Genetica (2011) 139:465–477
DOI 10.1007/s10709-011-9566-0

Population transcriptomics: insights from *Drosophila simulans*, *Drosophila sechellia* and their hybrids

François Wurmser · David Ogereau · Tristan Mary-Huard · Béatrice Loriod · Dominique Joly · Catherine Montchamp-Moreau

Received: 22 November 2010 / Accepted: 7 March 2011 / Published online: 19 March 2011

Abstract Sequence differentiation has been widely studied between populations and species, whereas interest in expression divergence is relatively recent. Using microarrays, we compared four geographically distinct populations of *Drosophila simulans* and a population of *Drosophila sechellia*, and interspecific hybrids. We observed few differences between populations, suggesting a slight population structure in *D. simulans*. This structure was observed in direct population comparisons, as well as in interspecific comparisons (hybrids vs. parents, *D. sechellia* vs. *D. simulans*). Expression variance is higher in the French and Zimbabwean populations than in the populations from the ancestral range of *D. simulans* (Kenya and Seychelles). This suggests a large scale phenomenon of decanalization following the invasion of a new environment. Comparing *D. simulans* and *D. sechellia*, we revealed 304 consistently differentially expressed genes, with striking overrepresentation of genes of the cytochrome P450 family, which could be related to their role in detoxification as well as in hormone regulation. We also revealed differences in genes involved in Juvenile hormone and Dopamine differentiation. We finally observed very few differentially expressed genes between hybrids and parental populations, with an overrepresentation of X-linked genes.

Keywords Expression differentiation · Population structure · Ecology · CYP450 · Hormone regulation · *Drosophila*

Introduction

Genetic variation at the sequence level has been widely studied within and between species, to identify the forces that drive evolution (e.g., natural selection, genetic drift). The interest has been now turning to changes in gene regulation leading to genetic isolation, and ultimately speciation. Indeed, gene expression can have a strong influence on downstream phenotypes, and therefore its variation is likely to be a target of natural selection (Pavey et al. 2010). Large-scale technologies such as microarrays provide genome wide information that can be used to assess expression evolution (Gilad and Borevitz 2006). Several comparisons are possible: among populations of the same species, among species, as well as between interspecific hybrids and their parents. Natural populations have been previously studied for regulatory variations in several organisms from yeast (Townsend et al. 2003) to hominids (Storey et al. 2007). Two studies have reported differences in gene expression between African and European populations of *Drosophila melanogaster* with respect to sex biased genes (Meiklejohn et al. 2003) and genes involved in toxicity resistance or flight musculature which are potentially involved in adaptation (Hutter et al.

Electronic supplementary material The online version of this article (doi:10.1007/s10709-011-9566-0) contains supplementary material, which is available to authorized users.

F. Wurmser (✉) · D. Ogereau · D. Joly · C. Montchamp-Moreau
Laboratoire Evolution, Génomes et Spéciation, CNRS UPR9034 Avenue de la Terrasse, Gif-sur-Yvette F-91198 Cedex, and Univ Paris-Sud, 91405 Orsay, France
e-mail: francois.wurmser@legs.cnrs-gif.fr

T. Mary-Huard
Statistics and Genomes Team, INRA UMR518/AgroParisTech, 16 rue Claude Bernard, 75231 Paris, France

B. Loriod
U928 INSERM, Parc scientifique de Luminy, 163 Avenue de Luminy, case 928, 13288 Marseille Cedex 09, France

2008). No evolutionary studies so far have compared expression between closely related species and their hybrids using a set of distinct parental populations (contrasting with studies based on a single strain for each population or species). This is the approach we have adopted in the present work, studying closely related *Drosophila* species belonging to the *melanogaster* subgroup. It will allow us to assess at the expression level, not only the extent of population structure but also the importance of misregulation and additivity in hybrids.

We used here different populations of *Drosophila simulans* and a population of the sibling species *Drosophila sechellia*. *Drosophila simulans* and *D. sechellia* have long been thought to have separated about 400.000 years ago (Hey and Kliman 1993; Kliman et al. 2000; Lachaise and Silvain 2004), but more recent data pointed to a split occurring around 250.000 years ago (McDermott and Kliman 2008). On average, the DNA sequence divergence between these two sister species is around 1.5% (McDermott and Kliman 2008).

While these two species are phylogenetically very close, they are ecologically strongly different. *D. simulans,* which originates from eastern Africa or Madagascar (Lachaise et al. 1988; Lachaise and Silvain 2004; Dean and Ballard 2004; Kopp et al. 2006), is now a cosmopolitan generalist. *D. sechellia* is an endemic specialist in the Seychelles islands, and breeds exclusively on *Morinda citrifolia*, a plant highly toxic for other drosophilids (including *D. simulans*) (R'Kha et al. 1991). According to Oleksiak et al. (2002), gene expression differences between populations arise mainly from genetic drift (they did not show more differences within population than between). Our design already takes into account the variation within populations, thus differences shown there may equally represent drift or adaptation. Notably we expect a higher expression variation in *D. simulans* because of the higher diversity of environments occupied by the species compared to *D. sechellia*. The geographic structure of *D. simulans* has been studied using nuclear gene sequences or microsatellites (Hamblin and Veuille 1999; Schöfl and Schlötterer 2006; Baudry et al. 2006). These studies have revealed a differentiation between east African populations, other African populations (notably Zimbabwean), and European populations. In contrast, *D. sechellia* harbours little nucleotide sequence variation, which ranks this species as the least genetically diverse drosophilid (Legrand et al. 2009).

Previous studies which have examined divergence in gene expression between *D. simulans* and closely related species, have used samples isolated from a single isofemale line (i.e. consisting of the offspring of a single wild-caught female) (Michalak and Noor 2003; Haerty and Singh 2006; Moehring et al. 2007). Thus, they have not taken into account the intraspecific variation. An exception from this is the recent parallel studies of sequence polymorphism and expression of six *D. simulans* lines originating from five different locations within the species range (Holloway et al. 2007; Lawniczak et al. 2008). The authors revealed interesting insights into the evolution of regulatory sequences between populations. However, each of their population was represented by only one (or two in the case of Madagascar) line, thus they could not take into account intra-population variation.

The present study includes four populations of *D. simulans* each represented by four isofemale lines, four isofemale lines of *D. sechellia* and four hybrid "populations" (crosses between *D. sechellia* males and *D. simulans* females). This design allowed us to take into account two different types of variances: intraspecific variance (between the different populations of *D. simulans)*, and intra-population variance (between the different lines which represent biological replicates within each population). We considered the four isofemale lines of *D. sechellia as a single* population as this species does not show geographic structure (Legrand et al. 2009). We used males because they are the most affected sex in *Drosophila* hybrids (since they are heterogametic they show more hybrid breakdown), therefore suitable to highlight differences linked with species divergence. Regulation breakdown in hybrids can occur due to incompatibilities between alleles at a given locus It can also result from negative epistasis between loci according to Dobzhansky-Muller's model of hybrid incompatibility (Dobzhansky 1936; Muller 1942). Differences in hybrid expression can also simply result from regulation divergence between genomes. Hybrids have been previously shown to harbour strong regulation breakdown compared to parents in different organisms (Gibson et al. 2004; Haerty and Singh 2006; Moehring et al. 2007). However, other studies have shown large scale additivity in the patterns of expression of the hybrids (Hughes et al. 2006; Rottscheidt and Harr 2007). These contrasting results may be caused by methodological differences (organ-specific vs. whole body/ differences in microarray platform) or by different degree of divergences and inbreeding between species (Rottscheidt and Harr 2007). These two non-mutually exclusive hypotheses will be considered here. We extracted RNA from whole body males to consider only general differences, and not tissue specific differences.

