**Potential effect of library PCR enrichment on measures of transcript abundance**

The PCR enrichment performed during library construction has the potential to alter the relative abundance of some transcripts and thus affect our ability to accurately measure differential gene expression. To avoid this potential concern, we followed the Illumina protocol and kept the number of cycles low (18 cycles) to keep the amplification within the exponential range. Because we know the initial amount of each of our five spike-in controls, we examined their relative abundance after amplification to determine whether higher copy number transcripts were preferentially amplified during this step. The figure below shows the spike-in abundance of each spike in control, for each of the genotypes we sequenced:

![Figure showing spike-in abundance for D. mauritiana and D. sechellia w, Q1(A), and 3Q1(A).](image)

Although we can not know with certainty that the amplification was uniform across all of the transcripts in each pool, analysis of our spike-in control RNAs show that the amplification had little effect on their relative proportions in each sample and that their amplified amounts remain relatively linear on a log scale.
FDR stringency on number of differentially expressed genes identified

To identify genes as differentially expressed, we used the Benjamini-Hochberg method to control FDR. With a significance threshold of \( P \leq 0.01 \), the corresponding FDR was 0.05. Although this level of stringency is common, we also tried using a more stringent FDR threshold of 0.01 to attempt to further narrow our set of potential expression candidates. The results of this comparison are shown in the table below:

<table>
<thead>
<tr>
<th>Genotype 1</th>
<th>Genotype 2</th>
<th>Number of genes identified as differentially expressed with FDR &lt; 0.05</th>
<th>Number of genes identified as differentially expressed with FDR &lt; 0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D. mauritiana</em></td>
<td><em>D. sechellia</em></td>
<td>2261</td>
<td>2259</td>
</tr>
<tr>
<td><em>D. mauritiana</em></td>
<td><em>Q1(A)</em></td>
<td>2404</td>
<td>2401</td>
</tr>
<tr>
<td><em>D. mauritiana</em></td>
<td><em>3Q1(A)</em></td>
<td>2593</td>
<td>2591</td>
</tr>
<tr>
<td><em>D. sechellia</em></td>
<td><em>Q1(A)</em></td>
<td>236</td>
<td>236</td>
</tr>
<tr>
<td><em>D. sechellia</em></td>
<td><em>3Q1(A)</em></td>
<td>410</td>
<td>410</td>
</tr>
</tbody>
</table>

These results show the number of differentially expressed genes varies only slightly by changing the FDR cutoff, and that changing the cutoff makes no difference for some of the comparisons.

Estimates of transcript splice isoform differences

MapSplice identifies splice junctions where multiple mRNA-seq reads require a gapped alignment to map to the reference genome; these alignment gaps correspond to one or more putative splice junctions. The quality and diversity of alignments that include each putative splice junction are used to calculate a significance score for each junction site. Splice junctions that received the highest significance scores are then reported as putative splice junctions by MapSplice. We used the number of reads that span these junctions and the number of reads that skip these junctions to calculate a splicing ratio for that junction in each of the four genotypes we studied. The splicing ratio was defined as the number of times the junction is skipped to the sum of the number of times it is skipped and
the number of times it is expressed. Thus, a splicing ratio of zero implies that the junction was always present in all transcript isoforms of that gene for a given genotype, and a splicing ratio of one implies that the junction was skipped in every transcript isoform of that gene that we were able to identify computationally.
File S2

