Divergence Between the *Drosophila pseudoobscura* and *D. persimilis* Genome Sequences in Relation to Chromosomal Inversions

Mohamed A. F. Noor,[*,1,2] **David A. Garfield,**[*,1] **Stephen W. Schaeffer**[†] **and Carlos A. Machado**[‡]

Biology Department, Duke University, Durham, North Carolina 27708, [†]Biology Department and Institute for Molecular Evolutionary Genetics, Pennsylvania State University, University Park, Pennsylvania 16802 and [‡]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721

ABSTRACT

As whole-genome sequence assemblies accumulate, a challenge is to determine how these can be used to address fundamental evolutionary questions, such as inferring the process of speciation. Here, we use the sequence assemblies of *Drosophila pseudoobscura* and *D. persimilis* to test hypotheses regarding divergence with gene flow. We observe low differentiation between the two genome sequences in pericentromeric and peritelomeric regions. We interpret this result as primarily a remnant of the correlation between levels of variation and local recombination rate observed within populations. However, we also observe lower differentiation far from the fixed chromosomal inversions distinguishing these species and greater differentiation within and near these inversions. This finding is consistent with models suggesting that chromosomal inversions facilitate species divergence despite interspecies gene flow. We also document heterogeneity among the inverted regions in their degree of differentiation, suggesting temporal differences in the origin of each inverted region consistent with the inversions arising during a process of divergence with gene flow. While this study provides insights into the speciation process using two single-genome sequences, it was informed by lower throughput but more rigorous examinations of polymorphism and divergence. This reliance highlights the need for complementary genomic and population genetic approaches for tackling fundamental evolutionary questions such as speciation.

THE availability of multiple annotated whole-genome sequence assemblies has allowed the scientific community to examine patterns of molecular evolution at an unparalleled scale. Despite this unprecedented opportunity, such data have been only rarely used to examine modes (or signatures) of the process of speciation (but see PATTERSON *et al.* 2006). Comparing genome sequences can potentially identify general patterns that are not obscured by diverse evolutionary forces that may act on single loci. However, such comparisons are complicated by having a sample size of one because typically only single genomes of each species have been sequenced. Hence, differences between the genome sequences may reflect divergence between the species or polymorphism within one or both species.

Over the last several years, the genomes of 12 Drosophila species have been sequenced, assembled, and annotated (GILBERT 2005), and more genome sequences are sure to follow. Among these are the co-occurring (sympatric) species *Drosophila pseudoobscura* (RICHARDS *et al.* 2005) and *D. persimilis*. These species are known to hybridize at low levels in nature (DOBZHANSKY 1973; POWELL 1983), and the signature of introgression is apparent in the sequences of some loci (MACHADO *et al.* 2002, 2007; MACHADO and HEY 2003). Introgression appears to be limited at loci within chromosomal inversions that distinguish these species (MACHADO *et al.* 2002, 2007), consistent with many recent theories regarding the role of chromosomal inversions in facilitating the persistence of hybridizing species (see reviews in ORTIZ-BARRIENTOS *et al.* 2002; BUTLIN 2005). Recently, MACHADO *et al.* (2007) examined polymorphism and divergence at 18 loci across the second chromosome of these species, and they observed that interspecies introgression is restricted ~2 Mb into collinear regions adjacent to the inversion. This finding suggests that the "islands of differentiation" between hybridizing species may be larger than the inverted regions alone, as predicted by some chromosomal speciation models (*e.g.*, NAVARRO and BARTON 2003).

Here, we examine the signature of speciation and divergence using the full genome sequences of *D. pseudoobscura* and *D. persimilis*. First, we examine general patterns of sequence divergence between these species across the lengths of their six chromosome arms. Using three "neutral" measures of divergence between these two single genome sequences, intergenic noncoding

[1]These authors contributed equally to this work.

[2]*Corresponding author:* Biology Department, Box 90338, Duke University, Durham, NC 27708. E-mail: noor@duke.edu

divergence, coding silent site, and intron divergence, we test whether introgression is restricted only within inverted regions or whether collinear regions near inversions also serve as islands of differentiation. Because of the close relationship of these species, we are able to use noncoding measures of divergence (Pollard *et al.* 2006). These three measures have different dynamics: coding silent site divergence can sometimes be affected by codon bias, and intron and intergenic divergence sometimes evolves under different constraints as well (*e.g.*, Ometto *et al.* 2006). Because the inverted regions have been important for the divergence of this species pair, we test whether specific classes of genes (based on Gene Ontology terms) are overrepresented in those regions and may have a functional connection to species differences. Finally, we discuss our findings in the contexts of theories of speciation and the utility of emerging whole-genome sequences for testing such theories.

D. pseudoobscura and *D. persimilis* bear five chromosomes: a metacentric X chromosome (having arms named XL and XR), three large telocentric autosomes (chromosomes 2–4), and one "dot" small autosome (chromosome 5). The chromosomal arms of these species correspond to the six "Muller's elements" (Muller 1940) as follows: Muller's element A, XL; B, chromosome 4; C, chromosome 3; D, XR; E, chromosome 2; and F, chromosome 5. The XL and second chromosomes of these species differ by fixed paracentric inversions spanning five to six major cytological sections (Dobzhansky and Tan 1936; Dobzhansky and Sturtevant 1938; Dobzhansky and Epling 1944), corresponding to just over 7 Mb of DNA sequence each (Machado *et al.* 2007). The XR chromosome arm is polymorphic for two arrangements in *D. persimilis*: the common standard and rare sex-ratio arrangements. The *D. persimilis* standard arrangement differs from the *D. pseudoobscura* arrangement by a paracentric inversion spanning 10 major cytological sections (almost half its length), corresponding to ∼13 Mb of DNA sequence. Rare strains of *D. persimilis* bear the sex-ratio arrangement, which has the same gene order as the *D. pseudoobscura* XR chromosome, but *D. persimilis* males bearing this arrangement produce all female offspring through a form of strong meiotic drive. The positions of the breakpoints of these three inversions are described in Machado *et al.* (2007), and in all three cases, the *D. persimilis* arrangement is derived and the *D. pseudoobscura* arrangement is the ancestral type (Dobzhansky and Tan 1936 and see results). The third chromosome is highly polymorphic for a very old gene arrangement polymorphism (Dobzhansky and Epling 1944; Lewontin *et al.* 1981; Schaeffer *et al.* 2003), including one arrangement shared between the species ("standard"). The strains of *D. pseudoobscura* and *D. persimilis* that had their genomes sequenced carry the "arrowhead" (Richards *et al.* 2005) and standard arrangements, respectively, and these differ by a single paracentric inversion (Dobzhansky and Sturtevant 1938; Dobzhansky and Epling 1944).

MATERIALS AND METHODS

Our analyses used comparative analysis freeze 1 (CAF1) of the *D. pseudoobscura* and *D. persimilis* genome sequences. While the lack of annotation of the *D. persimilis* genome sequence at the time of initial analysis complicated our efforts, the high degree of sequence similarity between the two species and their recent divergence made it possible to use the annotated coding sequences (CDS) from *D. pseudoobscura* (Richards *et al.* 2005) to predict intron–exon boundaries in the *D. persimilis* genome. Our annotation method consisted of two steps: (1) the identification of homologous genomic regions in the two species to identify the full-length genomic sequence for each *D. persimilis* gene with a *D. pseudoobscura* homolog and (2) an alignment between the *D. pseudoobscura* annotated CDS and the full-length *D. persimilis* gene identified in step 1 to delineate exon–intron boundaries. Filtering steps were applied at each step and at the end of the process to minimize the number of false homologies identified as well as to remove low-quality sequence reads (see below).

