Nonrecombinating Genes in a Recombination Environment: The Drosophila “Dot” Chromosome

Jeffrey R. Powell,*1 Kirstin Dion,1 Montserrat Papacete,2 Montserrat Aguadé,2 Saverio Vicario,3 and Ryan C. Garrick1

1Department of Ecology and Evolutionary Biology, Yale University
2Departamento de Genética, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain
3Università di Bari, CNR-ETB, Bari, Italy

*Corresponding author: E-mail: jeffrey.powell@yale.edu.
Associate editor: Hideki Innan

Abstract

Rate of recombination is a powerful variable affecting several aspects of molecular variation and evolution. A nonrecombinating portion of the genome of most Drosophila species, the “dot” chromosome or F element, exhibits very low levels of variation and unusual codon usage. One lineage of Drosophila, the willistoni/saltans groups, has the F element fused to a normally recombining E element. Here, we present polymorphism data for genes on the F element in two Drosophila willistoni and one D. insularis populations, genes previously studied in D. melanogaster. The D. willistoni populations were known to be very low in inversion polymorphism, thus minimizing the recombination suppression effect of inversions. We first confirmed, by in situ hybridization, that D. insularis has the same E + F fusion as D. willistoni, implying this was a monophyletic event. A clear gradient in codon usage exists along the willistoni F element, from the centromere distally to the fusion with E; estimates of recombination rates parallel this gradient and also indicate D. insularis has greater recombination than D. willistoni. In contrast to D. melanogaster, genes on the F element exhibit moderate levels of nucleotide polymorphism not distinguishable from two genes elsewhere in the genome. Although some linkage disequilibrium (LD) was detected between polymorphic sites within genes (generally <500 bp apart), no long-range LD between F element loci exists in the two willistoni group species. In general, the distribution of allele frequencies of F element genes display the typical pattern of expectations of neutral variation at equilibrium. These results are consistent with the hypothesis that recombination allows the accumulation of nucleotide variation as well as allows selection to act on synonymous codon usage. It is estimated that the fusion occurred ~20 Ma and while the F element in the willistoni lineage has evolved “normal” levels and patterns of nucleotide variation, equilibrium may not have been reached for codon usage.

Key words: recombination, molecular variation, Drosophila willistoni group, F element, codon usage.

Introduction

Since the seminal work of Kreitman (1983), numerous studies of DNA variation in Drosophila have revealed high levels of polymorphism (Moriyama and Powell 1996; Powell 1997; Begun et al. 2007). Generally, genes have nucleotide heterozygosity between 0.1% and 2% (proportion of nucleotide differences between two randomly drawn alleles) depending on function of the gene, exon, intron, flanking regions, etc. It was, therefore, notable when Berry et al. (1991) reported virtually no variation in Drosophila melanogaster and D. simulans in a gene (cubitus interruptus, ci) located on the fourth or “dot” chromosome, also known as the F element (nucleotide level is high due to differences between just three ~200 kb alleles each of which occurs in relatively high frequency. Further work on the melanogaster group species confirmed this low level of variation in F element genes and strong LD (Sheldahl et al. 2003; Wang et al. 2004; Arguello et al. 2010).

An exceptional pattern of codon usage has also been documented for genes on the F element compared with the rest of the Drosophila genome (Powell and Moriyama 1997; Vicario et al. 2007; Arguello et al. 2010). Codon usage in most of the genome exhibits strong bias indicative of selection for synonymous codon use, whereas genes on the F element appear to be under very little selection and codon usage reflects mutation bias. These observations are consistent with there being virtually no recombination in the F element when it is present as a very short dot. In classical mapping experiments, no recombination has been detected in this chromosome in D. melanogaster (Hochman 1976). However, results from DNA variation studies (Wang et al. 2002, 2004; Arguello et al. 2010) imply rare recombination estimated to be...
about 1/100 that of the rest of the Drosophila nuclear genome. Selective sweeps or background selection (Charlesworth et al. 1993) of nonrecombining DNA induces monomorphism and low recombination promotes LD. This makes selection at specific sites, such as among synonymous mutations (Hartl et al. 1994; Akashi 1995). Theoretical analyses of data from the dot chromosome of Drosophila indicate that the Hill–Robertson effect is very strong in influencing patterns and levels of nucleotide polymorphisms (Kaiser and Charlesworth 2008; Betancourt et al. 2009; Charlesworth et al. 2009).

Most species of Drosophila have a small, independently segregating, dot chromosome (F element); exceptions are the D. willistoni and D. saltans groups (Clayton and Wheeler 1975). In D. willistoni, the F element has been incorporated into another arm, element E, adjacent to the centromere (fig. 1). This has been confirmed by in situ gene hybridization to polytene chromosomes (Papaceit and Juan 1998). We confirm in this paper that this same fusion occurs in another member of the willistoni group (D. insularis) indicating the fusion very likely predates the radiation of the willistoni group and is a monophyletic event.

