Local Changes in GC/AT Substitution Biases and in Crossover Frequencies on Drosophila Chromosomes

Toshiyuki Takano-Shimizu
Department of Population Genetics, National Institute of Genetics, Mishima, Shizuoka-ken, Japan

I present here evidence of remarkable local changes in GC/AT substitution biases and in crossover frequencies on Drosophila chromosomes. The substitution pattern at 10 loci in the telomeric region of the X chromosome was studied for four species of the Drosophila melanogaster species subgroup. Drosophila orena and Drosophila erecta are clearly the most closely related species pair (the erecta complex) among the four species studied; however, the overall data at the 10 loci revealed a clear dichotomy in the silent substitution patterns between the AT-biased-substitution melanogaster and erecta lineages and the GC-biased-substitution yakuba and orena lineages, suggesting two or more independent changes in GC/AT substitution biases. More importantly, the results indicated a between-loci heterogeneity in GC/AT substitution bias in this small region independently in the yakuba and orena lineages. Indeed, silent substitutions in the orena lineage were significantly biased toward G and C at the consecutive yellow, lethal of scute, and asense loci, but they were significantly biased toward A and T at sta. The substitution bias toward G and C was centered in different areas in yakuba (significantly biased at EG:165H7.3, EG:171D11.2, and suppressor of sable). The similar silent substitution patterns in coding and noncoding regions, furthermore, suggested mutational biases as a cause of the substitution biases. On the other hand, previous study reveals that Drosophila yakuba has about 20-fold higher crossover frequencies in the telomeric region of the X chromosome than does D. melanogaster; this study revealed that the total genetic map length of the yakuba X chromosome was only about 1.5 times as large as that of melanogaster and that the map length of the X-telomeric y-sta region did not differ between Drosophila yakuba and D. erecta. Taken together, the data strongly suggested that an approximately 20-fold reduction in the X-telomeric crossover frequencies occurred in the ancestral population of D. melanogaster after the melanogaster-yakuba divergence but before the melanogaster-simulans divergence.

Introduction

Eukaryotic chromosomes are organized into discrete high-order domains of several different levels and categories, such as nucleosomes, transcription units, replication units, and larger chromosomal structural segments. Insulator elements, for example, define the borders of transcription regulation units and thus protect a unit against both positive and negative influences of neighboring units (Kellum and Schedl 1991; Chung, Whiteley, and Felsenfeld 1993). On a larger scale, chromosomes of warm-blooded vertebrates, particularly those of humans, are divided into GC-rich and GC-poor segments of sizes varying from 200 kb to >1 Mb, the so-called isochore structure (Bernardi et al. 1985), but the functional significance of these large-scale domains remains unclear. In addition, although no molecular evidence has been provided so far, Drosophila chromosomes may be divided into independent recombinational domains (Hawley 1980). Chromosomal organization of these kinds may play an important role in shaping the rates and patterns of changes in DNA sequences and proteins.

Genome composition organization is assumed to be determined and maintained by the action of mutation and selection, but the degree to which these two factors have contributed has been the subject of much controversy, especially with regard to the isochore structure

Key words: crossover frequencies, mutation pressure, GC/AT substitution bias, Drosophila, telomeric region.

Address for correspondence and reprints: Toshiyuki Takano-Shimizu, Department of Population Genetics, National Institute of Genetics, Mishima, Shizuoka-ken 411-8540, Japan.
E-mail: totakano@lab.nig.ac.jp.

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Downloaded from https://academic.oup.com/mbe/article-abstract/18/4/606/980087 by guest on 23 December 2018
23E12.2 (1A) through stubarista (2B1–2), and previous studies have suggested a lack of structural change in this region in all four species (Lemeunier and Ashburner 1976; Takano-Shimizu 1999). The results indicated a lineage dependency and a locus dependency of substitution rates and GC/AT substitution biases. Because the substitution biases in the noncoding regions were generally the same as those at synonymous sites, this is very likely due to local changes in mutational bias and mutation rates.

I have also reported that the crossover frequencies in the X-telomeric regions were more than 10 times as high in D. yakuba as in D. melanogaster (Takano-Shimizu 1999). This study further clarified that the crossover frequencies in other regions of the X chromosome and an autosomal region in D. yakuba were, on average, only 1.5 times as high as those in D. melanogaster and that there was no significant difference in crossover frequencies in the X-telomeric region between D. yakuba and D. erecta, implying an X-telomeric-region-specific drastic reduction in the crossover frequencies in the melanogaster lineage.

These regional variations might suggest the presence of region-dependent regulation mechanisms of mutation and recombination in Drosophila. It is posited that this between-loci variability and these between-lineages fluctuations of mutation pressures and crossover frequencies affected evolutionary patterns of genes in the regions involved. Together with previous findings (Akashi 1996; Takano 1998), the present results suggest that changes in mutation, recombination, and effective population size all contribute to the significant locus-lineage interaction in the synonymous substitution rates among the Drosophila lineages (Takano 1998; Zeng et al. 1998), but changes in mutation parameters (mutational bias and total mutation rate) presumably have the greatest effects.