Our populations of *D. simulans* were from Zimbabwe, Kenya, the Seychelles (the last two likely represent the ancestral range) (Lachaise et al. 1988; Lachaise and Silvain 2004; Dean and Ballard 2004; Kopp et al. 2006), and from France (a derived population). This approach allowed us to determine to what extent gene expression shows geographic structure in *D. simulans*, as well as differentiate between population effects (possibly linked with recent invasion) and species divergence. We observed a population structure in *D. simulans*, and consistent expression divergence

between *D. simulans* and *D. sechellia*. We notably observed a strong involvement of the Cytochrome P450 gene family, as well as genes regulating the juvenile hormone (JH).

Materials and methods

Drosophila stocks

We studied four populations of *D. simulans*. Each population was represented by four isofemale lines (biological replicates). The populations came from France (the Rhone Valley, collected in 2003), Kenya (Nairobi, collected in 2001), Zimbabwe (Mazoe, collected in 1997) and the Seychelles Islands (Mahé and Praslin, collected in 2003). The four *D. sechellia* lines also originated from Mahé and Praslin in the Seychelles archipelago (collected in 2003).

RNA samples

The lines were mass reared in uncrowded culture, on axenic standard medium at 25°C, with a natural light cycle. For each isofemale line, ≥6 replicate cultures were raised in vials containing each 8 males and 8 virgin females (10 of each for *D. sechellia*). For a given *D. simulans* population, four different crosses were performed, each between females of one of the line and males from a different line of *D. sechellia* (at least 3 replicates per cross). We thus obtained four "populations" of hybrids corresponding to the four *D. simulans* populations (Table 1). The experimental design thus included a total of nine "populations", each consisting of four biological replicates. Virgin male offspring from at least 3 replicate vials were collected within a few hours of emergence to create pools of 25 individuals that were transferred into fresh vials. Seven days later, each pool was frozen at −80°C. We used the Nucleospin RNA II kit from Macherey–Nagel to extract the RNA from the pools of 25 whole-body adult males, yielding to a total of about 3 μg RNA to hybridize on the arrays. RNA was then reverse transcribed in presence of alpha d'CTP p33.

Arrays

The arrays were nylon filters spotted with long amplicons from the species *D. melanogaster*. They were hybridised in the TAGC platform in Marseille. There were 7,041 spots: 5,931 different whole cDNA of *D. melanogaster* and 1,100 control spots, either negative, or positive controls (a cDNA of *Arabidopsis thaliana*: chlorophylle synthetase). cDNAs were cloned into a vector, amplified and spotted on the array. Each spotted fragment contained both the cDNA and a specific part of the vector for spotting normalization (first hybridisation). These arrays were hybridised twice. Firstly, hybridisation was performed with a P33 labeled oligonucleotide probe specific to the vector sequence spotted. As every molecule spotted contained this sequence, the radioactive signal (vector hybridisation signal) read was proportional to the quantity of spotted cDNA. Secondly, after deshybridisation of the vector probe, we proceeded with the hybridisation of the cDNA samples. A second radioactive signal was read. It will be further designated as complex hybridisation signal.

Data normalization

The normalization procedure was defined by the manufacturer of the arrays. Every spot for which the vector signal was smaller than 5 times the negative spots' median was eliminated from the analysis. For both signals, background (measured by the median of negative spots) was subtracted. Inter-spots/intra-arrays normalization was then performed by dividing the complex hybridisation signal by the vector hybridisation signal. The last normalization step was to divide the obtained expression value in each array by the corresponding median of all spots (or by the median of positive controls), effectively normalizing the signal between arrays. The two approaches (median of controls/median of all spots) led to the same results. Normalization quality was assessed visually by MA plot and box-plot of normalized expression values for each array (Supplementary Fig. 1). All genes with four or more missing data throughout the 36 arrays were discarded.

Statistical analysis

The statistical analysis was based on the whole set of 36 arrays. To determine differentially expressed genes, an Analysis of Variance (ANOVA) model was fitted for each gene (Kerr et al. 2000, 2002). The fitted model was the following:

$$Y_{ij} = \mu_i + E_{ij},$$

where i is the population index (i = 1,…,9: 4 *D. simulans*, 1 *D. sechellia* and 4 hybrid "populations"), j is the biological replicate index (j = 1,…,4), Y_{ij} is the normalized signal (log-transformed), μ_i is the mean expression for the gene in population i and E_{ij} is the residual variability. This model assumes a common variance for all populations, that is consistently estimated with $36 - 9 = 27$ degrees of freedom for most genes (a few genes have missing values due to normalization). The homogeneity of variance between groups was verified using Levene's test (Levene 1960). We showed a variance homogeneity for all genes but 20 (FDR = 0.1, see Supplementary Fig. 2 for the distribution of *P*-value). We tested the equality of mean expression between *D. sechellia* and *D. simulans*

Table 1 Experimental design to obtain hybrid males

♀ *Drosophila simulans*	Isofemale lines	♂ *Drosophila sechellia* Sech 1	Sech 2	Sech 3	Sech 4
France	F1				
	F2				
	F3				
	F4				
Zimbabwe	Z1				
	Z2				
	Z3				
	Z4				
Kenya	K1				
	K2				
	K3				
	K4				
Seychelles	S1				
	S2				
	S3				
	S4				

16 hybrids were obtained from crosses of *D. sechellia* males with *D. simulans* females. Each isofemale line of *D. sechellia* was involved in a cross with a different isofemale line of each population of *D. simulans*

populations as well as between *D. simulans* populations. As for the comparisons between a population of hybrids and its two parental populations, we performed tests to compare the mean expression of the hybrid to the mean expression of each parental population, and we also compared the mean expression of the hybrid to the average of the mean expressions of the two parental populations. All these comparisons were performed using usual contrast *t* tests within the ANOVA model. For each comparison, raw *P*-values were adjusted by the Benjamini-Hochberg method, which controls the false discovery rate (FDR) (Benjamini and Hochberg 1995). We used a FDR of 0.1.

A mixed-effects model does not suit our analysis, since estimates of some parental effects would only have been from two values. However, we assessed the effect of our analysis assuming correlated data by simulating data with parental effect. We did not find any increased number of false positive even with a high biological/technical variability ratio, the only consequence was a decreased power (Supplementary Fig. 3).

Patterns of inheritance: additivity, dominance, overdominance

To assess patterns of inheritance, we analyzed the distribution of dominance effects using the ratio d/a, where a is half the difference in expression between the parental populations (*D. sechellia* and respectively each population of *D. simulans*), and d is the expression difference between F1 hybrid and the parental average. If $d/a = 0$, it means perfect additivity ($d = 0$), if $|d/a|=1$, complete dominance and if $|d/a| > 1$, overdominance (Falconer and Mackay 1996). We performed this analysis for our four parental populations, only to those genes which were differentially expressed between the parents, to avoid any bias due to equality of expression between the parents.