Here, we present data on the DNA polymorphism in some of the same F element genes studied in D. melanogaster in two species of the willistoni group, D. willistoni and D. insularis. We also examined one E element gene just distal to the fusion. In essence, we ask the question of how the molecular evolution of ancestrally nonrecombining (or rarely recombining) genes has changed when moved to a recombining chromosome. We examine codon usage, estimates of recombination, and levels and patterns of nucleotide polymorphism.

**Materials and Methods**

**Material**

D. willistoni and D. insularis were collected in St. Lucia in December 2006. D. willistoni was also collected in Pichilingue, Ecuador, in May 2008. Species identifications were confirmed by species-specific DNA sequences in the Alcohol dehydrogenase (Adh) locus (Griffith and Powell 1997). DNA extractions were made from flies directly from the field or a single $F_1$ of isofemale lines. Thus, the alleles sampled were directly from the natural populations.

**In Situ Hybridization**

Polytene chromosome preparations for in situ hybridization were performed according to Montgomery et al. (1987). Prehybridization, hybridization, and detection were as described in Segarra and Aguadé (1992). The location of the hybridization signals in both D. willistoni and D. insularis was determined using the Schaeffer et al. (2008) photographic map of D. willistoni, given the high chromosomal similarity between species. Probes from the eyeless (ey), cubitus interruptus (ci), and Ankyrin (Ank) genes were obtained through biotin-16-dUTP labeling by nick translation of either gel-purified polymerase chain reaction (PCR) amplicons of D. willistoni (ey and ci) or cDNA (Ank) of D. melanogaster (Papaceit and Juan 1998).

**DNA Isolation and Sequencing**

In addition to Adh sequences to confirm identification of the sibling species of the willistoni group (Griffith and Powell 1997), five genes on the E + F chromosome, cubitus interruptus (ci), Ankyrin (Ank), eyeless (ey), toy, and salvador (sav), were amplified and sequenced for the D. willistoni and D. insularis samples. Sav, on element E, lies just distal to element F in D. willistoni and D. insularis (fig. 1). Primers used to amplify the above genes are detailed in supplementary table S1, Supplementary Material online, and were designed using the Comparative Assembly Freeze 1 of the D. willistoni genome (FlyBase) and software program Primer 3 (Rozen and Skaletsky 2000). PCR amplification and sequencing details are in supplementary materials, Supplementary Material online.

**Data Analyses**

To analyze codon usage along the F element in D. melanogaster and D. willistoni, we used genome annotation from the FlyBase repository (www.flybase.org) version 5.22 and 1.3, respectively. We calculated the $F_{OP}$ statistics (the frequency of optimal codon) from the coding sequence of the longest transcript of each gene, using the optimal codon for each species identified in Vicario et al. (2007). Extraction of annotations from the gff files and construction of the codon usage table and $F_{OP}$ calculations were performed with Python script written by Saviero Vicario and using the parsing routine of BioPython (Cock et al. 2009). Trends of $F_{OP}$ statistics along the F element were analyzed with a linear model approach within R statistical environment (R Development Core Team 2009).

Phase of segregating sites in nuclear allele sequences obtained from diploid PCR products was estimated using PHASE v2.1.1 (Stephens et al. 2001; Stephens and Donnelly 2003), with files formatted in SeqPHASE (Flot 2010). The “MR” model was used because it accommodates the possibility of intragenic recombination. For triallelic singlenucleotide polymorphisms, the parent-independent mutation model was used. Search settings were: 500 iterations as burn-in, 1,000 main iterations, and thinning interval = 1. PHASE has very low false–positive rates even when
haplotype pairs have confidence probabilities as low as 0.60 and/or some coalescent assumptions are violated and systematic bias in population parameter estimates seems to occur only when unresolved (low confidence) genotypes are omitted from the data set (Garrick et al. 2010). Initially, PHASE was run to identify multisite heterozygotes with confidence scores < 0.95, and PCR products from those individuals were cloned using the pGEM-T Easy Vector System (Promega) and resulting colonies were screened for the target insert using primers SP6 and T7 (Promega). For each individual, two positive colonies were sequenced. Additionally, indels within the ey locus prevented a number of D. willistoni and D. insularis individuals from being directly sequenced. The above cloning procedure was therefore employed to rectify this issue. Here, however, five to ten positive colonies per individual were sequenced, thus ensuring the accurate identification of alleles despite inherent PCR error. In subsequent runs of PHASE, cloned alleles were treated as “known” haplotypes. Each data set was run three times with a different starting seed, and consistency across runs was checked by eye. The haplotype pair with the highest confidence probability score from the replicate run with the best average goodness-of-fit value was accepted and used for calculation of summary statistics and assessment of LD.