Full-length *D. pseudoobscura* gene sequences and the full, unannotated *D. persimilis* genome sequence were downloaded from http://rana.lbl.gov/drosophila/. To identify homologous regions, BLAST (Altschul *et al.* 1990) was used to align each *D. pseudoobscura* gene to the *D. persimilis* genomic sequence, and the best alignments were identified. Because BLAST optimizes local alignments, syntenic and homologous sequence may be separated into multiple alignments. Custom scripts were thus written to connect multiple high-scoring alignments to span the full *D. pseudoobscura* genome sequence. Alignments were kept only if the assembled *D. persimilis* sequences conserved the original gene synteny and were found on the same *D. persimilis* scaffold. If necessary, the start and end of the *D. persimilis* sequence were extended up to 10 bp to ensure that the full sequence of the first and last exons was included in the alignment. To minimize confusion between paralogs or other stretches of similar sequence, the *D. persimilis* sequence identified was then BLASTed back against a database of the *D. pseudoobscura* gene sequences and kept only if the best hit identified was the original *D. pseudoobscura* gene. Sequences with large stretches of N's were discarded.

Each pair of homologous sequences along with the corresponding *D. pseudoobscura* CDS was then aligned using the TBA and MULTIZ programs (Blanchette *et al.* 2004). We found that these programs generated vastly improved alignments compared to CLUSTAL, which is not designed to deal with the gaps generated by aligning CDS and full gene sequences (Higgins *et al.* 2005). Any subsequence within the *D. persimilis* genomic sequence that aligned only with the *D. pseudoobscura* genomic sequence was labeled as an intron, while any *D. persimilis* subsequence that aligned with both the *D. pseudoobscura* CDS and genomic sequence was labeled as an exon and part of the *D. persimilis* CDS. The resulting CDS alignments were passed, using Perl scripts, to PAML for estimation of K_s and K_a using the codeml program with the pairwise distance estimation option (runmode = −2) (Yang 1997). Intronic sequences were pairwise aligned using the Smith–Waterman algorithm included with ClustalW, and the similarity of the resulting alignment was calculated using custom scripts after removing the 2 bases at the 5′ and 3′ bases ends of the intron, which may be under selection to maintain splice acceptor/donor sites.

As an additional quality-control step, we removed from the analysis any genes in which a stop codon was found in the first

two-thirds of the *D. persimilis* CDS. This eliminated most alignment errors, pseudogenes, and *D. persimilis* sequences of low quality in which erroneous base calls introduced artificial stop codons and frameshifts (D. GARFIELD, personal observation). All sequence processing was carried out using custom scripts written in Python unless otherwise noted above.

Despite all the quality-control steps above, we obtained some unrealistically high estimates of K_s for many genes that almost certainly result from misalignment, misannotation, or a frameshift (possibly resulting from sequencing error) within a sequence. Hence, we excluded 10% of the 5599 predicted genes remaining in the data set, consisting of genes with the highest K_s values, from further analyses except where specified (this 10% is referred to later as "outlier loci"). We chose 10% arbitrarily, but we feel it is conservative given that several of the excluded loci had believable K_s values.

We assembled the unassembled contigs of *D. pseudoobscura* into a linear order for our analyses and determined the distance from inversion breakpoints to a particular locus (see S. W. SCHAEFFER, unpublished results). Calculations were made such that there was assumed to be no gap between the scaffolds, which is surely incorrect but does not introduce a specific bias into our analyses. For analysis, we used the PAML outputs of K_s and K_a and our own calculation of the percentage of difference between intron sequences of the two genomes (excluding gaps). Statistical analyses were performed using StatView (SAS Institute, Cary, NC) software. Median divergence values within bins of 1 Mb were plotted—this size was necessary to consistently have at least 9–10 analyzed loci per bin. The set of genes located inside the inverted regions was characterized using the web-based tool FuncAssociate (BERRIZ *et al.* 2003), which finds overrepresented classes of genes in a list of genes of interest using gene ontology (GO) attributes. Analyses were conducted jointly for all inverted regions and independently for each inverted region. Further, we analyzed the set of genes from our data set that were candidates for positively selected genes to determine if any class of genes was overrepresented in that list. Analyses were conducted with genes showing K_a/K_s values >1 or >0.7, as the latter value has been shown to be a sensible indicator of positive selection using more sensitive methods (SWANSON *et al.* 2004). FuncAssociate allows correcting for multiple hypothesis testing using Monte Carlo simulations (BERRIZ *et al.* 2003).

Intergenic noncoding sequence divergence was subsequently estimated from total genome alignments generated by Mercator (http://www.biostat.wisc.edu/~cdewey/mercator/) and MAVID (BRAY and PACHTER 2004). These alignments were engineered by Anat Caspi (University of California, Berkeley, CA). Mercator identifies orthologous exons using the output of a gene prediction program to define anchor points for the alignment of intervening noncoding regions. The alignments of the *D. pseudoobscura* and *D. persimilis* whole-genome sequences were obtained from the internet (http://www.biostat.wisc.edu/~cdewey/fly_CAF1/) and included both coding and noncoding sequences. The coordinates of coding sequences for *D. pseudoobscura* were extracted from a synpipe analysis (BHUTKAR *et al.* 2006). Two adjustments were made to the list of coding sequences. First, genes that mapped within the coordinates of other genes were removed. Second, genes whose 5′ and 3′ ends overlapped were fused into a single coding region. The coding coordinates were used to infer the boundaries of the intergenic noncoding regions that would be extracted from the Mercator alignment. Gaps and ambiguous sites were excluded from the analysis. Average divergence values within bins of 0.25 Mb were plotted to observe local patterns in noncoding divergence across the chromosome.

RESULTS

General trends: We examined patterns of sequence divergence between *D. pseudoobscura* and *D. persimilis* along the lengths of chromosomes XL, XR, 2, 3, and 4. The first three chromosomes all differ by a fixed or nearly fixed single inversion between the two species, chromosome 3 is highly variable for arrangements (although the two sequenced lines differ by a single inversion), and chromosome 4 is homosequential along its length. By comparing the gene order of these species to the distant outgroup species *D. melanogaster*, we confirmed the cytological interpretation that the arrangements on all major chromosomes in *D. persimilis* are derived (DOBZHANSKY and TAN 1936; supplemental Figure 1 at http://www.genetics.org/supplemental/).

A total of 81 Mb of intergenic noncoding sequence was compared between *D. pseudoobscura* and *D. persimilis*. Noncoding divergence varied across the genome from 0.011 on chromosome 5 to 0.033 on XL (Table 1). Figure 1 shows that there is a large amount of regional variation in divergence levels across the six chromosomes. Chromosome 5 has low levels of differentiation across the entire (small) chromosome. Among the five major chromosome arms, estimates of variation based on mean or median divergence per noncoding region showed a similar trend: low differentiation near the telomeres and centromeres and higher differentiation in the central regions (Figure 1 and Table 1). The 3-Mb regions nearest the centromeres and telomeres of all of the major chromosomes except XR show significantly lower levels of divergence than the middle sections of the chromosomes with a Mann–Whitney *U*-test [chromosome (Chr) XL, ends = 0.018, middle = 0.031, $P <$ 0.0001; Chr XR, ends = 0.015, middle = 0.016, $P =$ 0.499; Chr 2, ends = 0.012, middle = 0.021, $P <$ 0.0001; Chr 3, ends = 0.010, middle = 0.017, $P <$ 0.0001; Chr 4, ends = 0.011, middle = 0.018, $P <$ 0.0001].