Variance comparison

To compare the genomic variability in two populations A and B, we propose the following test. This test looks for an excess of genes with higher (or lower) variance in one population relative to another. For a given gene g, we note $\sigma^2_{g,A}$ and $\sigma^2_{g,B}$ the gene expression variances in populations A and B. If population A harbours more genetic variability than population B, then, for most of the genes, the ratio $R_g = \frac{\sigma^2_{g,A}}{\sigma^2_{g,B}}$ will be higher than 1, whereas R_g should be higher than 1 for roughly 50% of the genes if the two populations are comparable. Therefore a test can be based on the number N_1^{AB} of genes for which the empirical ratio R_g of gene *g* is higher than 1. We note p_{AB} the true proportion of genes for which $\sigma^2_{g,A} > \sigma^2_{g,B}$. We test $H_0 = \{p_{AB} = 1/2\}$ (A and B are comparable) versus $H_1 = \{p_{AB} > 1/2\}$ (variability is higher in A). N_1^{AB} has a binomial distribution $B(G, p_{AB})$ with G the number of genes. The P-value of the test is thus:

$$P(N_1^{AB} > n_{1,obs}^{AB} | p_{AB} = 1/2).$$

This analysis takes into account the number of lines available to estimate each variance. We simulated data with different number of lines (n = 2, 5, 10 and 50) with equal population variance, to measure the impact of a small number of lines on variance estimates. This did not affect the *P*-value distribution and thus the error. Significant tests were around 5% (Supplementary Fig. 4). Further simulation showed this only affects the power of the test.

Gene ontology

Lists of differentially expressed genes were examined for statistical over/under representation of Gene Ontology

(Ashburner et al. 2000) terms using FuncAssociate (Berriz et al. 2003) with a reference background consisting of all genes in the arrays. Our array itself was compared to the whole genome, revealing several ontology biases in the construction of the array. This made essential the use of our array as background when examining differentially expressed genes for gene ontology bias. To further explore the terms and the corresponding genes, we used the Gene Ontology database provided by the Gene Ontology consortium (in May 2008).

Results

The experimental design allowed us to perform multiple comparisons, within *D. simulans*, and between *D. simulans* and *D. sechellia*, as well as their hybrids. By maximizing the biological source of variation (using biological and not merely technical replicates, Altman 2005) in populations (and species when applied), we revealed strongly significant variations. After all gene filtering during the normalization process, we assessed expression for 4,398 genes, which is about a fourth of the *Drosophila* genome. The differences observed are thus non exhaustive, but their consistency between all populations, through our large sampling, is supported by the power of the cross-design.

Comparison between populations of *D. simulans*

We did not detect any significant difference in gene expression between the three African populations. Contrasting with this result, all the comparisons between the French population and each of the three African populations showed differential expression. Respectively 6, 7 and 13 genes were found to be differentially expressed between the French population and the Kenyan, the Seychelles and the Zimbabwean populations (Supplementary Table 1). Six genes were differentially expressed in two pairwise comparisons. No gene was differentially expressed in all three pairwise comparisons. Out of the twelve genes which were over-expressed in the French population compared to at least one of the African population, four are cytochrome P450 genes (significantly over-represented, Fisher's exact test, $P < 0.05$).

We compared variability using a binomial test based on the fact that variance ratios of genes are expected to be half of the time above 1 under the assumption of similar variances, using only pairwise comparisons. *P*-values of the binomial test are provided in Table 2. The variance from the French population is significantly higher than the variance from any other population but the Zimbabwean ($P < 0.005$, Bonferroni corrected threshold). In terms of variance, we observe a differentiation of the French and the Zimbabwean populations compared to other African populations (from the zone of origin of *D. simulans*). All other pairwise comparisons revealed significant differences, even though there is a wide range of *P*-values. It is important to note that this analysis is independent from the test of variance homogeneity gene by gene performed with Levene's test. It is possible to have homogeneity gene by gene, whereas on a global scale, heterogeneity can be observed.

Drosophila simulans vs. *Drosophila sechellia*

The comparisons between *D. sechellia* and *D. simulans* yielded to 347, 337, 353 and 518 genes differentially expressed with the populations of Zimbabwe, Kenya, Seychelles and France, respectively. Details of over/under-expressed genes are shown in Table 3. The striking result is that 304 genes are consistently differentially expressed between all four populations of *D. simulans* and *D. sechellia* (Supplementary Table 2). We can therefore assume that these genes present constitutive expression differences between the two species.

Table 2 Above the diagonal: *P*-values of binomial tests under the assumption of equality of variance between the populations

			D. simulans			***D. sechellia***
		France	Zimbabwe	Seychelles	Kenya	Seychelles
D. simulans	F		4.78e-2	2.94e-69 *	3.75e-200 *	9.16e-128 *
	Z	F ≈ Z		1.20e-53 *	1.81e-148 *	1.03e-81 *
	SimS	SimS < F	SimS < Z		7.51e-41 *	1.83e-4 *
	K	K < F	K < Z	K < SimS		1.20e-25 *
D. sechellia	Sech	Sech < F	Sech < Z	Sech < SimS	Sech > K	

Below the diagonal: direction of the variance change

F France, *Z* Zimbabwe, *SimS* Seychelles, *K* Kenya, *Sech*: *D. sechellia*

* significant ($P < 0.005$, Bonferroni corrected threshold)

Table 3 Number of genes over-/under-expressed in *D. simulans* compared with *D. sechellia*

Population	Total	Over-expressed[a]	Under-expressed[b]
Zimbabwe	347	148	199
Kenya	337	144	193
Seychelles	353	158	195
France	518	214	304

[a] Genes over-expressed in *D. simulans* compared with *D. sechellia*

[b] Genes under-expressed in *D. simulans* compared with *D. sechellia*

Five terms were consistently over-represented in the subset of genes under-expressed in *D. sechellia* compared to every population of *D. simulans*. The molecular function "electron carrier activity" and the cellular components "vesicular fraction" and "microsome" refer to cytochrome P450 genes, as was assessed by examining the intersection of the genes with this annotation in *D. melanogaster*, and our set of differentially expressed genes.

The two other terms (namely "lipid metabolism" and "hormone catabolism") refer to three genes: *Juvenile hormone epoxide hydrolase 1* (*Jheh1*), *Juvenile hormone epoxide hydrolase 3* (*Jheh3*) and *Dopamine N-acetyltransferase* (*Dat*). These three genes are highly pleiotropic as they are directly involved in the regulation of key hormones: Juvenile hormone and Dopamine (DA). *Jheh1* and *Jheh3* are involved in JH regulation by degrading it. According to Gruntenko and Rauschenbach (2008), the JH titre can be assessed by the JH degradation level. Thus, we can assume that an over-expression of *Jheh1* and *Jheh3* in *D. simulans* compared with *D. sechellia* implies a lower JH titre in *D. simulans*. The gene *Dat* is involved in the degradation of DA.