**Materials and Methods**

**DNA Sequences and Analyses**

I determined partial sequences of the following genes directly from single-male-fly products of the polymerase chain reaction (PCR): EG:23E12.2, cinnamon (cin), EG:165H7.3, EG:171D11.2, suppressor of sable (su(s)), suppressor of white-apricot (su(wa)), and stubarista (sta) of D. yakuba (stock number 14021-0261.0 from the National Drosophila Species Resource Center at Bowling Green, Ohio), of D. erecta (a stock was provided by N. Inomata), and of D. orena (a stock was provided by C. C. Laurie). A lethal of scute (l'sc) sequence of a single male fly of D. orena, an asense (ase) partial sequence of a single male fly of D. orena, and an ase partial sequence of a single male fly of D. erecta were also determined. The primer sequences are available on request. The sequences reported in this article appear in the DNA Data Bank of Japan and in the European Molecular Biology Laboratory and GenBank sequence databases with the accession numbers AB032437–AB032463.

**Fig. 1.—**Assumed phylogenetic relationships (cladogram) among the five Drosophila species. Ancestral nucleotides at each segregating site were assumed to be those which required the smallest number of total substitutions, as in Casane et al. (1997). A substitution from A to T in the Drosophila yakuba branch was inferred at site 1, an A→C or C→A substitution in the internal branch from node I to node II at site 2, A→T substitutions in the Drosophila orena branch and in the Drosophila melanogaster branch at site 4, and a G→A substitution in the D. orena branch and a G→C substitution in the D. yakuba branch at site 5. The ancestral states could not be uniquely inferred at sites 3 and 6, and such sites were excluded from the analysis. Drosophila simulans sequences were available only for the y, l’sc, and ase loci, and the melanogaster lineage in this study included an internal branch from node I to node III.

The melanogaster sequences used were AL031884 (EG:23E12.2), L19876 (cin), X52892 (ase), AL009188 (EG:165H7.3), AL009147 (EG:171D11.2), M57889 (su(s)), X06589 (su(wa)), and AL031027 (sta). The other sequences are described in Takano-Shimizu (1999) and Takano (1998). In addition to the melanogaster, yakuba, erecta, and orena sequences, I added the simulans and teissieri sequences (Takano-Shimizu 1999) in the analysis of y and the simulans sequences (Takano 1998) in the analyses of l’sc and ase. The sequences were aligned with CLUSTAL W (Thompson, Higgins, and Gibson 1994), and the alignments were modified by eye in some cases. Sites with gaps, as well as regions with alignment difficulty, were excluded from the analysis.

I estimated the number of substitutions along the melanogaster, yakuba, erecta, and orena lineages based on the following parsimonious assumption (see fig. 1). The assumed phylogenetic relationships among the five species including Drosophila simulans are depicted in figure 1. The nucleotides at the nodes were assumed to be those which required the smallest number of total substitutions. Those sites at which a node nucleotide could not be determined uniquely were excluded. This is the same method as that used in Casane et al. (1997), but I also included the segregating sites, such as sites 4 and 5 in figure 1, where two substitutions at a single site were required. There were 984 sites at which a single substitution was inferred in a terminal branch (site 1 in fig. 1), 295 sites that had a single substitution in the internal branch (site 2), 50 three-nucleotide segregating sites with an inference of two substitutions (site 5), and 10 two-nucleotide segregating sites with an inference of two substitutions (site 4, only for y, l’sc, and ase, for which D. simulans sequences are available). One hundred sites were excluded from the analysis be-
cause of an inability to infer the ancestral states (sites 3, 6, and others).

When two or more substitutions occurred in one codon, and then two or more possible substitution pathways that had different numbers of replacement substitutions were present, I calculated weight factors according to Miyata and Yasunaga (1980). When a favored pathway had a weight factor of ≥0.75, that pathway was uniquely taken and those substitutions were included in the analysis. When two or more pathways had roughly equal weights and I could not decide whether a difference was synonymous or replacement, the substitution was excluded. This study also revealed a nonsense mutation in exon 3 of the \textit{vermilion} gene and, in particular, was useful in rejecting a hypothesis of ~1.4-fold shorter generation time length of \textit{D. melanogaster} compared with \textit{D. yakuba}.