DNA sequence diversity summary statistics were calculated for exons and introns separately for each locus, species, and population. Fu’s (1997) F and Tajima’s (1989) D neutrality tests were performed on the same data partitions. Significance of F and D was assessed by comparing the observed value with a null distribution simulated via the coalescent process under constant population size (1,000 replicates). Following Fu (1997), F was considered significant at the 5% level only if  P < 0.02. Neutrality of exons was also assessed by calculating dN/dS (ω), the ratio of nonsynonymous to synonymous polymorphisms. Nonrandom associations between polymorphisms at segregating sites within a gene were examined via tests of LD including both exons and introns for each locus; significance of LD was assessed using Fisher’s exact test (two tailed) and Bonferroni correction for multiple comparisons. All calculations described in this paragraph were performed using DnaSP V5.10 (Librado and Rozas 2009).

Results

In Situ Mapping

Because only D. willistoni had been studied for position of F element genes (Papaceit and Juan 1998), we performed in situ hybridization of these genes in D. insularis. Supplementary figure S1, Supplementary Material online, has photographs of results for three F element genes, ci, ey, and Ank. The results confirm that the F element is in the same position in D. insularis, attached to element E adjacent to the centromere. In addition, the order of the F element genes are the same as in D. willistoni (fig. 1 and supplementary fig. S1, Supplementary Material online here and fig. 2 in Papaceit and Juan 1998), indicating that the attachment of the F and E elements was a single event in the ancestral lineage of the present day willistoni group.

Codon Usage

Figure 2 shows the pattern of codon usage along the lengths of elements E and F in D. melanogaster and D. willistoni based on the genome sequences (Drosophila 12 Genomes Consortium 2007). The frequency of the “optimal” codon (FOP) for each species (defined in Vicario et al. 2007) is along the length of the chromosomal elements measured in base pairs. As we have shown previously (Powell and Moriyama 1997; Vicario et al. 2007) in species that have this element as a small dot such as D. melanogaster, genes on the nonrecombining F element show reduced usage of optimal codons. In D. willistoni, the F element genes on the E + F chromosome exhibit FOP more like the E element genes in this species, although with a clear gradient from the centromere to the fusion point (fig. 2).

It is well known that the degree of usage of optimal codons for a species is directly proportional to recombination rate (Kliman and Hey 1993; Comeron et al. 1999; Comeron and Kreitman 2002; Hey and Kliman 2002). Furthermore, in Drosophila, it is well established that recombination rate falls off along arms of chromosomes toward both ends, the centromere and telomere. Table 1 presents the regressions of FOP and position along the chromosomes for approximately 2 Mb adjacent to the centromere and 2 Mb at the telomere (2 Mb is the approximate size of the F element euchromatin). As expected, there is a significant positive correlation going from the centromere out 2 Mb and a negative correlation in the last 2 Mb toward the telomere for D. melanogaster. For D. willistoni, there is a highly significant correlation between position of the F element genes and FOP, but unlike in D. melanogaster, the centromeric end of the E element in D. willistoni exhibits no significant gradient in FOP. There remains a significant negative gradient at the telomeric end of element E in D. willistoni comparable in magnitude with that in D. melanogaster.

These results confirm that codon usage in F element genes in D. willistoni has begun to evolve to higher usage of optimal codons and that there still exists a steep gradient from the centromere toward the fusion point, most likely due to a gradient in recombination. Because the centromeric end of the E element shifted with the fusion event, the formerly centromeric end of the E element has lost its gradient in codon usage in D. willistoni.

Recombination Estimations

We used PHASE to estimate recombination between the four genes studied in D. willistoni and D. insularis. Table 2 presents the numerical results based on detailed data in supplementary table S2, Supplementary Material online. These estimates are for c which is 4Nc r, so it is conceivable that the difference between species is due to a difference in Nc. Given the very restricted distribution of D. insularis (only known from two Caribbean islands) and the wide distribution of D. willistoni (throughout the Neotropics),
it is highly unlikely the difference between species in the estimates of $c$ are due to $D. insularis$ having a greater $N_e$ than $D. willistoni$, at least for contemporary population sizes. Second, although the confidence intervals on these estimates are large and overlapping, it is notable that in both species, the pattern is in the expected direction of a gradient of recombination from the centromere outward. This is consistent with the gradient in codon usage documented above.