Hybrids vs. parental populations

Expression in male hybrids was compared with males of both parental species, and the mean expression of the parents. By this process, the differentially expressed genes were those different from the parents and therefore not showing the dominance of a parental allele on the other. This also excluded genetic additive effects since their expression had to be significantly different from the parents' mean. A significant difference could be due to underdominance, i.e. failed interaction between the two alleles, or misregulation through negative epistasis. We found few genes perturbed in hybrids. No gene disruption was detected in hybrids obtained with the populations of *D. simulans* from Kenya and from Seychelles. Significant over-expression was detected for four and five genes in hybrids offspring of the French and Zimbabwean populations, respectively (Table 4). X-linked genes were significantly over-represented in this set of genes (Fisher exact test, $P < 0.001$). One gene was common in both comparisons (*Cp110*). This gene is involved in centriole replication, although its precise function is unknown (Dobbelaere et al. 2008). Another gene (*r-cup*) is involved in male meiosis, and its disturbance could have a role in hybrid sterility (Barreau et al. 2008).

Table 4 Genes over-expressed in hybrids compared to parents

Offspring of	Gene localisation
F *D. simulans* × *D. Sechellia*	
Cp110	X
CG14785	X
CG4558	X
Es2	X
Z *D. simulans* × *D. Sechellia*	
Cp110	X
sm	2R
r-cup	X
CG3795	X
CG31108	3R

No genes were found differentially expressed with offspring of the Kenyan and Seychelles population

F French, *Z* Zimbabwean

We also examined patterns of dominance and additivity. For this analysis, we included only genes for which expression is different in the two parental populations. We used the ratio dominance over additivity (d/a). About 45% of genes showed additivity to partial dominance ($0 < d/a < 0.5$), about 26% showed partial to complete dominance ($0.5 < d/a < 1$), and about 29% showed overdominance (Fig. 1).

Discussion

Our multiple comparisons of transcriptomes revealed three main features: 1-geographic differentiation in *D. simulans;* 2- expression divergence between *D. simulans* and *D. sechellia* (about 7% of the genes), notably cytochrome P450 genes, and genes involved in hormone metabolism (JH and DA); 3- only eight genes misregulated in hybrids among which X-linked genes were over-represented.

Assessment of the use of interspecific array

We used arrays carrying cDNA from *D. melanogaster*. The use of interspecific arrays has been shown to introduce a bias in comparisons, due to differential hybridisation caused by sequence divergence (Gilad et al. 2005; Oshlack et al. 2007). However, *D. melanogaster*, one of the closest sister species of *D. simulans* and *D. sechellia*, is approximately as

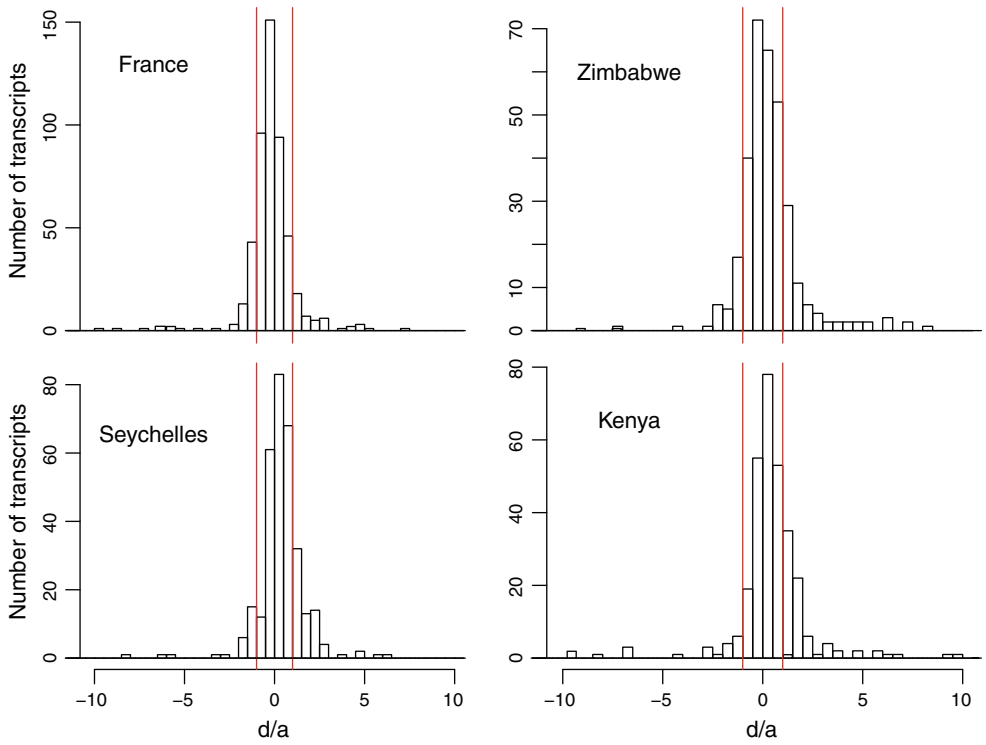

Fig. 1 Distribution of the ratio d/a for hybrids crossed from our four different populations of *D. simulans*. When d/a = 0, d = 0, we are thus showing perfect additivity. From 0 to 1 (or −1), we shift from complete additivity to complete dominance. Over 1 (under −1), we show overdominance

divergent from *D. simulans* and *D. sechellia*. The divergence between *D. sechellia* and *D. simulans* is thought to have occurred around 250.000 years ago, while the divergence between the two species and *D. melanogaster* dates back 2 to 3 million years (Lachaise et al. 1988; Hey and Kliman 1993; McDermott and Kliman 2008). We can thus assume the hybridisation signal to be similarly affected by the use of a *D. melanogaster* array. Moreover, in order to limit the possible bias, we chose a long cDNA array rather than oligonucleotide, which limited the effect of possible mismatches in the sequence. To assess whether there was a hybridisation bias, we looked for a correlation between: 1- the sequence divergence difference between each species and *D. melanogaster* and 2- the mean differences of expression between the two species. We assessed this in a large subset of genes included in the array (2,303 randomly chosen genes), using the sequences from the database (http://www.flybase.org). We observed no correlation (Spearman correlation coefficient: Rho = −0.0054; $P = 0.8056$). *D. simulans* and *D. sechellia* diverge by less than 2%, which is comparable with the level of divergence of disparate populations of *D. melanogaster*. 80% of divergent bases with *D. melanogaster* are shared between *D. simulans* and *D. sechellia* (Dworkin and Jones 2009). In this study, they found little evidence of any bias between *D. sechellia* and *D. simulans*. Although on a global scale, the influence of mismatch is likely negligible, we agree that individual genes can be affected. However, Mezey et al. (2008) showed that the effect of sequence divergence is mainly an increased variance leading to a decreased power of the test. Our tests are thus likely conservative. To assess a possible bias, we used sequences from Flybase to calculate the difference in coding sequence divergence (*D. melanogaster* vs. *D. simulans* minus *D. melanogaster* vs. *D. sechellia*). This was calculated first for differentially expressed genes between *D. sechellia* and *D. simulans*, and second for the random set of genes previously mentioned. Differentially expressed genes are represented by a subset of 232 genes taken from our list of 304 genes consistently differentially expressed between *D. sechellia* and *D. simulans*. We could not consider all genes because of incomplete/erroneous ortholog annotation in Flybase. However, in both sets (random and differentially expressed), genes were sorted the same way. Although the P-value is relatively low, the mean difference in divergence between random genes and differentially expressed genes is not significant (Mann–Whitney, W = 286,527, *P*-value = 0.06643). We assessed further the extent of the possible bias using a Q-Q plot of the distributions (Supplementary Fig. 5). This revealed that about 30 genes out of the 232 are biased toward a higher divergence compared to the random genes distribution. This bias can have two explanations. The first one is that