Measurements of Crossover Frequencies

Using three strains, I measured crossover frequencies of two different heterozygous females in four X-chromosomal regions and one autosomal region of \textit{D. yakuba: sans fille (snf)}–\textit{deltex (dx)}, \textit{forked (f)}–\textit{Shaker (Sh)}, \textit{Sh}–\textit{bangles and beads (bnb)}, \textit{Annexin X (AnnX)}–\textit{small optic lobes (so)} and \textit{lethal (2)} giant larvae \textit{(l(2)gl)}–\textit{decapentaplegic (dpp)}. The DNA samples used in this study were the same ones used in my previous study (Takano-Shimizu 1999). PCR-based molecular markers at 11 loci, including \textit{vermilion (v)} and \textit{rudimentary (r)}, are summarized in table 1. The \textit{snf} primers were designed from the published sequence (L29521) and the others from our sequence data (data not shown). Each of the three parental strains originated from a single-pair mating and was homozygous for the markers used. I determined genotypes of 12 randomly chosen F2 males from each of 16 bottles for the X loci and 15–19 males for the autosomal loci. Crossover frequencies were converted to map distances by the map function of Foss et al. (1993; with \(m = 4\)).
Table 2
Silent-site Sequence Divergence Between *Drosophila melanogaster* and *D. yakuba*, Between *D. erecta* and *D. orena*, and Between *D. melanogaster* and *D. orena*