Levels of Molecular Variation

Table 3 summarizes the levels of nucleotide variation found in $D. willistoni$ and $D. insularis$ (detailed data in supplementary material, Supplementary Material online). In both species, all genes studied on the $E + F$ chromosome (three or four in the $F$ element and one in the $E$ element adjacent to the fusion) exhibit a moderate amount of nucleotide polymorphism (with the exception of $ci$ in the St. Lucia $D. willistoni$ sample). These $\pi$ and $\theta$ estimates are about an order of magnitude greater than those found for $F$ element genes in $melanogaster$ species (Sheldahl et al. 2003; Wang et al. 2004; Arguello et al. 2010). $Adh$ on element $B$ in $D. willistoni$ and $D. insularis$ is comparable with these element $F$ genes in level of nucleotide variation with the exception of $Adh$ in the Ecuador population of $D. willistoni$. This locus in this population sample is a clear outlier exhibiting much higher exon variation than any other locus, 23 haplotypes in a sample of only 44 alleles ( supplementary table S2 , Supplementary Table 1. }

### Table 1. Regression Coefficients ($R$) between Frequency of Optimal Codon ($F_{OP}$) and Position in Base Pairs along Elements Starting from the Centromere. 2 Mb Is Used for Comparison as That Is the Approximate Length of the $F$ Element.

<table>
<thead>
<tr>
<th>Species</th>
<th>$F$ Element</th>
<th>First 2 Mb</th>
<th>Element $E$</th>
<th>Last 2 Mb</th>
<th>Element $E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>willistoni</td>
<td>+0.362***</td>
<td>+0.105 NS</td>
<td>−0.164*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>melanogaster</td>
<td>+0.022 NS</td>
<td>+0.155*</td>
<td>−0.187**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.**—NS, not significant, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. 

**Table 2.** Estimates from PHASE for Recombination between Genes. Figures are $\log_{10} 4N_e r$, where $r$ Is the Actual Estimated Recombination. Regions Are Listed from Centromere Toward the Distal Fusion Point.

<table>
<thead>
<tr>
<th>Species</th>
<th>Upper 95%</th>
<th>Ank-toy</th>
<th>toy-Sav</th>
</tr>
</thead>
<tbody>
<tr>
<td>inuslaris</td>
<td>−1.585</td>
<td>−1.404</td>
<td>−1.268</td>
</tr>
<tr>
<td>willistoni</td>
<td>−2.063</td>
<td>−1.844</td>
<td>−1.637</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Lower 95%</th>
<th>Ank-toy</th>
<th>toy-Sav</th>
</tr>
</thead>
<tbody>
<tr>
<td>inuslaris</td>
<td>−2.923</td>
<td>−2.285</td>
<td>−2.002</td>
</tr>
<tr>
<td>willistoni</td>
<td>−4.207</td>
<td>−1.919</td>
<td>−1.878</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Upper 95%</th>
<th>Ank-toy</th>
<th>toy-Sav</th>
</tr>
</thead>
<tbody>
<tr>
<td>willistoni</td>
<td>−6.018</td>
<td>−3.698</td>
<td>−2.611</td>
</tr>
<tr>
<td>Lower 95%</td>
<td>−7.995</td>
<td>−5.052</td>
<td>−3.846</td>
</tr>
</tbody>
</table>

**FIG. 2.** Frequency of optimal codons defined in Vicario et al. (2007) for protein-coding genes along the length of the $E$ and $F$ elements. In $Drosophila melanogaster$, these two elements exist as separate chromosomes (separated by vertical line) and are fused in $D. willistoni$ with the junction noted by the vertical line. The telomere is to the left for both the $E$ and $F$ elements of $D. melanogaster$ and $E + F$ of $D. willistoni$. Only about half the $E$ element is shown as assembly does not allow unambiguous continuation beyond this point for $D. willistoni$. 