sequence divergence affected hybridisation, leading to false positive. The second explanation is biological: genes with a higher coding sequence divergence can also be genes evolving faster in terms of expression changes. Although there might be some bias, cytochrome genes as well as *Jheh1*, *Jheh3*, and *Dat* show a divergence difference below 1.3%, situated in the unbiased part of the Q-Q plot. This bias may have two consequences: a decreased power for the test, and a slight increase in the false positive rate.

Differentiation between *D. simulans* populations

Expression differentiation

D. simulans originates from Eastern Africa or Madagascar (Lachaise et al. 1988; Lachaise and Silvain 2004; Dean and Ballard 2004; Kopp et al. 2006). This ancestral area is consistent with several observations of our study. We would like to point out the difference between the African populations of *D. simulans*, and the French population which are revealed with all comparisons. Each of the African population revealed around 350 expression differences with *D. sechellia*, while the French one revealed 518, thus showing a significantly stronger differentiation compared with *D. sechellia*. The direct comparison revealed no differential expression between African populations of *D. simulans*, while we detected several differentially expressed genes between the French population and each African population. This is consistent with the weak but existing population structure observed on microsatellites by Schöfl and Schlötterer (2006) and on nuclear genes by Baudry et al. (2006). Gene flow should be higher between the three African populations than with the French population, as expected from an isolation by distance model and the evolutionary history of the species. A low level of existing geographic differentiation was also described for morphological traits (Gibert et al. 2004). Among African populations, the study of Schöfl and Schlötterer (2006) shows a differentiation between sub-Saharan populations and non sub-Saharan populations. Within the sub-Saharan population a differentiation was observed between the lines from Zimbabwe and Malawi on the one hand, and the lines from Uganda on the other hand, the latter being geographically closer to the likely region of origin of *D. simulans*. Our study suggests a similar differentiation between the Zimbabwean populations and the population from the ancestral zone. Indeed, when we compare F1 interspecific hybrids with their parents, the Zimbabwean population leads to 5 differentially expressed genes (Table 4). This result is comparable with what is observed for the French population (4 genes differentially expressed), but contrasts with the lack of differences observed for the two populations close to the likely origin of *D. simulans*: the Kenyan and the Seychelles populations. The Zimbabwean population shows a very specific pattern. It has been shown to be clearly apart from the eastern populations at the genetic level, but it still has quite high polymorphism (Baudry et al. 2006). Although results in terms of expression differences are contrasted, this population will be further considered as derived.

The study of Baudry et al. (2006) revealed significant differentiation between Eastern African populations and both the French and the Zimbabwean population using four X-linked loci. The authors did not detect any differentiation between populations from Madagascar, Mayotte, Kenya and Tanzania, which is consistent with our observation of similar expression profiles for *D. simulans* of the Seychelles Islands and the Kenya. However we did not detect gene expression differences between the population from Zimbabwe and the two other African populations with direct comparisons. It is likely the differentiation observed on DNA sequences (consistent with hybrid and variance analysis of the present study) is not strong enough to appear on a direct expression comparison, or does not affect expression. This result is consistent with what is known of *D. simulans* biogeographic history: a strong intra-population variation and a relatively weak differentiation between populations (Lachaise et al. 1988; Lachaise and Silvain 2004).

Population structure in expression has been shown in other organisms. In humans, Storey et al. (2007) showed population structure in expression between European and Nigerian cell cultures, although these results are still controversial due to possible differences in the cell immortalisation process (Davis and Kohane 2009). Adaptation to local environment at the level of gene expression has also been shown in other organisms such as *Saccharomyces cerevisae* (Townsend et al. 2003) or on a teleost fish which showed adaptation to temperature (Oleksiak et al. 2002; Whitehead and Crawford 2006).

Population variance, an evidence for decanalization?

We observed a higher interspecific expression variance in derived populations compared with populations from the ancestral area (Eastern Africa/Madagascar). It is interesting to note that DNA sequence variation shows the opposite trend (Schöfl and Schlötterer 2004; Baudry et al. 2006). This resembles the pattern produced by the process of decanalization, i.e. the revelation in a new environment of existing cryptic variation (Gibson and Wagner 2000; Gibson and Dworkin 2004). This hypothesis predicts phenotypic constraints in the ancestral area due to stabilizing selection. At the genetic level, variation can accumulate and not be expressed in the phenotype, because of various buffering mechanisms such as epigenetic interactions,

environmental constraints, etc. When the genotype is transferred in a new environment (for example when there is invasion of a new area), the environment changes, and so does the selective pressure. This can in turn result in the expression of the cryptic genetic variation, which increases the phenotypic variance (in our case, the variance in expression level). This was notably observed for mesosternal bristle number in a drosophilid (Yassin et al. 2007). This process happens despite a reduction in sequence polymorphism due to bottleneck and founding effects (Lachaise et al. 1988; Lachaise and Silvain 2004). However, decanalization has been only described with specific traits, and never at a general scale. Whether this process can be seen by assessing expression variance in all genes is still unknown. Moreover, studying different traits in several populations of *D. simulans*, Capy et al. (1993) revealed morphometric clines, but no difference in variance of traits between derived populations and those from the ancestral range. Adopting a similar approach, we checked expression data from *D. melanogaster* for a similar pattern (data from Hutter et al. 2008). We could not detect any significant variance difference between European and African populations. As *D. melanogaster*'s invasion of the world is older than *D. simulans*'s, it is possible stabilizing selection has acted on regulation, effectively dropping the expression variance in derived populations. The variance difference in *D. simulans* seems biologically significant since it matches the invasion pattern of the species. This tends to eliminate the hypothesis of a technical bias. However, we could not confirm our observation, either on other expression data, or on phenotypic data. This result therefore deserves further investigation.

Gene expression divergence between species

Our study shows 304 genes consistently differentially expressed between the two species. We compared the genes differentially expressed in our study with those revealed by Dworkin and Jones (2009). Thirty-four percent of their differentially expressed genes were present in our array. Out of these, 16% (60 genes) were also in our list of 304 genes differentially expressed between *D. simulans* and *D. sechellia*. The discrepancies can be explained by differences in the design (we compared males while they compared mixed sexes), lines choice (ancient laboratory lines versus recent wild lines), and power of the arrays.