<table>
<thead>
<tr>
<th>GENE</th>
<th>NONCODING</th>
<th></th>
<th>CODING</th>
<th></th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>n</em></td>
<td><em>K</em></td>
<td><em>n</em></td>
<td><em>K</em></td>
<td><em>n</em></td>
</tr>
<tr>
<td>EG:23E12.2</td>
<td>melanogaster-yakuba</td>
<td>366</td>
<td>0.128 ± 0.020</td>
<td>173.5</td>
<td>0.267 ± 0.050</td>
</tr>
<tr>
<td></td>
<td>erecta-orena</td>
<td>0.054 ± 0.012</td>
<td>0.106 ± 0.027</td>
<td>0.070 ± 0.012</td>
<td></td>
</tr>
<tr>
<td></td>
<td>melanogaster-orena</td>
<td>0.125 ± 0.020</td>
<td>0.232 ± 0.044</td>
<td>0.157 ± 0.019</td>
<td></td>
</tr>
<tr>
<td>cin (1A8)</td>
<td>melanogaster-yakuba</td>
<td>226</td>
<td>0.215 ± 0.035</td>
<td>251.3</td>
<td>0.261 ± 0.038</td>
</tr>
<tr>
<td></td>
<td>erecta-orena</td>
<td>0.056 ± 0.016</td>
<td>0.071 ± 0.018</td>
<td>0.064 ± 0.012</td>
<td></td>
</tr>
<tr>
<td></td>
<td>melanogaster-orena</td>
<td>0.175 ± 0.031</td>
<td>0.326 ± 0.045</td>
<td>0.250 ± 0.027</td>
<td></td>
</tr>
<tr>
<td>y (1B1)</td>
<td>melanogaster-yakuba</td>
<td>731</td>
<td>0.108 ± 0.013</td>
<td>345</td>
<td>0.218 ± 0.030</td>
</tr>
<tr>
<td></td>
<td>erecta-orena</td>
<td>0.039 ± 0.008</td>
<td>0.207 ± 0.029</td>
<td>0.088 ± 0.010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>melanogaster-orena</td>
<td>0.119 ± 0.014</td>
<td>0.316 ± 0.039</td>
<td>0.175 ± 0.014</td>
<td></td>
</tr>
<tr>
<td>l’sc (1B3)</td>
<td>melanogaster-yakuba</td>
<td>122</td>
<td>0.107 ± 0.032</td>
<td>194.1</td>
<td>0.281 ± 0.047</td>
</tr>
<tr>
<td></td>
<td>erecta-orena</td>
<td>0.042 ± 0.019</td>
<td>0.160 ± 0.033</td>
<td>0.111 ± 0.021</td>
<td></td>
</tr>
<tr>
<td></td>
<td>melanogaster-orena</td>
<td>0.126 ± 0.035</td>
<td>0.306 ± 0.051</td>
<td>0.230 ± 0.032</td>
<td></td>
</tr>
<tr>
<td>ase (1B4)</td>
<td>melanogaster-yakuba</td>
<td>289.8</td>
<td>0.273 ± 0.038</td>
<td>289.8</td>
<td>0.273 ± 0.038</td>
</tr>
<tr>
<td></td>
<td>erecta-orena</td>
<td>0.211 ± 0.033</td>
<td>0.211 ± 0.033</td>
<td>0.391 ± 0.052</td>
<td></td>
</tr>
<tr>
<td></td>
<td>melanogaster-orena</td>
<td>0.391 ± 0.052</td>
<td>0.391 ± 0.052</td>
<td>0.391 ± 0.052</td>
<td></td>
</tr>
<tr>
<td>EG:165H7.3</td>
<td>melanogaster-yakuba</td>
<td>402</td>
<td>0.087 ± 0.016</td>
<td>152.4</td>
<td>0.229 ± 0.047</td>
</tr>
<tr>
<td></td>
<td>erecta-orena</td>
<td>0.018 ± 0.007</td>
<td>0.080 ± 0.024</td>
<td>0.034 ± 0.008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>melanogaster-orena</td>
<td>0.046 ± 0.011</td>
<td>0.208 ± 0.043</td>
<td>0.087 ± 0.013</td>
<td></td>
</tr>
<tr>
<td>EG:171D11.2</td>
<td>melanogaster-yakuba</td>
<td>197</td>
<td>0.198 ± 0.036</td>
<td>92.4</td>
<td>0.264 ± 0.065</td>
</tr>
<tr>
<td></td>
<td>erecta-orena</td>
<td>0.081 ± 0.021</td>
<td>0.080 ± 0.031</td>
<td>0.080 ± 0.017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>melanogaster-orena</td>
<td>0.178 ± 0.033</td>
<td>0.249 ± 0.063</td>
<td>0.199 ± 0.030</td>
<td></td>
</tr>
<tr>
<td>su(s) (1B13)</td>
<td>melanogaster-yakuba</td>
<td>254</td>
<td>0.196 ± 0.032</td>
<td>140.3</td>
<td>0.278 ± 0.055</td>
</tr>
<tr>
<td></td>
<td>erecta-orena</td>
<td>0.032 ± 0.011</td>
<td>0.092 ± 0.027</td>
<td>0.053 ± 0.012</td>
<td></td>
</tr>
<tr>
<td></td>
<td>melanogaster-orena</td>
<td>0.153 ± 0.027</td>
<td>0.274 ± 0.055</td>
<td>0.193 ± 0.025</td>
<td></td>
</tr>
<tr>
<td>su(wa) (1E1-4)</td>
<td>melanogaster-yakuba</td>
<td>579</td>
<td>0.080 ± 0.012</td>
<td>16</td>
<td>0.216 ± 0.130</td>
</tr>
<tr>
<td></td>
<td>erecta-orena</td>
<td>0.037 ± 0.008</td>
<td>0.067 ± 0.069</td>
<td>0.038 ± 0.008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>melanogaster-orena</td>
<td>0.096 ± 0.014</td>
<td>0.137 ± 0.100</td>
<td>0.097 ± 0.013</td>
<td></td>
</tr>
<tr>
<td>sta (2B1-2)</td>
<td>melanogaster-yakuba</td>
<td>558</td>
<td>0.151 ± 0.018</td>
<td>195.9</td>
<td>0.124 ± 0.027</td>
</tr>
<tr>
<td></td>
<td>erecta-orena</td>
<td>0.060 ± 0.011</td>
<td>0.042 ± 0.015</td>
<td>0.055 ± 0.009</td>
<td></td>
</tr>
<tr>
<td></td>
<td>melanogaster-orena</td>
<td>0.135 ± 0.017</td>
<td>0.130 ± 0.028</td>
<td>0.134 ± 0.014</td>
<td></td>
</tr>
<tr>
<td>Mean^b</td>
<td>melanogaster-yakuba</td>
<td>0.141 ± 0.017</td>
<td>0.241 ± 0.015</td>
<td>0.182 ± 0.019</td>
<td></td>
</tr>
<tr>
<td></td>
<td>erecta-orena</td>
<td>0.047 ± 0.006</td>
<td>0.112 ± 0.019</td>
<td>0.081 ± 0.016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>melanogaster-orena</td>
<td>0.128 ± 0.014</td>
<td>0.257 ± 0.026</td>
<td>0.192 ± 0.028</td>
<td></td>
</tr>
</tbody>
</table>

^a Number of sites compared.
^b The unweighted arithmetic means and their standard errors of means were calculated from the estimates at 10 (or 9) loci.

Crossover frequencies in the y–*sta* region of *D. yakuba* and *D. erecta* were also compared. This experiment was done by using four isofemale strains for each species: four *yakuba* strains provided by C. C. Laurie (115, Elgan Lodge, David, and Coyne) and four *erecta* strains, one from the National Drosophila Species Resource Center at Bowling Green, Ohio (stock number 14021-0224.0), one from N. Inomata, and two (E220-5 and E1541) from C. C. Laurie. These four *yakuba* strains were different from the three strains used in the previous study (Takano-Shimizu 1999); thus, a more reliable estimate for within-species variation in crossover frequency was obtainable. The basic experimental design was the same as that in Takano-Shimizu (1999). Two pairs of strains were chosen to carry different alleles at both the y and the sta loci. The experiments were done in two sets, and one pair of *yakuba* strains and one pair of *erecta* ones were tested in each set. Reciprocal *G₀* crosses were made for each pair on the same day in each set, and the *F₁* heterozygous females were crossed separately with two other strains. *G₁* crosses were repeated twice on different days, yielding eight crosses for each parental pair. Crosses were done at 22°C.