Supporting information

Introgression effects on size and shape

Introgression hybrids NENEH2(A), DEE1(B), and 2H3(B) all possess significant differences in posterior lobe area compared to D. sechellia w (F_{14,110}=2.31, P=0.008; F_{14,131}=3.96, p<0.001; F_{14,124}=2.18, P=0.012, respectively), but posterior lobe area in each of these introgression hybrid genotypes was significantly larger than that of D. sechellia w. Also, although introgression 2H3(B) had a statistically significant effect on posterior lobe shape, the distribution of principal components appeared to move away from that of D. mauritiana (see Figure S1). The posterior lobe shape of introgression hybrid 2H3(B) appears to possess wider neck and baseline regions compared to D. sechellia w. To test this apparent difference in the posterior lobe width, we measured the length of the artificial baseline used to enclose our posterior lobe outlines in D. sechellia w and 2H3(B). Introgression hybrid 2H3(B) does indeed possess significantly longer baselines than D. sechellia w (\bar{X}_{sech}=0.076±0.001 mm, \bar{X}_{2H3(B)}=0.089±0.001 mm, P<0.0001). This 2H3(B) posterior lobe shape that lies outside of the range of the pure species shapes might result from either transgressive segregation or Dobzhansky-Muller incompatibilities between D. mauritiana and D. sechellia alleles as discussed in the text.

Sensitivity of morphometric techniques

Introgressions 4C2(A) and Q1(A) both possess significantly smaller posterior lobe areas, but have no significant affects on area-normalized posterior lobe shape. It seems possible that because these introgressions possess shapes not significantly different from that of D. sechellia w, this aspect of their morphology might lessen the statistical effect of their small but significant differences in posterior lobe area when size and shape are combined when calculating elliptical Fourier coefficients. (Similar results were also obtained for introgressions that possess significantly larger posterior lobe areas than D. sechellia w, but no significant difference in posterior lobe shape). To determine if our morphometric methods were in fact sensitive enough to detect differences in morphology when posterior lobe size and shape are combined, we dissected and measured 20 F1 hybrid males obtained from crosses between D. sechellia w females and D. mauritiana males. These males show substantial differences in both posterior lobe size and posterior lobe shape compared to D. mauritiana and D. sechellia w (Figure 1). Compared to D. sechellia w, F1 hybrid males have significantly reduced posterior lobe area (\bar{X}=3.110±0.0929 \times 10^{-3} \text{ mm}^2, P=3.75 \times 10^{-14}) and also show significant differences in posterior lobe shape when we normalize area (F_{14,61}=4.40, P=2.40 \times 10^{-7}). When we included
both posterior lobe area and shape in the elliptical Fourier coefficients that describe F₁ hybrid male morphology we also obtained a significant result ($F_{14.61} = 2.93$, $P = 1.83 \times 10^{-3}$). However, although the difference in morphology when both size and shape are analyzed together is statistically significant, the magnitude of this difference is lower than those of either area or shape when each is analyzed individually. Thus, it appears that the nonsignificant test results we obtained using the combined size and shape morphologies for $4C2(A)$ and $Q1(A)$ might reflect the relative sensitivity of our morphometric assays.
**Figure S1** Distribution of the first two principal components calculated using area-normalized elliptical Fourier coefficients for introgression hybrids that show significant effects on posterior lobe shape. *D. sechellia* is shown with red triangles, *D. mauritiana* is shown with blue circles, and introgression hybrid genotypes are shown with black diamonds. Ellipses show 70 percent normal-probability contours for each sample.
Figure S2  Mapping performance. Increasing numbers of sequence mismatches were tested to identify the number of mismatches for each end of a paired-end read that allowed a substantial proportion of transcriptome reads to map uniquely to the D. melanogaster genome. D. mauritiana P-insertion line Q1 is shown in black, D. sechellia w is shown in gray, 3Q1(A) is shown in blue, and Q1(A) is shown in red. Roughly 90 percent of end reads map uniquely using ≤12 mismatches.
Table S1  Genomic distribution of the *D. mauritiana-D. sechellia* introgression lines.

Table S1 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.111.130815/DC1.
Table S2  Number of genes differentially expressed using various mismatch criteria.

Table S2 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.111.130815/DC1.
Table S3  Differential gene expression in *D. mauritiana* versus *D. sechellia* male genital imaginal disc.

Table S3 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.111.130815/DC1.

Genes with *D. mauritiana*-biased expression are not highlighted and genes with *D. sechellia*-biased expression are highlighted in gray. "MAUR / SECH" shows the ratio of the average normalized *D. mauritiana* to *D. sechellia* expression values across biological replicates. Expression levels were estimated using a Generalized Poisson (GP) model. Differentially expressed genes were identified using *P*<0.01 after FDR correction of 0.05 as statistical criteria.
Table S4  Differential gene expression in Q1(A) and 3Q1(A) D. mauritiana-D. sechellia introgression hybrid male genital imaginal disc.