4N_{e}r$
Table 3. Summary of Variation in Six Genes Studied in this Paper.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Species</th>
<th>Population</th>
<th>N</th>
<th>bps</th>
<th>Haps</th>
<th>S</th>
<th>Hd</th>
<th>θ</th>
<th>π</th>
<th>bps</th>
<th>Haps</th>
<th>S</th>
<th>Hd</th>
<th>θ</th>
<th>π</th>
</tr>
</thead>
<tbody>
<tr>
<td>ci</td>
<td>Drosophila insularis</td>
<td>STL 166</td>
<td>781</td>
<td>7</td>
<td>0.750</td>
<td>1.6</td>
<td>2.1</td>
<td>55</td>
<td>3</td>
<td>2</td>
<td>0.247</td>
<td>6.4</td>
<td>5.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D. willistoni</td>
<td>STL 148</td>
<td>781</td>
<td>3</td>
<td>0.440</td>
<td>0.5</td>
<td>0.8</td>
<td>55</td>
<td>3</td>
<td>2</td>
<td>0.027</td>
<td>6.5</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>STL 104</td>
<td>781</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>55</td>
<td>2</td>
<td>1</td>
<td>0.019</td>
<td>3.5</td>
<td>0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ECU 44</td>
<td>781</td>
<td>3</td>
<td>0.553</td>
<td>0.6</td>
<td>0.8</td>
<td>55</td>
<td>2</td>
<td>1</td>
<td>0.045</td>
<td>4.2</td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ank</td>
<td>D. insularis</td>
<td>STL 96</td>
<td>371</td>
<td>10</td>
<td>0.694</td>
<td>3.7</td>
<td>3.2</td>
<td>106</td>
<td>4</td>
<td>4</td>
<td>0.656</td>
<td>7.3</td>
<td>14.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D. willistoni</td>
<td>STL 134</td>
<td>371</td>
<td>5</td>
<td>0.059</td>
<td>2.0</td>
<td>2.0</td>
<td>106</td>
<td>2</td>
<td>1</td>
<td>0.382</td>
<td>2.3</td>
<td>4.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>STL 90</td>
<td>371</td>
<td>3</td>
<td>0.044</td>
<td>1.1</td>
<td>1.1</td>
<td>106</td>
<td>2</td>
<td>1</td>
<td>0.470</td>
<td>2.5</td>
<td>6.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ECU 44</td>
<td>371</td>
<td>4</td>
<td>0.133</td>
<td>1.9</td>
<td>0.5</td>
<td>106</td>
<td>2</td>
<td>1</td>
<td>0.045</td>
<td>2.9</td>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>toy</td>
<td>D. insularis</td>
<td>STL 104</td>
<td>590</td>
<td>8</td>
<td>0.647</td>
<td>2.6</td>
<td>2.2</td>
<td>258</td>
<td>11</td>
<td>8</td>
<td>0.728</td>
<td>6.4</td>
<td>6.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D. willistoni</td>
<td>STL 140</td>
<td>590</td>
<td>9</td>
<td>0.455</td>
<td>2.8</td>
<td>1.9</td>
<td>261</td>
<td>6</td>
<td>6</td>
<td>0.297</td>
<td>4.8</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>STL 96</td>
<td>590</td>
<td>7</td>
<td>0.506</td>
<td>2.6</td>
<td>2.0</td>
<td>261</td>
<td>3</td>
<td>2</td>
<td>0.120</td>
<td>1.7</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ECU 44</td>
<td>590</td>
<td>3</td>
<td>0.321</td>
<td>2.0</td>
<td>1.4</td>
<td>261</td>
<td>5</td>
<td>5</td>
<td>0.559</td>
<td>5.1</td>
<td>3.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ey</td>
<td>D. willistoni</td>
<td>STL 92</td>
<td>748</td>
<td>12</td>
<td>0.767</td>
<td>2.6</td>
<td>3.6</td>
<td>569</td>
<td>8</td>
<td>9</td>
<td>0.728</td>
<td>3.1</td>
<td>3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>STL 48</td>
<td>748</td>
<td>9</td>
<td>0.715</td>
<td>3.0</td>
<td>3.6</td>
<td>569</td>
<td>6</td>
<td>7</td>
<td>0.616</td>
<td>2.8</td>
<td>3.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ECU 44</td>
<td>748</td>
<td>6</td>
<td>0.707</td>
<td>1.8</td>
<td>2.5</td>
<td>569</td>
<td>7</td>
<td>8</td>
<td>0.789</td>
<td>3.2</td>
<td>4.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sav</td>
<td>D. insularis</td>
<td>STL 92</td>
<td>983</td>
<td>9</td>
<td>0.779</td>
<td>1.8</td>
<td>1.8</td>
<td>80</td>
<td>3</td>
<td>3</td>
<td>0.382</td>
<td>7.5</td>
<td>8.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D. willistoni</td>
<td>STL 78</td>
<td>983</td>
<td>9</td>
<td>0.728</td>
<td>2.1</td>
<td>1.3</td>
<td>187</td>
<td>4</td>
<td>3</td>
<td>0.124</td>
<td>3.3</td>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>STL 38</td>
<td>983</td>
<td>7</td>
<td>0.694</td>
<td>2.0</td>
<td>1.4</td>
<td>187</td>
<td>2</td>
<td>1</td>
<td>0.053</td>
<td>1.3</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ECU 40</td>
<td>983</td>
<td>4</td>
<td>0.731</td>
<td>1.0</td>
<td>1.1</td>
<td>187</td>
<td>3</td>
<td>2</td>
<td>0.188</td>
<td>2.5</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adh</td>
<td>D. insularis</td>
<td>STL 132</td>
<td>612</td>
<td>10</td>
<td>0.694</td>
<td>2.7</td>
<td>1.7</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D. willistoni</td>
<td>STL 144</td>
<td>612</td>
<td>28</td>
<td>0.797</td>
<td>8.3</td>
<td>3.9</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>STL 100</td>
<td>612</td>
<td>13</td>
<td>0.609</td>
<td>3.5</td>
<td>2.2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ECU 44</td>
<td>612</td>
<td>23</td>
<td>0.874</td>
<td>9.0</td>
<td>4.9</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note.—STL, St. Lucia, ECU, Ecuador; N, number of alleles sampled; bps, number of aligned base pairs; Haps, number of haplotypes; S, number of segregating sites; Hd, haplotype (gene) diversity; θ, Watterson’s estimate × 10^5; π = obs. nucleotide heterozygosity × 10^5.