On most chromosome arms, this same general pattern was also observed in K_s and intron divergence (Figure 2): median (or mean) divergence was lower near the ends (toward the centromere and telomere) than in the middle of the chromosome (3 Mb at both ends *vs.* the remainder: Mann–Whitney *U*-test, $P < 0.001$ for all major chromosomes). This pattern was least apparent in divergence measures on the telomeric side of the XL chromosome arm. Numbers of genes examined on each chromosome are presented in supplemental Table 1 (http://www.genetics.org/supplemental/) and supplemental Table 2 (http://www.genetics.org/supplemental/) presents all the raw data.

Multiple phenomena can explain this pattern of lower divergence near the chromosome ends, and we hypothesize that the effect of distance from the inverted region on introgression may be one contributing factor. The inversions differentiating these species are paracentric and located near the middle of most of the

TABLE 1

Estimates of divergence between *D. persimilis* and *D. pseudoobscura* in intergenic noncoding sequences on the six chromosome arms (Muller's elements)

Muller's element	XL (A)	XR (D)	2 (E)	3 (C)	4 (B)	5 (F)	Total
Total sites[a]	12,876,771	17,793,177	19,879,428	12,444,131	18,237,488	642,465	81,873,460
Divergent sites[b]	430,843	393,725	490,799	230,862	456,568	6,785	2,009,582
Total divergence[c]	0.033	0.022	0.025	0.019	0.025	0.011	0.025
Regions[d]	1,635	2,571	3,011	2,362	2,149	94	11,822
Median[e]	0.027	0.015	0.020	0.015	0.017	0.004	0.018

[a] Number of homologous sites compared excluding gaps and ambiguous bases.
[b] Number of bases that differed within the total sites.
[c] Divergent sites/total sites.
[d] Number of intergenic noncoding segments.
[e] Estimates of the median across the *n* intergenic regions.

chromosome arms, so the end regions with lower divergence are further from the inversion than central parts of the chromosomes. This pattern can thus result from introgression occurring more readily in the end regions than in regions closer to the inversions. However, contrary to this hypothesis being a sufficient explanation, the same pattern of lower differentiation near the chromosome ends was observed in chromosome 4, which does not differ by any inversion between the species. We speculate instead that this pattern is at least in part a remnant of the correlation between variation and local recombination rate observed within
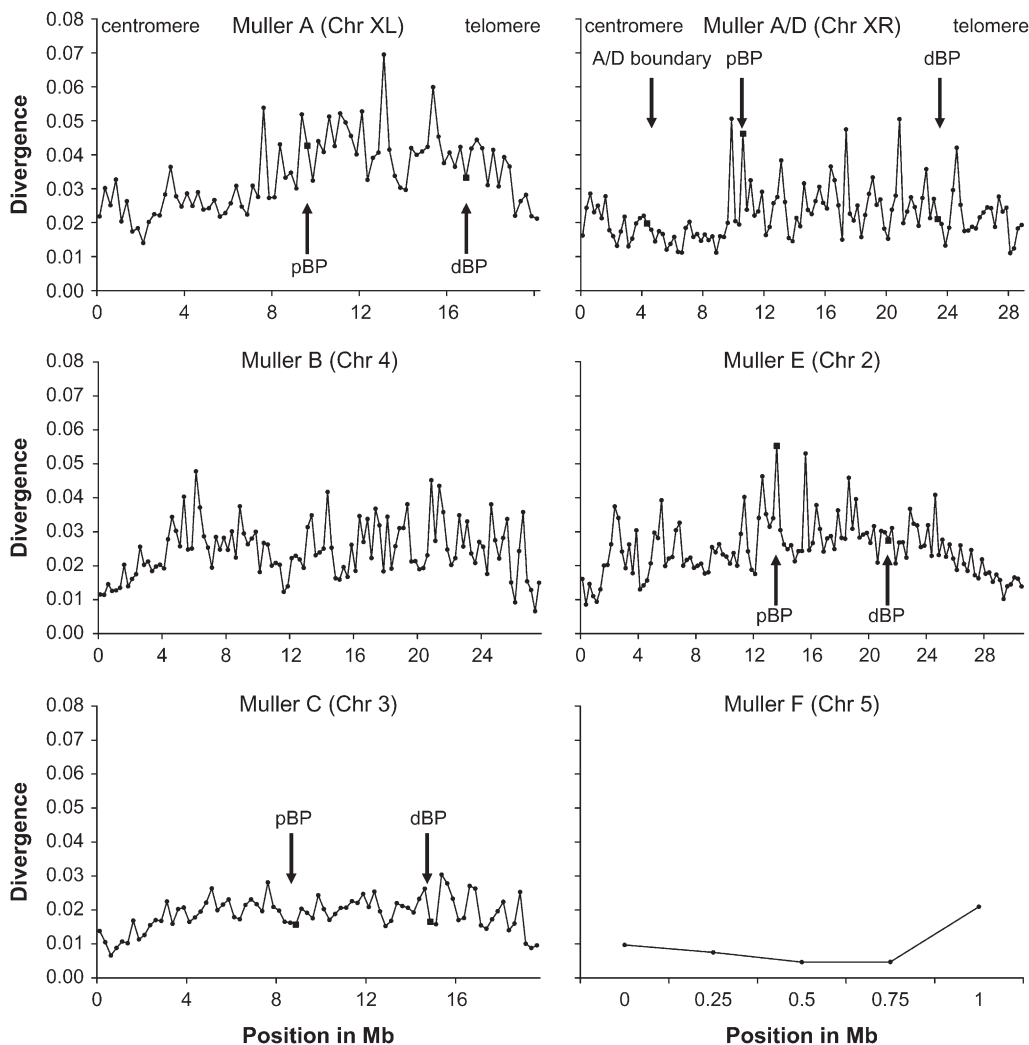

FIGURE 1.—Levels of intergenic noncoding divergence between *D. pseudoobscura* and *D. persimilis* within 0.25-Mb windows in the six Muller elements. The regions that contain breakpoint regions are indicated with a square point on the curve and an arrow.
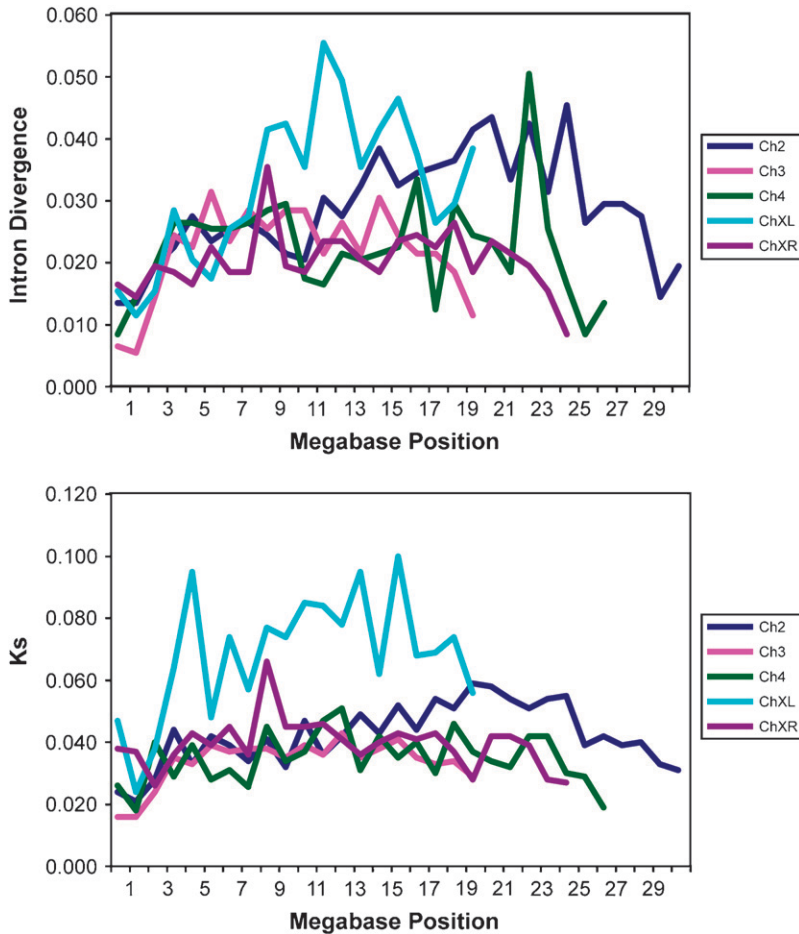
FIGURE 2.—Median values per megabase of divergence between *D. pseudoobscura* and *D. persimilis* genome sequences along the five major chromosome arms. They are plotted along the *x*-axis centromere to telomere (left to right). The top graph shows mean percentage of divergence between intron sequences, and the bottom graph shows median K_s values.