Table S4 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.111.130815/DC1.

Set of 100 candidate genes that are differentially expressed between D. mauritiana and D. sechellia and differentially expressed between introgression hybrids 3Q1(A) or Q1(A) and D. sechellia, but not differentially expressed between introgression hybrids 3Q1(A) or Q1(A) and D. mauritiana. Genes that were expressed at D. mauritiana levels in 3Q1(A), but not in Q1(A) are not highlighted, genes that were expressed at D. mauritiana levels in Q1(A), but not in 3Q1(A) are highlighted in yellow, and genes that were expressed at D. mauritiana levels in both introgression hybrids are highlighted in gray. Expression levels were estimated using a Generalized Poisson (GP) model. Differentially expressed genes were identified using \( P<0.01 \) after FDR correction of 0.05 as statistical criteria. Genes known to reside within either the Q1(A) or 3Q1(A) introgression regions at the current level of introgression breakpoint mapping resolution are marked by (*) next to their gene name.
Table S5  Putative transcript splice junctions in male genital imaginal disc transcripts.

Table S5 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.111.130815/DC1.

Splice junction start and end coordinates were determined using the D. melanogaster genome as a reference and coordinates correspond to the D. melanogaster gene annotations. Empty fields in the introgression hybrid genotypes indicate the locus was not differentially expressed in that genotype. "ND" indicates that a splice junction at those coordinates was not detected by MapSplice, "M" indicates that the splice junction at that site had a D. mauritiana-like splicing ratio in at least one of the introgression hybrids, and "S" indicates that the splice junction at that site had a D. sechellia-like splicing ratio in at least one of the introgression hybrids. The splicing ratio was defined as the number of times the junction is skipped to the sum of the number of times it is skipped and the number of times it is expressed.
Table S6  Sex-biased gene expression in the *D. melanogaster* genital imaginal disc.

Table S6 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.111.130815/DC1.

Genes with male-biased expression are not highlighted and genes with female-biased expression are highlighted in gray. Genes that were significantly differentially expressed were identified using \( P<0.01 \) after FDR correction of 0.05 (\( q<0.05 \)) as statistical criteria. "MALE / FEMALE" shows the ratio of the average normalized male to female expression values across microarrays. Male-biased genes known to reside within *D. mauritiana* introgressions with large morphological effects at the current level of resolution of introgression breakpoint mapping are marked by (*) next to their gene name.
Table S7  Male- and species-biased gene expression in the male genital imaginal disc.

Table S7 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.111.130815/DC1.

Gene list of 27 genes that show significant differential expression between the *D. mauritiana* and *D. sechellia* male genital disc and also show significant male-biased expression in the *D. melanogaster* genital disc. Expression levels shown for *D. mauritiana* and *D. sechellia* were estimated using a Generalized Poisson (GP) model. "MAUR / SECH" shows the ratio of the average normalized *D. mauritiana* to *D. sechellia* expression values across biological replicates. Differentially expressed genes between *D. mauritiana* and *D. sechellia* were identified using *P*<0.01 after FDR correction of 0.05 as statistical criteria. "MALE / FEMALE" shows the ratio of the average normalized *D. melanogaster* male to female expression values across microarrays. Differentially expressed genes between *D. melanogaster* males and females were identified using *P*<0.01 after FDR correction of 0.05 (*q*<0.05) as statistical criteria. Genes included within the 100 candidates expressed at *D. mauritiana* levels in Q1(A) or 3Q1(A) are marked next to their gene name by: (*) = expressed at *D. mauritiana* levels in Q1(A), but not 3Q1(A); (†) = expressed at *D. mauritiana* levels in 3Q1(A), but not Q1(A); ($) = expressed at *D. mauritiana* levels in both Q1(A) and 3Q1(A).