Material online). dN/dS ratios do not indicate this excess in exon variation due to nonsynonymous variation (supplementary table S3, Supplementary Material online). There is no obvious explanation for this exceptional sample.

Excluding D. willistoni Adh in Ecuador, it appears that the genes on the F element in the willistoni lineage have equivalent levels of nucleotide variation as the two non-F element genes. Mean θ (× 10^5) for exons of F element genes is 1.9 (SD = 1.07), whereas Adh (excluding Ecuador) and Sav have mean θ of 2.2 (SD = 0.95). This level of variation in F element genes in D. willistoni and D. insularis is much higher than the exceptionally low variation found in these genes in species of the melanogaster subgroup (Berry et al. 1991; Hilton et al. 1994; Wang et al. 2002). (Note that we do not include data on the ey locus for D. insularis because it became clear that this gene is in multiple copies in this species and thus unambiguously determining orthology and paralogy is difficult. We simply exclude these data.)

It is also notable that for all loci studied, both on the F element and sav, D. insularis introns display considerably greater nucleotide diversity than that found in D. willistoni introns. This is consistent with the higher rate of recombination estimated for D. insularis compared with D. willistoni (table 2).

Pattern of Molecular Variation
The pattern of nucleotide variation is also very different in the fused E + F in D. willistoni and D. insularis compared with species where these elements are separate chromosomes. Wang et al. (2002) documented very strong LD along the length of the F element in D. melanogaster. In tests of LD, we found several cases of significant (after Bonferroni correction for multiple tests) associations within genes, sites that averaged 50–400 bps apart (supplementary table S4, Supplementary Material online). However, we found no significant LD between loci.

Examples of the distribution of allele frequencies are in figure 3 with similar figures for all loci in all populations in supplementary figure S2, Supplementary Material online. All loci exhibit the expected pattern of a most frequent allele followed by a few alleles in moderate frequency trailing off to singletons. The exception is ci in D. willistoni that exhibits very little variation. The patterns of distributions in figure 3 are the expected at equilibrium for neutral alleles. Tests of the conformance to these expectations are Tajima (1989) and Fu (1997), results of which are in supplementary table S5, Supplementary Material online. The Ank locus exons (but not introns) give some indication of nonconformance to neutral expectations as do toy introns. The exceptional variation in Adh in the Ecuador sample also yields tests at odds with neutrality. However, as to whether these tests indicate recent bottleneck or some other demographic perturbation is not at all clear; for example, the negative Ank F’s and D for exons are twice accompanied by positive values for introns in the same genes. Generally, the F element genes in the willistoni group species display levels of accumulated nucleotide variation and its distribution (pattern) along haplotypes is not distinct from the two genes on other elements in these species (fig. 3).
Discussion

Timing of the Fusion

The willistoni group is part of the willistoni–saltans lineage in subgenus Sophophora. All members of this lineage lack the dot chromosome (Clayton and Wheeler 1975). Here, we show that two members of the willistoni group have the fusion of elements E and F and the order of genes is identical. Because D. insularis and D. willistoni have the deepest branch within the sibling species of the willistoni group (Gleason et al. 1998; Robe et al. 2010), it is most parsimonious to assume the fusion was a monophyletic event in the ancestral lineage of this group. Given the absence of the dot in the saltans group, it is likely the fusion predated the split between the willistoni and saltans groups. These groups are estimated to have last shared a common ancestor ~20 Mya (Powell et al. 2003; O’Grady P, personal communication).