**Results**

**Silent-Site Sequence Divergence**

Sequence divergences at 10 loci in the telomeric region of the X chromosome were surveyed for the four closely related Drosophila species *D. melanogaster, D. yakuba, D. erecta,* and *D. orena*. Most molecular data (Solignac, Monnerot, and Mounolou 1986; Cariou 1987; Lachaise et al. 1988; Jeffs, Holmes, and Ashburner 1994; Shibata and Yamazaki 1995) support the hypothesis that *D. erecta* and *D. orena* are a sister species pair (the *erecta* complex) and that *D. yakuba* is more closely related to *D. melanogaster* than the *erecta* complex (fig. 1). To show their close relationships, the numbers of per-site silent substitutions were estimated using Kimura’s (1980) two-parameter method, and those for three species pairs are given in table 2. “Silent” substitutions in this paper refer to synonymous substitutions in coding
Table 3
Total Numbers of Substitutions in 10 Genes on the Drosophila melanogaster, D. yakuba, D. erecta, and D. orena Lineages

<table>
<thead>
<tr>
<th></th>
<th>MELANOGASTER</th>
<th>YAKUBA</th>
<th>ERECTA</th>
<th>ORENA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A/T → G/C</td>
<td>A/T b</td>
<td>Others c</td>
<td>A/T → G/C</td>
</tr>
<tr>
<td>Noncoding</td>
<td>(3,431) a</td>
<td>57</td>
<td>105***</td>
<td>(220) e</td>
</tr>
<tr>
<td>Coding (6,937)</td>
<td>50</td>
<td>102***</td>
<td>(188) e</td>
<td>36</td>
</tr>
<tr>
<td>Synonymous</td>
<td>(1,850.7) a</td>
<td>107</td>
<td>207***</td>
<td>(408) e</td>
</tr>
<tr>
<td>Silent sum</td>
<td>16</td>
<td>28</td>
<td>(71) e</td>
<td>27</td>
</tr>
<tr>
<td>Replacement</td>
<td>5086.3 a</td>
<td>57</td>
<td>105***</td>
<td>(220) e</td>
</tr>
</tbody>
</table>

NOTE.—The results of a G-test with Williams’ correction for the numbers of two classes of substitutions, A/T → G/C and G/C → A/T, in the noncoding substitutions, in the synonymous ones, and in their sums are given. The substitution classes with significantly larger numbers are marked by asterisks: * P < 0.05; ** P < 0.005; *** P < 0.001.

a A or T to G or C substitutions.
b G or C to A or T substitutions.
c G → C and A → T substitutions.
d Total number of sites compared.
e Total number of substitutions on each branch.

regions and all substitutions in noncoding regions. The average number of silent substitutions per site was 0.2 even for the most distantly related species pair, D. melanogaster and D. orena. These degrees of sequence divergence are expected to allow reliable inference of the substitution patterns and to give reasonable numbers of total substitutions.

Sequence divergences were studied in both coding and noncoding regions at all 10 loci except for ase. Surprisingly, although most noncoding regions studied were introns, silent-site sequence divergences were higher in coding regions than in noncoding regions (table 2). Bauer and Aquadro (1997) have also found the same tendency in sequence divergence between D. melanogaster and D. simulans. However, the current result appears to be at least partly attributable to the elimination of alignment-ambiguous sites, which included high-sequence-divergence regions. When such sites were included individually for each pair of species in the calculation (allowing different numbers of total sites for different pairs), the silent-site sequence divergence in the EG:165H7.3 noncoding regions, for instance, was 0.253 ± 0.023 between D. melanogaster and D. yakuba and 0.069 ± 0.011 between D. erecta and D. orena. These estimates are very comparable with the synonymous-site sequence divergences in the EG:165H7.3 coding regions (0.229 ± 0.047 for the former species pair and 0.080 ± 0.024 for the latter). Thus, the present data did not necessarily imply a significant difference in silent substitution rates between the coding and the noncoding regions. Careful examination will be necessary to settle this issue.

GC/AT Silent-Substitution Biases

The total numbers of silent substitutions estimated in the 10 genes are given in table 3. A clear dichotomy in the silent substitution patterns existed between the AT-biased-substitution melanogaster and erecta lineages and the GC-biased-substitution yakuba and orena lineages. The numbers of A/T → G/C (A or T to G or C) and G/C → A/T (G or C to A or T) substitutions significantly deviated from the expectation of a 1:1 ratio in all four cases. Clearly, this classification is not consistent with the actual phylogenetic relationships of these four species (fig. 1). Consequently, the results implied two or more independent changes in GC/AT silent-substitution biases.