species (*e.g.*, BEGUN and AQUADRO 1992), and we discuss this interpretation in detail in the DISCUSSION.

This low divergence in regions of low recombination was also observed on the dot fifth chromosome. Virtually no crossing over occurs on this chromosome, and genes on it have a median K_s of 0.005 and median intron divergence of 0.000 (over half were identical) differentiating the two genome sequences, significantly lower than the other chromosome arms studied (individually or in combination) irrespective of whether outlier loci were excluded or not (excluded 10% outlier data, Mann–Whitney *U*-test, $P < 0.0001$; all data, Mann–Whitney *U*-test, $P < 0.0001$).

We also observed that the proximal region of chromosome XR has relatively low levels of intergenic noncoding divergence compared to the rest of the chromosome. The region of XR nearest the centromere includes genes that are on Muller A in the other sequenced Drosophila genomes (SEGARRA *et al.* 1995). The mechanism that led to the movement of genes from the A element to the D element is not clear at this time, but this genome rearrangement may have influenced the pattern of divergence.

Divergence and distance from inversion breakpoint: fixed inversion differences: We sought to examine the effect of distance from the inversion breakpoints on

divergence between the two genome sequences. Approximate positions of the breakpoints of the inversions in chromosomes XR, XL, and 2 are presented in MACHADO *et al.* (2007). In all three chromosome arms bearing fixed or nearly fixed inversion differences, the inverted regions exhibited greater divergence than the collinear regions (Table 2). Proximal and distal sides of the inversion did not differ to a notable degree. Given our observation of a centromeric/telomeric effect (see above), we excluded the 5 Mb on both ends of all the chromosomes from further analyses. This exclusion was conservative, as the effect did not seem to extend that far on chromosome 4 (Figures 1 and 2).

We then used Spearman's rank correlations to test the hypothesis that divergence decreases with increasing distance from the inversion breakpoint into collinear regions. On the second chromosome, the intergenic noncoding sequence divergence matched this prediction, but similar trends in K_s and intron divergence were not statistically significant (intergenic noncoding divergence, centromeric side, $\rho = -0.159$, $P < 0.0001$; telomeric side, $\rho = -0.131$, $P = 0.006$; K_s, centromeric side, $\rho = -0.071$, $P = 0.1870$; telomeric side, $\rho = -0.093$, $P = 0.2065$; intron divergence, centromeric side, $\rho = -0.059$, $P = 0.3221$; telomeric side, $\rho = -0.091$, $P = 0.2597$). These trends were also not significant when

<div style="display:flex">
<div>

TABLE 2

Estimates of intergenic noncoding divergence in proximal (centromeric side), inverted, and distal (telomeric side) regions in four chromosomes with inversion differences between *D. pseudoobscura* and *D. persimilis*

Muller's element	Proximal	Inverted	Distal
A (XL)			
Total sites	6,007,785	4,449,178	2,419,808
Divergent sites	164,947	187,203	78,693
Total divergence	0.027	0.042	0.033
C (3)			
Total sites	5,544,486	3,811,977	3,087668
Divergent sites	96,039	77,496	57,327
Total divergence	0.017	0.020	0.017
D (XR)			
Total sites	7,008,960	7,239,062	3,545,155
Divergent sites	131,709	186,451	75,565
Total divergence	0.019	0.026	0.019
E (2)			
Total sites	6,181,220	4,919,099	8,779,109
Divergent sites	138,541	145,458	206,800
Total divergence	0.022	0.030	0.022

</div>
<div>

TABLE 3

Median values for intergenic noncoding divergence, K_s, and intron divergence between three regions of chromosomes XL, XR, and 2

	IN	NEAR	FAR
Chr XL, intergenic div	0.036	0.029	0.023
Chr XR, intergenic div	0.017	0.015	NA
Chr 2, intergenic div	0.026	0.026	0.020
Chr 3, intergenic div	0.018	0.017	0.017
Chr XL, K_s	0.059	0.052	0.036
Chr XR, K_s	0.028	0.031	NA
Chr 2, K_s	0.042	0.039	0.033
Chr 3, K_s	0.031	0.033	0.034
Chr XL, intron div	0.042	0.028	0.017
Chr XR, intron div	0.017	0.017	NA
Chr 2, intron div	0.032	0.032	0.023
Chr 3, intron div	0.020	0.023	0.021

</div>
</div>

tested using mean or median values for stretches (of varying size) for multiple adjacent genes. On the XL chromosome arm, divergence was significantly related to distance from the inversion on the centromeric side (intergenic noncoding divergence, $\rho = -0.312$, $n = 354$, $P < 0.0001$; K_s, $\rho = -0.287$, $P = 0.0013$; intron divergence, $\rho = -0.331$, $P = 0.0012$). We did not examine the telomeric side of XL, as the inversion breakpoint appears to be within 5 Mb of the telomeric region. Intergenic noncoding sequence divergence had a significant negative correlation with breakpoint distance on the centromeric side of the XR chromosome arm, but no significant relationship was detected for other comparisons on this arm (intergenic noncoding divergence, centromeric side, $\rho = -121$, $P = 0.004$; telomeric side, $\rho = 0.064$, $P = 0.721$; K_s, centromeric side, $\rho = 0.100$, $P = 0.4797$; telomeric side, $\rho = 0.137$, $P = 0.5615$; intron divergence, centromeric side, $\rho = -0.135$, $P = 0.3811$; telomeric side, $\rho = -0.036$, $P = 0.8937$), although the range of distances examined was limited because of the large size of this inversion. None of these relationships changed if a parametric linear regression or logarithmic regression was performed (data not shown).

The regressions above assume a generally constant relationship between divergence and distance from the inversions. However, on the basis of MACHADO *et al.*'s (2007) findings, there may be three distinct types of regions: regions within the inversions (IN), regions outside the inversions but within 2 Mb of a breakpoint (NEAR), and regions >2 Mb outside the inversion (FAR). Hence, we examined whether these three regions differed from each other via a Kruskall–Wallis test within each chromosome arm. As the inversion on

the XR chromosome is very large, and because we excluded the 5 Mb near the centromere and telomere, there was virtually no "FAR" region left to test on this chromosome arm.