Codon Usage

As first documented by Kliman and Hey (1993) and since confirmed by a number of studies, codon usage bias is directly proportional to recombination rate, that is, genes in regions of higher recombination have higher codon usage bias than genes in regions of low recombination. This is likely due to the ineffectiveness of selection at single sites due to genetic hitchhiking (Hill and Robertson 1966). Our results are consistent with this. Because the F element became attached to the E element, codon usage has begun to evolve to that optimal for D. willistoni, although evidently the ~20 My since this fusion has not been long enough to bring the F element genes into equilibrium. The frequencies of optimal codons in the D. willistoni F element genes have not yet reached the pattern typical of ~2Mb adjacent to centromeres, for example, compare with element E in D. melanogaster (fig. 2). The slope of F element genes is much steeper in D. willistoni compared with either the E element telomeric region or centromeric regions in D. melanogaster.

That ~20 My is not long enough to bring about complete conformance of F element genes to optimal codon usage confirms that very small selection coefficients are associated with selection among synonymous codons (e.g., Hartl et al. 1994; Akashi 1995; Maside et al. 2004). The fact that codon usage has shifted for some amino acids in D. willistoni (Anderson et al. 1993; Powell et al. 2003; Vicario et al. 2007) suggests that the shift in codon usage predates

![Fig. 3. Examples of distribution of allele frequencies determined by concatenating exons and introns. One example is given for each of the six genes studied, three for each species. The counts are shown on the ordinate and haplotypes arbitrarily numbered along the abscissa. Details in supplementary material, Supplementary Material online including similar plots for all loci in all samples and haplotype sequences.](image-url)
the $E + F$ fusion, that is, the codon usage of the $F$ element genes have not risen to the level of the shift in codon usage in the rest of the genome. The split of the saltans/willistoni lineage from the obscura/melanogaster lineage in subgenus Sophophora is estimated to have occurred 30–40 Mya (Russo et al. 1995; Powell et al. 2003; O’Grady P. personal communication). Thus, there were 10–20 My between the origin of the saltans/willistoni lineage and the split between saltans and willistoni. Given that both these groups have the same pattern of codon usage shift (Rodriguez-Trelles et al. 2000; Powell et al. 2003), this shift must have begun during the 10–20 My when they shared a common ancestor. In other words, 30–40 My is long enough to evolve the unique codon usage pattern in genes in most of the genomes of the saltans/willistoni groups, but $\sim 20$ My is long enough to only partially achieve this shift in the $F$ element genes.

**Nucleotide Diversity**

Unlike codon usage, the amount and pattern of nucleotide diversity of $F$ element genes in *D. willistoni* and *D. insularis* appear to have converged on a normal pattern for these species. Very few genes have been studied for polymorphism, but of those that have, there is little indication of differences between $F$ element genes and the rest of the genome with regard to level of polymorphism. Similarly, except for linked sites within genes, there is no outstanding LD. There is no indication of LD between loci, unlike that found in *D. melanogaster* (Wang et al. 2002). For the most part, the distributions of allele frequencies (fig. 3 and supplementary material, Supplementary Material online) also indicate patterns consistent with expectations of accumulation of neutral variation at equilibrium.

**Recombination and Population History**

Our estimates of recombination (table 2) are consistent with the codon usage (fig. 2). There is a parallel gradient in estimated recombination rates with usage of optimal codons, although it is noted that these recombination estimates have large confidence intervals and this pattern is not statistically significant. Consistent with the higher estimated recombination rate in *D. insularis* compared with *D. willistoni*, the former species has higher levels of nucleotide diversity in introns compared with the latter species (table 3). What is not clear is why *D. insularis* has a higher recombination rate than *D. willistoni*, one to four orders of magnitude greater (table 2). This could be due to the high number of paracentric inversion known in *D. willistoni*. However, the populations of *D. willistoni* studied here, St. Lucia and Pichilingue, were chosen precisely because they were known to be low in inversion polymorphism (Dobzhansky 1957; da Cunha et al. 1959; Ayala et al. 1971) to minimize the effect of inversions suppressing recombination. Nevertheless, one inversion in the $E + F$ element of *D. willistoni* (the III-A) is very close to the fusion point but appears not to include the $F$ element (pictured in Dobzhansky et al. 1950; compared with fig. 10 in Schaeffer et al. 2008). This inversion is absent or rare in most populations of this species (including those studied in this report), although heterozygotes for the III “A or B” inversion does reach $\sim 50\%$ in some Peruvian and Brazilian populations (da Cunha et al. 1959; these authors only report the frequencies of inversion heterozygotes rather than gene arrangement frequencies). Nothing is known of naturally occurring inversions in *D. insularis*.