Cytochrome P450 gene family: detoxification or hormone regulation

Cytochrome P450 is a large family of 83 functional genes in *D. melanogaster* (Tijet et al. 2001). The down-regulation of cytochrome P450 genes in *D. sechellia* compared with *D. simulans* (also observed by Dworkin and Jones 2009) can probably be explained by the role of this gene family in detoxification. It could be related to the specialization of the species on the toxic plant *Morinda citrifolia* (Dworkin and Jones 2009). The strict association between *D. sechellia* and its host could have reduced the variety of toxins *D. sechellia* is exposed to, releasing selective constraints on detoxification genes such as the cytochrome P450 gene family. It is also possible that some differences in the need of detoxification genes arose from different environmental conditions (not related to the specialization on *M. citrifolia*). However, we observe cytochrome P450 regulation divergences in comparisons of *D. sechellia* with all four populations of *D. simulans*, despite the fact that they come from four different geographic areas. If the latter hypothesis was supported, we would likely observe differences between all *D. simulans* populations.

Interestingly, the cytochrome P450 genes are also significantly over-represented in the twelve genes that are over-expressed in the French population compared with at least one African population (Supplementary Table 1). Four genes are cytochrome P450s, and one (*Walrus*) has a similar molecular function. It is possible the French population encounters a wider set of toxins due to anthropization processes (Dworkin and Jones 2009) leading to stronger constraints on the cytochrome P450 gene family. However, without data concerning pollutants at the sites of sampling, it is difficult to verify this hypothesis. Alternatively, the divergence of expression of these genes could be related with the involvement of these genes in hormone metabolism (Feyereisen 1999; Tijet et al. 2001), and especially JH regulation.

Divergence of hormonal regulation

We have observed divergence of expression for three genes involved in hormone (notably juvenile hormone and Dopamine) regulation: *jheh1*, *jheh3*, and *dat*. Although the role of JH has been widely described in females of *D. melanogaster* (Gruntenko and Rauschenbach 2008; Liu et al. 2008), it is still poorly known in adult males. However, a physiologic approach has shown a role of JH in seminal fluid protein accumulation in the male reproductive accessory glands (Wolfner et al. 1997), a role also supported by the mutant-based study of Wilson et al. (2003). Mutants with weak receptivity to JH show lower protein accumulation in these glands, and this can in turn affect male fertility. Mutant males also show very little interest in courtship, suggesting a role (either direct or indirect via the perturbation of accessory glands' protein synthesis) of JH in courtship behavior. Dopamine (DA) also plays a role in courtship behavior, a role that could be consistent with the misregulation of *Dat*. The changes in the regulatory

pathway of JH between the two species suggests a change of reproductive behavior in males, which could possibly correspond to a change in females, i.e. coevolution of both sexes via sexual selection. While highly speculative, this hypothesis could be a source (as much as a consequence) of the reproductive/behavioural isolation between the two species. DA has been shown to be involved in JH regulation in females, but has apparently little effect on JH in males as has been shown in *D. virilis* (Gruntenko and Rauschenbach 2008). It is however likely that a change in DA level will affect reproduction in males, but this won't be via JH.

Sterile hybrids, yet weak gene expression differences

We observed 8 genes over-expressed in interspecific hybrids compared to their parental populations, and none under-expressed. Six of these genes are located on the X chromosome. Genes were differentially expressed only for offspring of the two most differentiated hybrid populations: Zimbabwe and France. One gene was common in the two comparisons (*Cp110*).

Impact of the X-chromosome

An interesting result of our analysis is that out of the 8 genes mis-regulated in hybrids, six are located on the X chromosome, a number significantly higher than expected under the assumption of random localization of the differentially expressed genes. This observation is consistent with the so-called "faster-X" effect, which is a commonly mentioned but still controversial characteristic of speciation (Betancourt et al. 2002; Thornton and Long 2002; Musters et al. 2006; Begun et al. 2007; Masly and Presgraves 2007; Presgraves 2008; Vicoso and Charlesworth 2009). According to this theory, X-linked genes evolve more rapidly than genes on autosomes, perhaps due to higher efficiency of selection on the hemizygous X in males. X-linked genes have a specific evolution, due to their presence two-thirds of the time in females and one-third in males and due to their smaller population size than autosomes. However, using introgression, Hollocher and Wu (1996) found no higher density of sterility factor in the X chromosome than on autosomes. This suggests that the X-linked disturbance causing sterility is linked to divergence in regulation and not directly to sequence divergence of X-linked genes.

A low hybrid/parent regulation differentiation

The weak hybrid/parents differentiation observed in our study can be somewhat surprising, compared to results obtained in other studies. For example, using gene expression data on testes, Haerty and Singh (2006) found 241 genes differentially expressed between hybrid and parents. Three differences in the experimental design can explain this discrepancy. First, Haerty and Singh (2006) did not differentiate additive effects. Thus, it is possible that some of their differentially expressed genes represent in fact an averaged expression between the two parental genomes, a possibility we have excluded here. Second (and main) point, their study was on testes. They focused on an organ that is strongly affected as hybrid males between these two species are sterile (Cabot et al. 1994; Joly et al. 1997). Therefore they must have revealed genes involved in this sterility. Our goal was more to examine a global divergence. Therefore, we adopted a whole-body approach, which limited the detection of organ specific divergences, but highlighted more ubiquitous and global changes. Testes have a very specific expression pattern, likely perturbed in hybrids, but these particular differences would be hidden by our approach (Wolgemuth and Watrin 1991; Grimes 2004). Finally, the *D. simulans* line used by Haerty and Singh (2006) was a laboratory strain originating from Arizona, a population geographically far from the African native area of the species where we collected our samples.

The study of Michalak and Noor (2003) revealed 51 genes differentially expressed between hybrids and parents using *D. mauritiana* and *D. simulans*. None of these genes were found differentially expressed in our study. However, it has been previously shown that factors involved in hybrid sterility are different between *simulans/sechellia* vs. *simulans/mauritiana* hybrids (Coyne et al. 1991; Cabot et al. 1994). Another study detected 220 differentially expressed genes (Moehring et al. 2007). No gene is commonly differentially expressed between their study and ours. In this study, there is no correction for multiple testing, potentially allowing for a large number of false positives. If we adopted the same approach, our number of differentially expressed genes would jump to an average of about 67 (Supplementary Table 3), which is similar ($Chi^2 = 1.97$, $df = 1$, $P = 0.16$) to what is observed by Moehring et al. (2007). Interestingly, all these studies, as well as others, in *Drosophila* (Michalak and Noor 2003, 2004; Ranz et al. 2004; Landry et al. 2005; Haerty and Singh 2006) or various other organisms (Wang et al. 2006; Malone et al. 2007; Renaut et al. 2009; Mavarez et al. 2009) detected a large number of genes under-expressed in hybrids compared to parents, and only a few over-expressed. Contrasting with this observation, the present study only showed over-expressed genes in hybrids with FDR correction applied. Without the FDR correction, the representation of over/under-expressed genes is not consistent between the different hybrid populations. In fact, we have strong contrasts between the four different comparisons of hybrids with parents (Supplementary Table 3). This could be related with differences in allele profile between the

parental *D. simulans* populations, therefore allowing for different misregulation patterns in hybrids. Furthermore, the populations we used may lead to hybrids with different properties, owing to the fact that we chose African populations, while other studies have chosen more recently derived American isofemale lines. This aspect remains to be more thoroughly explored.