Within chromosomes XL and 2, we found significant heterogeneity in intergenic noncoding divergence (Chr XL, $H = 172.69$, $P < 0.0001$; Chr 2, $H = 138.32$, $P < 0.0001$), K_s (Chr XL, $H = 25.72$, $P < 0.0001$; Chr 2, $H = 28.45$, $P < 0.0001$), and intron divergence (Chr XL, $H = 43.92$, $P < 0.0001$; Chr 2, $H = 32.37$, $P < 0.0001$) among the three regions. For XR, there was significant heterogeneity among the regions in the intergenic noncoding divergence (Chr XR, $H = 65.60$, $P < 0.0001$); however, we failed to observe a difference between these regions in K_s or intron divergence (Table 3). Breaking these down into pairwise comparisons of groups within the chromosomes using the Mann–Whitney U-tests, we frequently observed highly significant differences between IN and FAR, significant differences between NEAR and FAR, and weaker or no differences between IN and NEAR (Tables 3 and 4).

If there has been gene flux between these species subsequent to the origins of these inversions, we predict that regions immediately adjacent to inversion breakpoints should be most differentiated (NAVARRO *et al.* 1997; SCHAEFFER and ANDERSON 2005). Consistent with this expectation, 11 of the 12 intergenic noncoding regions immediately adjacent to the mapped fixed inversion difference breakpoints have higher levels of divergence than the median divergence observed for the respective chromosome (Table 1; supplemental Table 3 at http://www.genetics.org/supplemental/). Four of these regions have some of the largest divergence values seen within their chromosome. We did not test for this pattern in K_s and intron divergence because there was often a very long stretch of noncoding DNA between the inversion breakpoint and the nearest predicted gene.

TABLE 4

Statistical significance (determined by Mann–Whitney U-test) of pairwise differences in divergence between regions of chromosomes

	IN *vs.* FAR	NEAR *vs.* FAR	IN *vs.* NEAR
Chr XL, intergenic div	<0.0001	<0.0001	<0.0001
Chr XR, intergenic div	NA	NA	0.013
Chr 2, intergenic div	<0.0001	<0.0001	0.684
Chr 3, intergenic div	0.064	0.066	0.712
Chr XL, K_s	<0.0001	0.0099	0.090
Chr XR, K_s	NA	NA	0.424
Chr 2, K_s	<0.0001	0.012	0.761
Chr 3, K_s	0.404	0.975	0.479
Chr XL, intron div	<0.0001	0.0010	0.0093
Chr XR, intron div	NA	NA	0.969
Chr 2, intron div	<0.0001	0.0001	0.320
Chr 3, intron div	0.826	0.184	0.117

Divergence and distance from inversion breakpoint: polymorphic inversion: We also examined whether divergence was greater far from the third chromosome inversion breakpoints than within the third chromosome inversion or near its breakpoints. In contrast to the results above, we found essentially no difference between IN, NEAR, and FAR on the third chromosome (Tables 3 and 4), reflecting a difference between inversions that are fixed between species and inversions that are polymorphic within species (with an abundant arrangement that is shared). For intergenic noncoding regions, chromosome 3 has the lowest level of noncoding divergence of the major chromosomal arms. Chromosome XR has a similar level of overall divergence, but the range of values in XR is much greater than that in chromosome 3. Two of the four noncoding regions immediately adjacent to the third chromosome breakpoints are greater than the median value for the chromosome, but these values are not significantly different from the average.

Differences in relative rates of nonsynonymous to synonymous change: We investigated whether the ratio of nonsynonymous to synonymous nucleotide divergence (K_a/K_s: also called "ω") exhibited any differences between the three regions IN, NEAR, and FAR. Within chromosomes XL, XR, and 3, we failed to observe significant heterogeneity in K_a/K_s (XL, $H = 3.8$, $P = 0.15$; XR, $U = 16,752$, $P = 0.59$; 3, $H = 1.0$, $P = 0.61$) among the regions. However, we did observe significant heterogeneity among the three regions of the second chromosome ($H = 10.6$, $P = 0.0050$).

Examining the variation among the three regions of the second chromosome by Mann–Whitney U-tests, we found significant differences between IN and NEAR ($P = 0.0026$) and between NEAR and FAR ($P = 0.0023$), but curiously not between IN and FAR ($P = 0.7543$). The highest K_a/K_s on this chromosome was observed in

region NEAR (supplemental Table 4 at http://www.genetics.org/supplemental/).

Differences between inverted regions among chromosomes: We examined measures of divergence between the fixed and nearly fixed inverted regions of the three chromosomes (XL, XR, 2) studied. In a model assuming contact between the species and some gene flow through the divergence process, this comparison could indicate the relative age of the three inversions. Kruskall–Wallis tests indicated significant heterogeneity ($P < 0.0001$) in both intergenic divergence, K_s, and intron divergence between the inverted regions of the three chromosomes, with chromosome XL being the most divergent and XR being the least (Tables 2 and 3). Each pairwise comparison also displayed a significant difference wherein XL > 2 > XR for these measures ($P < 0.0001$). This pattern and its statistical significance were robust to inclusion of the excluded outlier loci and to limiting the data set to loci at specific distances from the inversion breakpoints.

Gene ontology analyses: Finally, we looked for functional categories that may be overrepresented among the genes in the inverted regions and those likely evolving under positive selection. We applied these analyses to the full sets of genes, without excluding the outliers. Two GO attributes are significantly overrepresented in the list of 2631 genes from the three inversions, FAD binding/flavin adenine dinucleotide binding and choline dehydrogenase activity (supplemental Table 5 at http://www.genetics.org/supplemental/). However, the number of genes associated with these GO attributes is very small (16 and 7, respectively) and almost all the genes are organized in tandem in the same genomic region of the XL inversion.

As our analyses have shown (see above), significant differences in the average level of silent divergence among inverted regions suggest different times of origin for the three fixed inversions. For this reason, we conducted separate analyses for each inverted region. Each inverted region rendered a set of significantly overrepresented GO attributes (supplemental Table 5 at http://www.genetics.org/supplemental/). Five GO attributes are significantly overrepresented in the XL inverted region (supplemental Table 5). The two terms with the highest significance are the same GO terms that were significant in the joint analyses (see above), but this region also exhibits an overrepresentation of genes with oxidoreductase activity (see DISCUSSION). The second chromosome inverted region has two overrepresented GO terms. Interestingly, genes that are structural constituents of the cuticle are overrepresented in the XR inverted region. Previous studies have shown differences in cuticular hydrocarbons between these two species (NOOR and COYNE 1996).

Finally, no GO attributes are significantly overrepresented among the set of genes from our data set that were candidates for positively selected genes. Considering just those genes with $K_a/K_s > 1.0$ or $K_a/K_s > 0.7$ made no difference (data not shown).

DISCUSSION

We used three estimates of divergence between the genome sequence assemblies of *D. pseudoobscura* and *D. persimilis* to test a variety of hypotheses regarding a model of divergence with gene flow for the history of these two species. First, we observed a general pattern of low differentiation between the two genome sequences in pericentromeric and peritelomeric regions. This pattern was observed in all major chromosomes examined. We focused most of our analyses on the collinear regions of those three chromosome arms bearing inversions that differentiate the two species, excluding these pericentromeric and peritelomeric regions. As in the polymorphism/divergence study of MACHADO *et al.* (2007), we observed that loci far from the second chromosome inversion breakpoints were less divergent than loci within 2 Mb of the breakpoints, and loci close to the inversion breakpoints were essentially no less divergent than loci within the inverted regions. Since the present study used only single sequence representatives of the two species (each from an inbred isofemale line), our analyses are also not complicated by differences between the species in nucleotide variation resulting from the fixation of the novel inversions in *D. persimilis.* We further expand upon the result of this previous study by documenting the same relationships on the XL and XR chromosome arms, but failing to note this pattern on the polymorphic third chromosome. However, we did not observe a consistent difference in the ratio of nonsynonymous to synonymous divergence (K_a/K_s) between loci inside *vs.* outside the inverted regions. Finally, we observed that loci within the XL chromosome arm inversion were more divergent than those within the second chromosome inversion, which in turn were more divergent than loci within the XR chromosome arm inversion. We discuss these results in turn.

