- Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- The UniProt Consortium, 2017 UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45: D158–D169. <https://doi.org/10.1093/nar/gkw1099>
- Townsend, J. P., 2004 Resolution of large and small differences in gene expression using models for the Bayesian analysis of gene expression levels and spotted DNA microarrays. *BMC Bioinformatics* 5: 54. <https://doi.org/10.1186/1471-2105-5-54>
- Turatsinze, J. V., M. Thomas-Chollier, M. Defrance, and J. van Helden, 2008 Using RSAT to scan genome sequences for transcription factor binding sites and *cis*-regulatory modules. *Nat. Protoc.* 3: 1578–1588. <https://doi.org/10.1038/nprot.2008.97>
- Tzou, P., E. De Gregorio, and B. Lemaitre, 2002 How *Drosophila* combats microbial infection: a model to study innate immunity and host-pathogen interactions. *Curr. Opin. Microbiol.* 5: 102–110. [https://doi.org/10.1016/S1369-5274\(02\)00294-1](https://doi.org/10.1016/S1369-5274(02)00294-1)
- Voigt, S., S. Laurent, M. Litovchenko, and W. Stephan, 2015 Positive selection at the *Polyhomeotic* locus led to decreased thermosensitivity of gene expression in temperate *Drosophila melanogaster*. *Genetics* 200: 591–599. <https://doi.org/10.1534/genetics.115.177030>
- Wang, J., L. Kean, J. Yang, A. K. Allan, S. A. Davies *et al.*, 2004 Function informed transcriptome analysis of *Drosophila* renal tubule. *Genome Biol.* 5: R69. <https://doi.org/10.1186/gb-2004-5-9-r69>
- Wilches, R., 2014 Evolution of genes related to temperature adaptation in *Drosophila melanogaster* as revealed by QTL and population genetics analyses. Dissertation. Ludwig-Maximilians-Universität, Munich, Germany.
- Wittkopp, P. J., K. Vaccaro, and S. B. Carroll, 2002 Evolution of yellow gene regulation and pigmentation in *Drosophila*. *Curr. Biol.* 12: 1547–1556. [https://doi.org/10.1016/S0960-9822\(02\)01113-2](https://doi.org/10.1016/S0960-9822(02)01113-2)
- Wittkopp, P. J., B. K. Haerum, and A. G. Clark, 2004 Evolutionary changes in *cis* and *trans* gene regulation. *Nature* 430: 85–88. <https://doi.org/10.1038/nature02698>
- Wray, G. A., 2007 The evolutionary significance of *cis*-regulatory mutations. *Nat. Rev. Genet.* 8: 206–216. <https://doi.org/10.1038/nrg2063>
- Yang, J., C. McCart, D. J. Woods, S. Terhzaz, K. G. Greenwood *et al.*, 2007 A *Drosophila* systems approach to xenobiotic metabolism. *Physiol. Genomics* 30: 223–231. <https://doi.org/10.1152/physiolgenomics.00018.2007>
- Zheng, W., F. Rus, A. Hernandez, P. Kang, W. Goldman *et al.*, 2018 Dehydration triggers ecdysone-mediated recognition-protein priming and elevated anti-bacterial immune responses in *Drosophila* Malpighian tubule renal cells. *BMC Biol.* 16: 60. <https://doi.org/10.1186/s12915-018-0532-5>

Communicating editor: D. Begun