Intermediate additivity of expression

We observed about 45% of genes showing additivity in expression in the hybrid. This suggests an intermediate pattern compared to what was observed in other studies, from a few percent (Gibson et al. 2004; Haerty and Singh 2006; Moehring et al. 2007) up to 71% (Hughes et al. 2006; Rottscheidt and Harr 2007). The possible reasons for these discrepancies are numerous, as detailed by Rottscheidt and Harr (2007): amount of inbreeding of the parental lines, methodological differences and phylogenic distance between the parents. The last argument probably explains why, using the same method, we observed more dominance than Hughes et al. (2006). Their parental lines are isofemale lines of the same species, while we used populations of two closely related species; it is therefore expected that we would observe more expression disturbances in the form of overdominance (Hughes et al. 2006). It is worth noting that this overdominance is probably still quite small for each individual gene, since we observed very few genes showing expression outside the range of their parental populations.

Cis-regulation, a major player?

Our observations of very few cases of misexpression in hybrids (=very little perturbation of *trans*-regulation) and a large set of expression changes between the two species suggest a major role of *cis*-regulation on the divergence between the two species. This is consistent with an evolutionary model of stronger constraints on *trans*-regulation: the pleiotropic role of transcription factors makes them likely to be more constrained than *cis*-factors, which usually affect only one locus, or even one allele of a given locus. This observation is consistent with other studies on *Drosophila* species (Wittkopp et al. 2004, 2008a, b).

Acknowledgments FW was supported by a PhD fellowship from the "Institut Ecologie et Environnement" of the "Centre National de la Recherche Scientifique". The authors wish to thank Dr Wilfried Haerty for advice both during the analysis and about the manuscript. We also would like to thank Dr Amir Yassin and Dr Francesco Catania for fruitful discussions improving this manuscript. We thank Dr Jean-Jacques Daudin for advice in the statistical analysis. We also thank Dr Michel Piovant for giving access to the array platform. We thank people who provided us with the biological material: Dr Bruno Le Rü, Dr Roland Allemand and Dr Daniel Lachaise (deceased in 2006).

Conflict of interest The authors declare no conflict of interest.

References

Altman N (2005) Replication, variation and normalisation in microarray experiments. Appl Bioinformatics 4:33–44

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H et al (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet 25:25–29

Barreau C, Benson E, Gudmannsdottir E, Newton F, White-Cooper H (2008) Post-meiotic transcription in *Drosophila* testes. Development 135:1897–1902

Baudry E, Derome N, Huet M, Veuille M (2006) Contrasted polymorphism patterns in a large sample of populations from the evolutionary genetics model *Drosophila simulans*. Genetics 173:759–767

Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP et al (2007) Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol 5:e310

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B 57:289–300

Berriz GF, King OD, Bryant B, Sander C, Roth FP (2003) Characterizing gene sets with funcassociate. Bioinformatics 19:2502–2504

Betancourt AJ, Presgraves DC, Swanson WJ (2002) A test for faster X evolution in drosophila. Mol Biol Evol 19:1816–1819

Cabot EL, Davis AW, Johnson NA, Wu CI (1994) Genetics of reproductive isolation in the *Drosophila simulans* clade: complex epistasis underlying hybrid male sterility. Genetics 137:175–189

Capy P, Pla E, David J (1993) Phenotypic and genetic variability of morphometrical traits in natural populations of *Drosophila melanogaster* and *D. simulans*. I. geographic variations. Genet Sel Evol 25:517–536

Coyne JA, Rux J, David JR (1991) Genetics of morphological differences and hybrid sterility between *Drosophila sechellia* and its relatives. Genet Res 57:113–122

Davis AR, Kohane IS (2009) Expression differences by continent of origin point to the immortalization process. Hum Mol Genet 18:3864–3875

Dean MD, Ballard JWO (2004) Linking phylogenetics with population genetics to reconstruct the geographic origin of a species. Mol Phylogenet Evol 32:998–1009

Dobbelaere J, Josué F, Suijkerbuijk S, Baum B, Tapon N et al (2008) A genome-wide RNAi screen to dissect centriole duplication and centrosome maturation in *Drosophila*. PLoS Biol 6:e224

Dobzhansky T (1936) I. Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. Genetics 21:113–135

Dworkin I, Jones CD (2009) Genetic changes accompanying the evolution of host specialization in *Drosophila sechellia*. Genetics 181:721–736

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Longman, London

Feyereisen R (1999) Insect p450 enzymes. Annu Rev Entomol 44:507–533

Gibert P, Capy P, Imasheva A, Moreteau B, Morin JP et al (2004) Comparative analysis of morphological traits among *D. melanogaster* and *D.* simulans: genetic variability, clines and phenotypic plasticity. Genetica 120:165–179

Gibson G, Dworkin I (2004) Uncovering cryptic genetic variation. Nat Rev Genet 5:681–690

Gibson G, Wagner G (2000) Canalization in evolutionary genetics: a stabilizing theory? Bioessays 22:372–380

Gibson G, Riley-Berger R, Harshman L, Kopp A, Vacha S et al (2004) Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. Genetics 167:1791–1799

Gilad Y, Borevitz J (2006) Using DNA microarrays to study natural variation. Curr Opin Genet Dev 16:553–558

Gilad Y, Rifkin SA, Bertone P, Gerstein M, White KP (2005) Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. Genome Res 15:674–680

Grimes SR (2004) Testis-specific transcriptional control. Gene 343: 11–22

Gruntenko NE, Rauschenbach IY (2008) Interplay of JH, 20e and biogenic amines under normal and stress conditions and its effect on reproduction. J Insect Physiol 54:902–908

Haerty W, Singh RS (2006) Gene regulation divergence is a major contributor to the evolution of Dobzhansky-Muller incompatibilities between species of *Drosophila*. Mol Biol Evol 23: 1707–1714

Hamblin MT, Veuille M (1999) Population structure among African and derived populations of *Drosophila simulans*: evidence for ancient subdivision and recent admixture. Genetics 153:305–317

Hey J, Kliman RM (1993) Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. Mol Biol Evol 10:804–822

Hollocher H, Wu CI (1996) The genetics of reproductive isolation in the *Drosophila simulans* clade: X vs. autosomal effects and male vs. female effects. Genetics 143:1243–1255

Holloway AK, Lawniczak MK, Mezey JG, Begun DJ, Jones CD (2007) Adaptive gene expression divergence inferred from population genomics. PLoS Genet 3:2007–2013

Hughes KA, Ayroles JF, Reedy MM, Drnevich JM, Rowe KC et al (2006) Segregating variation in the transcriptome: *cis* regulation and additivity of effects. Genetics 173:1347–1355

Hutter S, Saminadin-Peter SS, Stephan W, Parsch J (2008) Gene expression variation in African and European populations of *Drosophila melanogaster*. Genome Biol 9:R12

Joly D, Bazin C, Zeng LW, Singh RS (1997) Genetic basis of sperm and testis length differences and epistatic effect on hybrid inviability and sperm motility between *D. simulans* and *D. sechellia*. Heredity 78:354–362