Although the individual tests for the coding and noncoding silent substitutions on the erecta lineage did not find a significant deviation from the 1:1 ratio for the numbers of A/T→G/C and G/C→A/T, both the noncoding regions and the coding regions manifested the same GC/AT silent substitution patterns on each of the melanogaster, yakuba, and erecta lineages (G’ = 0.18, P > 0.5 for 57 A/T→G/C and 105 G/C→A/T vs. 50 and 102 in melanogaster; G’ = 0.29, P > 0.5 for 72 and 39 vs. 70 and 44 in yakuba; G’ = 0.01, P > 0.9 for 16 and 26 vs. 16 and 27 in erecta). These results suggest mutational bias as a cause of biased substitutions in these lineages. On the other hand, the numbers of A/T→G/C and G/C→A/T silent substitutions on the orena lineage gave a statistically significant heterogeneity between the noncoding and the coding regions (G’ = 11.7, P < 0.001), and I discuss its possible cause below.

The numbers of substitutions at each locus are given in table 4 and are graphically presented in figure 2. I tested the numbers of A/T→G/C and G/C→A/T silent substitutions on each lineage for goodness of fit to the 1:1 ratio individually at each locus, applying the sequential Bonferroni technique (Rice 1989). For the yakuba lineage, significant substitution biases toward G and
C were found at \(EG:165H7.3, 171D11.2\), and \(su(s)\) (\(G' = 8.5, P = 0.0035\) for \(EG:165H7.3\); \(G' = 7.5, P = 0.0062\) for \(171D11.2\); \(G' = 22.8, P < 0.001\) for \(su(s)\)). A lack of significant heterogeneity in the two classes of silent substitutions between noncoding and coding regions (\(P\) from Fisher's exact test > 0.3 for 11:4 vs. 8:1 at \(EG:165H7.3\); \(P\) from Fisher’s exact test > 0.5 for 10:2 vs. 4:1 at \(EG:171D11.2\); \(P\) from Fisher’s exact test > 0.5 for 15:1 vs. 10:1 at \(su(s)\)) suggests again that the substitution biases are attributed to mutational biases.

The \(orena\) lineage also exhibited a locus-dependent \(A/T\) to \(G/C\) substitution bias, where a significant bias toward \(G\) and \(C\) was found at the \(y\), \(l'sc\), and \(ase\) loci (\(G' = 55.0, P < 0.001\) for \(y\); \(P\) from Fisher’s exact test < 0.001 for \(l'sc\); and \(G' = 32.3, P < 0.001\) for \(ase\)); however, the number of \(G/C\) to \(A/T\) substitutions was significantly larger than that of \(A/T\) to \(G/C\) substitutions at \(sta\) (\(G' = 12.1, P < 0.001\)). As mentioned above, the total numbers of \(A/T\) to \(G/C\) and \(G/C\) to \(A/T\) silent substitutions on the \(orena\) lineage gave a statistically significant heterogeneity between the noncoding and the coding regions. This finding, however, did not suggest the action of natural selection. The present data lack sufficient noncoding data at the two significantly GC-biased \(l'sc\) and \(ase\) loci: only 122 noncoding sites were compared at \(l'sc\), and none were compared at \(ase\). This deficiency of noncoding data at these loci appears to be a cause of the significant noncoding-vs.-coding heterogeneity. Indeed, when the \(l'sc\) and \(ase\) loci were excluded from the analysis, the data could not reject the independence of coding and noncoding regions (\(G' = 3.4, P > 0.05\)). Moreover, no significant heterogeneity between noncoding and coding regions existed at the remaining two loci of significantly biased substitution (\(P\) from Fisher’s exact test = 0.23 for 16:2 vs. 38:1 at \(y\); \(P\) from Fisher’s exact test = 0.43 for 2:10 vs. 0:6 at \(sta\)). This suggests that the substitution biases in this lineage can also be attributed to mutational biases.

As illustrated in figure 2, the substitution bias toward \(G\) and \(C\) was centered in different areas in \(yakuba\) (\(EG:165H7.3, 171D11.2\), and \(su(s)\)) and in \(orena\) (\(y\), \(l'sc\), and \(ase\)); as mentioned below, \(D. melanogaster\) and \(D. erecta\) did not show a GC-biased-substitution pattern at all. These results, combined with the phylogenetic relationships of the four species (fig. 1), strongly suggest at least two between-lineages changes in GC/AT substitution biases in the X-telomeric region. Because \(Drosophila\) \(pseudoobscura\) (an outgroup species for the \(m-\)
the following three loci: $G' = 7.6, P = 0.0059$ at $cin$; $G' = 11.6, P = 0.0007$ at $l'sc$; $G' = 10.4, P = 0.0013$ at $su(w^c)$; however, the degree of bias varied among genes. For instance, the ratio of the number of $G/C \rightarrow A/T$ substitutions to that of $A/T \rightarrow G/C$ substitutions in the noncoding regions at $su(s)$ and $su(w^c)$ loci was significantly larger than that at the other seven loci ($G' = 4.4, P < 0.05$ for $32 G/C \rightarrow A/T$ and $9 A/T \rightarrow G/C$ vs. $73$ and $48$). As for $D. yakuba$ and $D. orena$, this result may suggest locus-dependent GC/AT substitution biases in $D. melanogaster$.