Because PHASE estimates recombination as the parameter $4N_r r$, where $r$ is the recombination value, it is possible that the difference between species for these estimates are due to differences in $N_r$. We argued above that it is very unlikely that *D. insularis* has a larger contemporary $N_r$ than *D. willistoni* given the narrow distribution of the former and wide distribution of the latter (Ehrman and Powell 1982). However, it is conceivable that ancestral $N_r$s are very different from those today. Comparisons of $\theta$ and $\pi$ (table 3) and the allele frequency plots (fig. 3 and supplementary fig. S2, Supplementary Material online) are inconsistent among the loci in regard to whether they indicate a bottleneck (or expanding population size). The statistical comparison of $\theta$ and $\pi$, Tajima’s test, should be negative in a recently bottlenecked expanding population, yet as noted above, the loci are ambiguous in this regard with some loci exhibiting significantly negative measures in exons and positive measures in introns of the same genes. Previous findings of surprisingly high nonsynonymous polymorphism in the Adh locus (Griffith and Powell 1997) led these authors to speculate *D. willistoni* may have recently experienced a bottleneck; the present data do not confirm an excess of nonsynonymous Adh polymorphisms in this species (supplementary table S3, Supplementary Material online).

One might expect to find $F$ element genes to show patterns of bottleneck and expansion given that the fusion was monophyletic, which would have meant a bottleneck of one when the fusion occurred. Such a bottleneck is no longer evident in the data, not surprising if the fusion took place at least 20 Mya; with five generations per year, $\sim 10^8$ generations have elapsed since the fusion.

**Codon Usage Bias Intensity in willistoni and melanogaster**

In figure 2, we presented the frequency of optimal codon usage along about half the length of the $E$ element for *D. melanogaster* and *D. willistoni*. It is evident that the scatter of data points for *D. melanogaster* is greater than that for *D. willistoni*. In fact, what this shows is that the lower bound for both species is about the same, $\sim 0.35$, but that the range above this lower bound is greater for *D. melanogaster*. That is, there are more genes with extreme bias in *D. melanogaster* than in *D. willistoni*. Because $F_{OP}$ is an average across all amino acids and there has been a change in optimal codon in *D. willistoni* for about a third of the amino acids (Powell et al. 2003; Vicario et al. 2007), the period of time that *D. willistoni* has been under selection for the shift in optimal codons for these amino acids may cause this parameter averaged across all amino acids not to reach extremes. Alternatively, it may indicate that the effective population size in the lineage leading to
**Conclusions and Caveats**

It is clear that the level and pattern of variation in genes on the F element in the willistoni group is much different from that found in these genes in species where they are on a (nearly) nonrecombining chromosome, the dot. Because the genes studied are the same as those studied in D. melanogaster, it is unlikely the differences are due to gene-specific constraints. Moving these genes into a recombining part of the genome has allowed accumulation of variation with no LD over distances above about 500 bp as well as allowed these genes to evolve codon usage closer to the optimal for the species. Have willistoni group F element genes come to an equilibrium? There is no indication that F element genes are less variable than genes in other parts of the genome in the willistoni group (table 3) or that the pattern of distribution of allele frequencies is different (fig. 3). It is less clear if codon usage in F element genes has reached an equilibrium. Given that this element is adjacent to the centromere in the E + F chromosome, a strong gradient in codon usage exists (fig. 2) that likely reflects a gradient in recombination. The steepness of this gradient is larger for the E + F chromosome than that in comparable regions indicating equilibrium may not yet have been reached for codon usage.

Although we estimated that the fusion of F and E elements likely occurred ≥20 Mya, it is conceivable that this estimate is off. It is based on the assumptions of a molecular clock (Powell et al. 2003; O’Grady P, personal communication). In addition, we are assuming the most parsimonious clock (Powell et al. 2003; O’Grady P, personal communication). In addition, we are assuming the most parsimonious clock (Powell et al. 2003; O’Grady P, personal communication). In addition, we are assuming the most parsimonious clock (Powell et al. 2003; O’Grady P, personal communication). In addition, we are assuming the most parsimonious clock (Powell et al. 2003; O’Grady P, personal communication). In addition, we are assuming the most parsimonious clock (Powell et al. 2003; O’Grady P, personal communication). In addition, we are assuming the most parsimonious clock (Powell et al. 2003; O’Grady P, personal communication). In addition, we are assuming the most parsimonious clock (Powell et al. 2003; O’Grady P, personal communication).

**Supplementary Material**

Supplementary tables S1–S5 and figures S1 and S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

For assistance in collecting Drosophila, we thank Alessandro Powell, Sam Genecin, and Gisella Caccone for the St. Lucia sample, and Andrea Acuario and Violeta Raphael for the Ecuador sample. Patrick O’Grady shared his findings with regard to timing of the splits of Drosophila lineages. Richard Kliman made useful comments on a draft of this manuscript. This work was supported by the US National Institutes of Health (RO1 GM0077533).

**References**


Dobzhansky Th. 1957. Genetics of natural populations. XXVI. Chromosomal variability in island and continental populations of Drosophila willistoni from Central America and the West Indies. Evolution 11:280–293.