Kerr MK, Martin M, Churchill GA (2000) Analysis of variance for gene expression microarray data. J Comput Biol 7:819–837

Kerr M, Afshari C, Bennett L, Bushel P, Martinez J et al (2002) Statistical analysis of a gene expression microarray experiment with replication. Stat Sinica 12:203–217

Kliman RM, Andolfatto P, Coyne JA, Depaulis F, Kreitman M et al (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. Genetics 156:1913–1931

Kopp A, Frank A, Fu J (2006) Historical biogeography of *Drosophila simulans* based on Y-chromosomal sequences. Mol Phylogenet Evol 38:355–362

Lachaise D, Silvain JF (2004) How two afrotropical endemics made two cosmopolitan human commensals: the *D. melanogaster-D. simulans* palaeogeographic riddle. Genetica 120:17–39

Lachaise D, Cariou M, David J, Lemeunier F, Tsacas L et al (1988) Historical biogeography of the *Drosophila melanogaster* species subgroup. Evolutionary biology, pp 159–225

Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG et al (2005) Compensatory *cis-trans* evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. Genetics 171:1813–1822

Lawniczak MKN, Holloway AK, Begun DJ, Jones CD (2008) Genomic analysis of the relationship between gene expression variation and DNA polymorphism in *Drosophila simulans*. Genome Biol 9:R125

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

Levene H (1960) Contributions to probability and statistics: essays in honor of Harold Hotelling. In: Olkin I (ed). Stanford University Press, Stanford, pp 278–292

Liu Z, Li X, Prasifka JR, Jurenka R, Bonning BC (2008) Overexpression of *Drosophila* juvenile hormone esterase binding protein results in anti-JH effects and reduced pheromone abundance. Gen Comp Endocr 156:164–172

Malone JH, Chrzanowski TH, Michalak P (2007) Sterility and gene expression in hybrid males of *X. laevis* and *X. muelleri*. PLoS One 2:e781

Masly JP, Presgraves DC (2007) High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. PLoS Biol 5:e243

Mavarez J, Audet C, Bernatchez L (2009) Major disruption of gene expression in hybrids between young sympatric anadromous and resident populations of brook charr (*Salvelinus fontinalis mitchill*). J Evol Biol 8:1708–1720

McDermott SR, Kliman RM (2008) Estimation of isolation times of the island species in the *Drosophila simulans* complex from multilocus DNA sequence data. PLoS One 3:e2442

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL (2003) Rapid evolution of male-biased gene expression in *Drosophila*. Proc Natl Acad Sci USA 100:9894–9899

Mezey JG, Nuzhdin SV, Ye F, Jones CD (2008) Coordinated evolution of co-expressed gene clusters in the *Drosophila* transcriptome. BMC Evol Biol 8:2

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

Michalak P, Noor MA (2004) Association of misexpression with sterility in hybrids of *D. simulans* and *D. mauritiana*. J Mol Evol 59:277–282

Moehring AJ, Teeter KC, Noor MA (2007) Genome-wide patterns of expression in *Drosophila* pure species and hybrid males. II. Examination of multiple-species hybridisations, platforms, and life cycle stages. Mol Biol Evol 24:137–145

Muller HJ (1942) Isolating mechanisms, evolution, and temperature. Biol Symp 6:71–125

Musters H, Huntley MA, Singh RS (2006) A genomic comparison of faster-sex, faster-X, and faster-male evolution between *Drosophila melanogaster* and *Drosophila pseudoobscura*. J Mol Evol 62:693–700

Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. Nat Genet 32:261–266

Oshlack A, Chabot AE, Smyth GK, Gilad Y (2007) Using DNA microarrays to study gene expression in closely related species. Bioinformatics 23:1235–1242

Pavey SA, Collin H, Nosil P, Rogers SM (2010) The role of gene expression in ecological speciation. Ann N Y Acad Sci 1206:110–129

Presgraves DC (2008) Sex chromosomes and speciation in *Drosophila*. Trends Genet 7:336–343

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

Ranz JM, Namgyal K, Gibson G, Hartl DL (2004) Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. Genome Res 14:373–379

Renaut S, Nolte AW, Bernatchez L (2009) Gene expression divergence and hybrid misexpression between lake whitefish

species pairs (*Coregonus spp.* salmonidae). Mol Biol Evol 26:925–936

Rottscheidt R, Harr B (2007) Extensive additivity of gene expression differentiates subspecies of the house mouse. Genetics 177:1553–1567

Schöfl G, Schlötterer C (2004) Patterns of microsatellite variability among X chromosomes and autosomes indicate a high frequency of beneficial mutations in non-African *D. simulans*. Mol Biol Evol 21:1384–1390

Schöfl G, Schlötterer C (2006) Microsatellite variation and differentiation in African and non-African populations of *Drosophila simulans*. Mol Ecol 15:3895–3905

Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J et al (2007) Gene-expression variation within and among human populations. Am J Hum Genet 80:502–509

Thornton K, Long M (2002) Rapid divergence of gene duplicates on the *Drosophila melanogaster* X chromosome. Mol Biol Evol 19:918–925

Tijet N, Helvig C, Feyereisen R (2001) The cytochrome p450 gene superfamily in *Drosophila melanogaster*: annotation, intron-exon organization and phylogeny. Gene 262:189–198

Townsend JP, Cavalieri D, Hartl DL (2003) Population genetic variation in genome-wide gene expression. Mol Biol Evol 20:955–963

Vicoso B, Charlesworth B (2009) Effective population size and the faster-X effect: an extended model. Evolution 63:2413–2426

Wang J, Tian L, Lee H, Wei NE, Jiang H et al (2006) Genomewide non-additive gene regulation in *Arabidopsis* allotetraploids. Genetics 172:507–517

Whitehead A, Crawford DL (2006) Neutral and adaptive variation in gene expression. Proc Natl Acad Sci USA 103:5425–5430

Wilson TG, DeMoor S, Lei J (2003) Juvenile hormone involvement in *Drosophila melanogaster* male reproduction as suggested by the methoprene-tolerant (27) mutant phenotype. Insect Biochem Mol 33:1167–1175

Wittkopp PJ, Haerum BK, Clark AG (2004) Evolutionary changes in *cis* and *trans* gene regulation. Nature 430:85–88

Wittkopp PJ, Haerum BK, Clark AG (2008a) Regulatory changes underlying expression differences within and between *Drosophila* species. Nat Genet 40:346–350

Wittkopp PJ, Haerum BK, Clark AG (2008b) Independent effects of *cis*- and *trans*-regulatory variation on gene expression in *Drosophila melanogaster*. Genetics 178:1831–1835

Wolfner MF, Partridge L, Lewin S, Kalb JM, Chapman T et al (1997) Mating and hormonal triggers regulate accessory gland gene expression in male *Drosophila*. J Insect Physiol 43:1117–1123

Wolgemuth DJ, Watrin F (1991) List of cloned mouse genes with unique expression patterns during spermatogenesis. Mamm Genome 1:283–288

Yassin A, Abou-Youssef AY, Bitner-Mathe B, Capy P, David JR (2007) Mesosternal bristle number in a cosmopolitan drosophilid: an x-linked variable trait independent of sternopleural bristles. J Genet 86:149–158