Although the pooled data revealed significantly AT-biased silent-site substitutions on the $erecta$ lineage (table 3), the same tests for the numbers of two classes of substitutions on this lineage did not give a significant deviation from the 1:1 ratio at all 10 loci.

In sum, the present data clearly indicate that the GC/AT substitution bias is not only lineage-dependent, but also locus-dependent.

GC/AT Replacement-Substitution Biases

The pooled numbers of $A/T \rightarrow G/C, G/C \rightarrow A/T$, and other replacement substitutions are also given in table 3. Only the $orena$ lineage showed significantly GC-biased substitutions, in agreement with the silent-site substitutions. However, the GC/AT replacement substitution biases seemed not to be fully consistent with the biases at silent sites. The $yakuba$ lineage showed the GC-biased silent substitutions, but the number of $G/C \rightarrow A/T$ replacement substitutions in this lineage was, at face value, larger than that of $A/T \rightarrow G/C$ replacement substitutions. Indeed, a significant heterogeneity in the numbers of $A/T \rightarrow G/C$ and $G/C \rightarrow A/T$ silent substitutions existed between silent and replacement substitutions ($G' = 6.7, P < 0.01$ for $142 A/T \rightarrow G/C$ and $83 G/C \rightarrow A/T$ silent substitutions vs. $8$ and $15$ replacement substitutions). This ratio of $A/T \rightarrow G/C$ and $G/C \rightarrow A/T$ replacement substitutions on the $yakuba$ lineage was very similar to that on the $melanogaster$ lineage ($16 A/T \rightarrow G/C$ and $28 G/C \rightarrow A/T$ replacement substitutions). The GC/AT substitution biases on the $erecta$ lineage also differed between silent and replacement substitutions ($G' = 4.1, P < 0.05$). Selection on replacement substitutions might strongly oppose the GC/AT mutational bias, and differential selection on replacement and silent substitutions might have played a role in shaping the substitution rates and patterns. In this case, the higher replacement substitution rate in the $melanogaster$ lineage, as mentioned below, may be due to a relaxation of purifying selection in addition to a higher total mutation rate. However, this conclusion should be treated with caution, because the number of replacement sites compared varied among the loci, and there might be a between-loci variation in GC/AT substitution bias at replacement sites, as in the case of silent sites. In fact, the lack of noncoding data at the $l'sc$ and $ase$ loci that showed the strongly GC-biased synonymous substitutions resulted in the significant heterogeneity in the total numbers of $A/T \rightarrow G/C$ and $G/C \rightarrow A/T$ silent substitutions between the coding and the noncoding regions. The small num-

lanogaster species subgroup) has about the same A+T content of introns (61% A+T) as $D. melanogaster$ (60%) and $D. simulans$ (61%) (Akashi and Schaeffer 1997), locus-dependent GC-biased substitutions appear to have occurred independently in the $yakuba$ and $orena$ lineages.

In contrast to $yakuba$ and $orena$, the number of $G/C \rightarrow A/T$ substitutions generally outnumbered that of $A/T \rightarrow G/C$ substitutions on the $melanogaster$ lineage (Akashi 1996; a significant bias was observed individually at
bers of replacement substitutions at each locus did not allow further investigation on this issue.

Comparison of melanogaster vs. yakuba Lineage Substitution Rates

The melanogaster lineage accumulated a significantly larger number of replacement substitutions than the yakuba lineage (χ² in Tajima’s [1993] 1D test = 10.0, P < 0.005 for 71 vs. 38). The number of substitutions in noncoding regions and that of synonymous substitutions were also significantly larger in the melanogaster lineage than in the yakuba lineage (χ² in Tajima’s 1D test = 12.1, P < 0.001 for 220 vs. 153 noncoding substitutions; χ² = 8.7, P < 0.005 for 188 vs. 135 synonymous substitutions). The higher protein sequence evolutionary rate in the melanogaster lineage may be attributable to higher total mutation rates in the melanogaster lineage than in the yakuba lineage or, as mentioned below, to the much lower crossover frequencies in the X-telomeric region in melanogaster than in D. yakuba. The latter hypothesis is based on the theoretical finding that a decrease in recombination frequencies results in reduced efficacy of natural selection, followed by accelerated substitution rates of slightly deleterious mutations (Birky and Walsh 1988; Charlesworth 1994; McVean and Charlesworth 1999, 2000; Stephan, Charlesworth, and McVean 1999).