We interpret at least some of the lower differentiation between the *D. pseudoobscura* and *D. persimilis* telomeric and centromeric genome sequences as reflective of ancestral patterns of polymorphism rather than the process of divergence between these species. If two *D. pseudoobscura* genome sequences were compared, we expect to see lower "divergence" between them in low-recombination regions because polymorphism levels are typically lower in pericentromeric and peritelomeric regions, both in *D. pseudoobscura* and in general (*e.g.,* AQUADRO and BEGUN 1993; NACHMAN and CHURCHILL 1996; HAMBLIN and AQUADRO 1999; ORTIZ-BARRIENTOS *et al.* 2006). Similarly, outside of inverted regions, these two species share many nucleotide polymorphisms. There may have been insufficient time for fixed differences to accumulate in regions of low recombination to overcome the intraspecific correlation between exchange and variation. Given that many polymorphisms are shared between these species, the low divergence at the ends of the chromosome arms between the two

genome sequences may reflect this reduced shared polymorphism resulting from low crossover rates. We cannot rule out, however, that some of this effect may result from greater introgression in these regions.

This observation illustrates a potential problem with the use of single genome sequences from closely related species to examine the process of divergence or speciation. While one may be tempted to infer sequence differences as reflecting the species divergence, ancestral polymorphism patterns are ignored and can conflate the interpretations of patterns documented. As observed in studies of single or small numbers of genes, some *D. pseudoobscura* are more similar at some loci to *D. persimilis* individuals than to some other *D. pseudoobscura* individuals (MACHADO and HEY 2003; MACHADO *et al.* 2007), so intraspecific polymorphism must be considered.

We also observed generally lower levels of noncoding divergence on chromosomes 3 and 5 than on the others. This pattern also may result from the lower levels of recombination experienced by these chromosomes. Meiotic exchanges are rare for the homolog of chromosome 5 (element F) in *D. melanogaster* (QUESNEVILLE *et al.* 2005, p. 449). The lower level of divergence on chromosome 3 and the lack of differentiation near the inversion breakpoints are likely a consequence of the gene arrangement history for this chromosome. First, the wealth of gene arrangement polymorphism in populations of *D. pseudoobscura* and *D. persimilis* is likely to limit nucleotide diversity due to the reduced amount of recombination experienced by the third chromosome (DOBZHANSKY and EPLING 1944). Second, the arrowhead arrangement, the chromosome type that was sequenced in *D. pseudoobscura*, was recently derived from a standard arrangement, the chromosome type that was sequenced in *D. persimilis* (AQUADRO *et al.* 1991; S. W. SCHAEFFER, unpublished data). The divergence data between *D. pseudoobscura* and *D. persimilis* are consistent with a very recent origin of the two arrangements because there has been insufficient time for regions near the inversion breakpoints to differentiate (NAVARRO *et al.* 2000) and show a negative relationship between divergence and distance to the nearest breakpoint.

Role of chromosomal inversions in divergence of species: Chromosomal inversions and other means of restricting recombination have been suggested to be important in maintaining differentiation between hybridizing species (see reviews in ORTIZ-BARRIENTOS *et al.* 2002; AYALA and COLUZZI 2005; BUTLIN 2005). Several studies have shown that introgression between hybridizing species is reduced or absent in such regions of restricted recombination (*e.g.,* inside inversions), while other regions may show comparatively high levels of introgression or homogenization (RIESEBERG *et al.* 1999; NOOR *et al.* 2001b; FEDER *et al.* 2003; BROWN *et al.* 2004; PANITHANARAK *et al.* 2004; STUMP *et al.* 2005; BASSET *et al.* 2006). Hence, genomic regions such as inversions may represent islands of differentiation.

However, the "inside inversion" *vs.* "within collinear region" dichotomy may be overly simplistic for investigating introgression and divergence. Crossing over may be substantially reduced for several megabases from the inversion breakpoint into collinear regions, at least when assessed in a single-generation cross. Depending on the extent of crossover reduction, the level of hybridization, and the number of generations, this crossover reduction may cause the islands of differentiation to be substantially larger than the inverted region alone.

Our analysis of the two single whole-genome sequences of *D. pseudoobscura* and *D. persimilis* provided further evidence consistent with the hypothesis of islands of differentiation substantially larger than the inverted region on chromosome 2 (Machado *et al.* 2007). Our analysis also provided novel evidence for this hypothesis on the two arms of the X chromosome. Given the lack of a significant difference in intraspecific variability among these regions (after excluding the pericentromeric and peritelomeric regions: Hamblin and Aquadro 1999; Machado *et al.* 2007), the simplest explanation for our observation is that greater interspecies introgression has occurred between *D. pseudoobscura* and *D. persimilis* in collinear regions >2 Mb from the inversion breakpoints than close to or within the inverted regions. From the results reported, we cannot completely exclude the possibility that regions closer to (or within) inversion breakpoints have diverged faster because of differences in local mutation rate. However, this hypothesis would be inconsistent with results of detailed studies that have focused on handfuls of these loci (Machado *et al.* 2002, 2007; Machado and Hey 2003) and the lack of difference in intraspecific polymorphism across chromosome regions. We tested this hypothesis further by measuring K_s for coding sequence alignments of *D. melanogaster* to both *D. pseudoobscura* and *D. persimilis*. We examined 1706 genes on these three chromosome arms, and the Kruskall–Wallis test for IN *vs.* NEAR *vs.* FAR showed no significant difference among these regions within any chromosome (data not shown). Hence, there is no evidence that the differences between these regions in divergence is related to mutation rate differences.

In contrast, we failed to detect a relationship between divergence and inverted regions on chromosome 3. Unlike the chromosomes XL, XR, and 2, this chromosome has a very old and rich inversion polymorphism, and both the standard and arrowhead arrangements are found in *D. pseudoobscura*. Because these two arrangements are found within one of the two species, the opportunity for gene flux within (or near) inverted regions will necessarily be high, as intraspecific inversion heterozygotes will appear many orders of magnitude more often than interspecific inversion heterozygotes (hybrids). Gene flux between arrangements within *D. pseudoobscura* has been documented in a recent polymorphism study of seven gene regions on chromosome 3 (Schaeffer and Anderson 2005).

Navarro and Barton (2003) posited a model of chromosomal inversions facilitating differentiation of partially isolated populations wherein the inversions were fixed by positive selection. Following fixation, the different arrangements accumulate alleles that are adaptive for their population but incompatible or detrimental in the alternate population. Such alleles fail to accumulate outside of inversions because they would be eliminated by migration, recombination, and selection. Although other recombinational models of chromosomal speciation have been posited (Noor *et al.* 2001b; Rieseberg 2001), this model yields a unique prediction: the molecular signature of positive selection should be stronger in inverted regions than in collinear regions. We tested, and failed to find evidence, for the molecular signature of positive selection (defined by high K_a/K_s estimates in inverted regions) predicted by the Navarro and Barton (2003) model (supplemental Table 4 at http://www.genetics.org/supplemental/).

Given that the fixed chromosomal inversions have been important for the divergence of this species pair and harbor nearly all the reproductive isolation factors that could be genetically mapped (Noor *et al.* 2001a,b; Brown *et al.* 2004), we predicted that an analysis of Gene Ontology term representation would provide lists of gene classes that may have been important in the evolution of phenotypic differences between the species. The inversions harbor 2631 genes, ~19% of the genes in the genome, assuming a total of 14,000 genes. Although two GO attributes were overrepresented in the 2631 genes from the three inversions (supplemental Table 5 at http://www.genetics.org/supplemental/), we do not consider this a relevant result because the two GO terms are associated with very few genes in the genome (7–16) and almost all of these genes are located close to each other in the same genomic region of the XL inversion. However, when the analyses were conducted separately for each inverted region we observed additional overrepresented GO attributes (supplemental Table 5). Among those, two stand out. First, one of the overrepresented classes of genes in XL corresponds to genes with oxidoreductase activity, a result that matches recent results for this species pair using a microarray survey (C. Machado, unpublished data). Second, genes that are structural constituents of the cuticle are overrepresented in the XR inverted region, a result that is also significant given that previous studies have shown differences in cuticular hydrocarbons between these two species (Noor and Coyne 1996). Although these analyses are potentially useful for identifying candidate classes of genes important for phenotypic divergence that may have been "captured" by the fixed inversions, they depend on the available GO classifications. It will be thus important to repeat these analyses as GO classifications are refined.

Insights into divergence in this species group: It seems clear that chromosomal inversions have affected

levels of sequence divergence between *D. pseudoobscura* and *D. persimilis*. Further, we have shown here and elsewhere (Machado *et al.* 2007) that an effect associated with the inverted regions extends beyond the inversion breakpoints. The inverted regions of these species are associated with a variety of known barriers to gene flow between them, including hybrid male sterility, hybrid inviability, a hybrid courtship dysfunction, and behavioral discrimination (Noor *et al.* 2001a,b; Brown *et al.* 2004).

The sequence of events leading to the formation of these species (or their "genotypic clusters" *sensu* Mallet 1995) remains unknown, however. One possibility is that the progenitor of these species was split into two isolated (allopatric) populations, during which time the population that eventually became *D. persimilis* fixed three inversions that distinguish it from the other population. Meanwhile, both species evolved incompatibilities that prevent gene flow. Finally, the two incipient species came together, and while homogenization occurred in collinear regions, the inverted regions remained distinct and allowed for the persistence of the two species. This homogenization could have eliminated reproductive isolation factors that were present in collinear regions (Brown *et al.* 2004).

A second possibility is that some or much of the divergence between these populations occurred despite some gene flow between the two incipient species in sympatry. Local adaptation (possibly through preexisting clines) may have fixed the three inversions in what became *D. persimilis*, and alleles involved in incompatibilities between the two populations could then have accumulated within these inverted regions (Navarro and Barton 2003; Kirkpatrick and Barton 2006).

The first possibility, complete isolation followed by inversion formation and incompatibilities, predicts that the average divergence between the species in the three inverted regions should be fairly similar because it should match the time of the original population split. In contrast, the second possibility (divergence with gene flow) makes no such assumption, and divergence within the inversions should reflect the approximate time frame within which they arose. Our data here seem to support this latter model, as we observe significant differences between the XL, the XR, and the chromosome 2 inverted regions in divergence. In addition, the least divergent *D. pseudoobscura* XR arrangement is found in rare *D. persimilis* individuals exhibiting sex-chromosome meiotic drive, suggesting this inversion either is very old or may have spread between the species early in divergence. These *D. persimilis* gene arrangements may have been part of the east–west inversion cline transect observed in *D. pseudoobscura* (Dobzhansky and Epling 1944). That the three fixed inversions of *D. persimilis* are derived from the gene arrangements within *D. pseudoobscura* also supports the suggestion that *D. persimilis* emerged from within a chromosomally differentiated *D. pseudoobscura* population. Subsequent

(or continued) sympatry would permit exchange in collinear regions and possibly to a much more limited extent in some inverted regions (Schaeffer and Anderson 2005).

However, several caveats apply such that we cannot yet confidently accept the extreme divergence-with-gene-flow model. First, these comparisons assume a constant mutation rate and fixation probability, and local differences in these parameters are certainly plausible. Second, these comparisons assume that there is no gene flux between the inverted arrangements. Such gene flux has been detected between inversions in intraspecific chromosomal polymorphisms (Schaeffer and Anderson 2005), presumably through double crossovers or gene conversion. However, such flux between the fixed inversion differences separating these species will likely be reduced over time as incompatibility genes emerge and the low fitness of hybrids reduces the opportunities for exchange (Dobzhansky 1973; Powell 1983). In our data set, we failed to observe lower sequence divergence between central regions of inversions than between peripheral regions near breakpoints (data not shown), as would be predicted if there was gene flux via double crossover or gene conversion (Navarro *et al.* 1997; Schaeffer and Anderson 2005). It is possible that ancestral gene conversions are present in the data, resulting in some regions showing low levels of divergence, but we cannot tell without more extensive data. As a result of these complications, it is premature to accept a fully (or primarily) sympatric divergence model in these species.

Prospects: The accumulation of whole-genome sequence assemblies and ease of computational analysis will certainly be a boon for studies of speciation and the process of divergence. In this study, using two genome sequences, we examined patterns of divergence between a species pair that has been a classic system to study the process of speciation. We noted several caveats and areas where one could easily have been misled. For instance, although significantly lower divergence at some genomic regions could be interpreted to be the result of increased levels of introgression, lower polymorphism in those same regions can also affect the observed divergence between the single genome sequences compared. Disentangling the relative contribution of each factor (divergence *vs.* polymorphism) using single genome sequences is not possible. However, our analyses were highly informed by lower throughput but more rigorous examinations of polymorphism and divergence within the two focal species (Hamblin and Aquadro 1999; Machado *et al.* 2002, 2007; Machado and Hey 2003). The combination of impressive advances in computational analysis and the acquisition of genomic data with "old-school" reductionist benchwork can play a major role in informing genomic analysis of fundamental evolutionary questions, including understanding the origin of species (Noor and Feder 2006).

LITERATURE CITED

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

Aquadro, C. F., and D. J. Begun, 1993 Evidence for and implications of genetic hitchhiking in the *Drosophila* genome, pp. 159–178 in *Mechanisms of Molecular Evolution*. Sinauer Associates, Sunderland, MA.

Aquadro, C. F., A. L. Weaver, S. W. Schaeffer and W. W. Anderson, 1991 Molecular evolution of inversions in *Drosophila pseudoobscura*: the amylase gene region. Proc. Natl. Acad. Sci. USA **88:** 305–309.

Ayala, F. J., and M. Coluzzi, 2005 Chromosome speciation: humans, Drosophila, and mosquitoes. Proc. Natl. Acad. Sci. USA **102:** 6535–6542.

Basset, P., G. Yannic, H. Brunner and J. Hausser, 2006 Restricted gene flow at specific parts of the shrew genome in chromosomal hybrid zones. Evolution **60:** 1718–1730.

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356:** 519–520.

Berriz, G. F., O. D. King, B. Bryant, C. Sander and F. P. Roth, 2003 Characterizing gene sets with FuncAssociate. Bioinformatics **19:** 2502–2504.

Bhutkar, A., S. Russo, T. F. Smith and W. M. Gelbart, 2006 Techniques for multi-genome synteny analysis to overcome assembly limitations. Genome Inform. **17:** 152–161.

Blanchette, M., W. J. Kent, C. Riemer, L. Elnitski, A. F. Smit et al., 2004 Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. **14:** 708–715.

Bray, N., and L. Pachter, 2004 MAVID: constrained ancestral alignment of multiple sequences. Genome Res. **14:** 693–699.

Brown, K. M., L. M. Burk, L. M. Henagan and M. A. F. Noor, 2004 A test of the chromosomal rearrangement model of speciation in *Drosophila pseudoobscura*. Evolution **58:** 1856–1860.

Butlin, R. K., 2005 Recombination and speciation. Mol. Ecol. **14:** 2621–2635.

Dobzhansky, T., 1973 Is there gene exchange between *Drosophila pseudoobscura* and *Drosophila persimilis* in their natural habitats? Am. Nat. **107:** 312–314.

Dobzhansky, T., and C. Epling, 1944 *Contributions to the Genetics, Taxonomy, and Ecology of Drosophila pseudoobscura and Its Relatives*. Carnegie Institute, Washington, DC.

Dobzhansky, T., and A. H. Sturtevant, 1938 Inversions in the chromosomes of *Drosophila pseudoobscura*. Genetics **23:** 28–64.

Dobzhansky, T., and C. C. Tan, 1936 Studies of hybrid sterility. III. A comparison of the gene arrangement in two species, *Drosophila pseudoobscura* and *Drosophila miranda*. Zeit. Induk. Abst. Verer. **72:** 88–114.

Feder, J. L., J. B. Roethele, K. Filchak, J. Niedbalski and J. Romero-Severson, 2003 Evidence of inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. Genetics **163:** 939–953.

Gilbert, D. G., 2005 DroSpeGe, a public database of Drosophila species genomes. (http://insects.eugenes.org/DroSpeGe/).

Hamblin, M. T., and C. F. Aquadro, 1999 DNA sequence variation and the recombinational landscape in *Drosophila pseudoobscura*: a study of the second chromosome. Genetics **153:** 859–869.

Higgins, D. G., G. Blackshields and I. M. Wallace, 2005 Mind the gaps: progress in progressive alignment. Proc. Natl. Acad. Sci. USA **102:** 10411–10412.

Kirkpatrick, M., and N. Barton, 2006 Chromosome inversions, local adaptation and speciation. Genetics **173:** 419–434.

Lewontin, R. C., J. A. Moore, W. B. Provine and B. Wallace (Editors), 1981 *Dobzhansky's Genetics of Natural Populations I–XLIII*. Columbia University Press, New York.

Machado, C. A., and J. Hey, 2003 The causes of phylogenetic conflict in a classic Drosophila species group. Proc. R. Soc. Lond. Ser. B **270:** 1193–1202.

Machado, C. A., R. M. Kliman, J. A. Markert and J. Hey, 2002 Inferring the history of speciation from multilocus sequence data: the case of *Drosophila pseudoobscura* and its close relatives. Mol. Biol. Evol. **19:** 472–488.

Machado, C. A., T. S. Haselkorn and M. A. F. Noor, 2007 Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. Genetics **175:** 1289–1306.

Mallet, J., 1995 A species definition for the modern synthesis. Trends Ecol. Evol. **10:** 294–299.

Muller, H. J., 1940 Bearings of the Drosophila work on systematics, pp. 185–268 in *New Systematics*, edited by J. Huxley. Clarendon Press, Oxford.

Nachman, M. W., and G. A. Churchill, 1996 Heterogeneity in rates of recombination across the mouse genome. Genetics **142:** 537–548.

Navarro, A., and N. H. Barton, 2003 Accumulating postzygotic isolation in parapatry: a new twist on chromosomal speciation. Evolution **57:** 447–459.

Navarro, A., E. Betran, A. Barbadilla and A. Ruiz, 1997 Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. Genetics **146:** 695–709.

Navarro, A., A. Barbadilla and A. Ruiz, 2000 Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in Drosophila. Genetics **155:** 685–698.

Noor, M. A. F., and J. L. Feder, 2006 Speciation genetics: evolving approaches. Nat. Rev. Genet. **7:** 851–861.

Noor, M. A. F., and J. A. Coyne, 1996 Genetics of a difference in cuticular hydrocarbons between *Drosophila pseudoobscura* and *D. persimilis*. Genet. Res. **68:** 117–123.

Noor, M. A. F., K. L. Grams, L. A. Bertucci, Y. Almendarez, J. Reiland et al., 2001a The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. Evolution **55:** 512–521.

Noor, M. A. F., K. L. Grams, L. A. Bertucci and J. Reiland, 2001b Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. USA **98:** 12084–12088.

Ometto, L., D. DeLorenzo and W. Stephan, 2006 Contrasting patterns of sequence divergence and base composition between Drosophila introns and intergenic regions. Biol. Lett. **2:** 604–607.

Ortiz-Barrientos, D., J. Reiland, J. Hey and M. A. F. Noor, 2002 Recombination and the divergence of hybridizing species. Genetica **116:** 167–178.

Ortiz-Barrientos, D., A. S. Chang and M. A. F. Noor, 2006 A recombinational portrait of the *Drosophila pseudoobscura* genome. Genet. Res. **87:** 23–31.

Panithanarak, T., H. C. Hauffe, J. F. Dallas, A. Glover, R. G. Ward et al., 2004 Linkage-dependent gene flow in a house mouse chromosomal hybrid zone. Evolution **58:** 184–192.

Patterson, N., D. J. Richter, S. Gnerre, E. S. Lander and D. Reich, 2006 Genetic evidence for complex speciation of humans and chimpanzees. Nature **441:** 1103–1108.

Pollard, D. A., A. M. Moses, V. N. Iyer and M. B. Eisen, 2006 Detecting the limits of regulatory element conservation and divergence estimation using pairwise and multiple alignments. BMC Bioinformatics **7:** 376.

Powell, J. R., 1983 Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from *Drosophila*. Proc. Natl. Acad. Sci. USA **80:** 492–495.

Quesneville, H., C. M. Bergman, O. Andrieu, D. Autard, D. Nouaud et al., 2005 Combined evidence annotation of transposable elements in genome sequences. PLoS Comput. Biol. **1:** 166–175.

RICHARDS, S., Y. LIU, B. R. BETTENCOURT, P. HRADECKY, S. LETOVSKY *et al.*, 2005 Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution. Genome Res. **15:** 1–18.

RIESEBERG, L. H., 2001 Chromosomal rearrangements and speciation. Trends Ecol. Evol. **16:** 351–358.

RIESEBERG, L. H., J. WHITTON and K. GARDNER, 1999 Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics **152:** 713–727.

SCHAEFFER, S. W., and W. W. ANDERSON, 2005 Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. Genetics **171:** 1729–1739.

SCHAEFFER, S. W., M. P. GOETTING-MINESKY, M. KOVACEVIC, J. R. PEOPLES, J. L. GRAYBILL *et al.*, 2003 Evolutionary genomics of inversions in Drosophila pseudoobscura: evidence for epistasis. Proc. Natl. Acad. Sci. USA **100:** 8319–8324.

SEGARRA, C., E. R. LOZOVSKAYA, G. RIBO, M. AGUADE and D. L. HARTL, 1995 P1 clones from Drosophila melanogaster as markers to study the chromosomal evolution of Muller's A element in two species of the obscura group of Drosophila. Chromosoma **104:** 129–136.

STUMP, A. D., M. C. FITZPATRICK, N. F. LOBO, S. TRAORE, N. SAGNON *et al.*, 2005 Centromere-proximal differentiation and speciation in Anopheles gambiae. Proc. Natl. Acad. Sci. USA **102:** 15930–15935.

SWANSON, W. J., A. WONG, M. F. WOLFNER and C. F. AQUADRO, 2004 Evolutionary expressed sequence tag analysis of Drosophila female reproductive tracts identifies genes subjected to positive selection. Genetics **168:** 1457–1465.

YANG, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:** 555–556.

Communicating editor: L. HARSHMAN