On the other hand, it is unlikely that a difference in generation time lengths is responsible for these substitution rate differences. The average egg-to-adult developmental period of D. yakuba (11.2 ± 0.1 days for females and 11.4 ± 0.1 days for males) was even shorter than that of D. melanogaster (13.4 ± 0.1 days for females and 13.6 ± 0.1 days for males) at 22°C (see Materials and Methods). Because developmental time depends on temperature, this comparison is not guaranteed to reflect a precise between-species difference in the wild. Drosophila yakuba has a widespread distribution in the Afrotropical region; D. melanogaster has spread its distribution from the same central Africa very recently and is now a cosmopolitan species (Lachaise et al. 1988). Thus, it cannot be expected that the average temperature in the habitats of D. melanogaster is higher enough than that in the habitats of D. yakuba to make a big between-species difference in generation time lengths. At any rate, there is no evidence to support the hypothesis that D. melanogaster has a ~1.4-fold shorter generation time length than D. yakuba.

When Tajima’s 1D test for the numbers of all silent substitutions in the melanogaster and yakuba lineages was performed individually for each locus with the critical probability from the sequential Bonferroni method (Rice 1989), only the ase and su(w⁰) loci showed a significant departure from the molecular-clock hypothesis (χ² = 9.0, P = 0.003 for 38 in melanogaster vs. 16 in yakuba at ase; χ² = 8.0, P = 0.005 for 32 in melanogaster vs. 13 in yakuba at su(w⁰)).

In conclusion, the melanogaster lineage showed higher substitution rates in the X-telomeric region than did the yakuba lineage, which may be explained by higher total mutation rates or much lower crossover frequencies in this region in D. melanogaster than in D. yakuba or both.

Comparison of orena vs. erecta Lineage Substitution Rates

Tajima’s (1993) 1D test for the numbers of total silent substitutions in the orena and erecta lineages showed a significant departure from the molecular-clock hypothesis at the y, l’sc, and ase loci (χ² = 20.0, P < 0.001 for y; χ² = 8.2, P < 0.005 for l’sc; χ² = 10.0, P < 0.002 for ase), for which I applied the sequential Bonferroni technique (Rice 1989). The pooled data from the 10 loci also showed higher substitution rates in the orena lineage than in the erecta lineage for synonymous substitutions in the coding regions (χ² = 30.8, P < 0.001) but not for noncoding substitutions (χ² = 2.0, P > 0.1) or replacement substitutions (χ² = 1.3, P > 0.1). As with the GC/AT substitution bias, the lack of significant difference in noncoding substitutions probably arose because the l’sc, and ase loci that showed a significant deviation from the molecular-clock hypothesis lacked sufficient noncoding data. Indeed, the difference in the numbers of synonymous substitutions between the two lineages was not significant when the y, l’sc, and ase loci were excluded (χ² = 2.9, P > 0.05 for 36 in orena vs. 23 in erecta).

Based on the findings of the significantly GC-biased silent substitutions at the y, l’sc, and ase loci of D. orena and of the significant silent-substitution-rate differences between the orena and the erecta lineages at the same loci, it is posited that the change in mutation matrix (mutational bias or both bias and total mutation rates) led to higher substitution rates in the orena lineage than in the erecta lineage.

Genetic Map of the D. yakuba X Chromosome

Table 5 summarizes the estimated map distances in the five regions obtained in this study and those in the table.
Two X-telomeric regions studied in Takano-Shimizu (1999). Figure 3 presents a comparison of the estimated genetic map lengths of the X chromosome between *D. yakuba* and *D. melanogaster* (FlyBase 1999). Previous studies suggest a lack of structural change in all seven regions in Table 5 between these two species (Lemeunier and Ashburner 1976). The estimated crossover frequencies were generally higher in *D. yakuba* than in *D. melanogaster*; ratios of *yakuba*-to-*melanogaster* map lengths varied greatly among the regions studied. Nonetheless, the results clearly demonstrate a remarkable difference in crossover frequencies between the two species specifically in the X-telomeric region. Except for the X-telomeric regions, position effects on the chromosomes can account for the heterogeneity in the ratios of map lengths of the two species. A previous cytological study showed that the chromosomal positions of the *AnnX, sol, l(2)gl*, and *dpp* loci, as well as the *y, su(w*) and *sta* loci, did not differ between the two species: *AnnX* and *sol* are located in proximal regions of the X chromosome, and *l(2)gl* and *dpp* are located on the tip of the second chromosome (Lemeunier and Ashburner 1976). These *AnnX-sol* and *l(2)gl-dpp* regions gave equal ratios of the lengths in the two species (1.6). In contrast, the chromosomal positions of the other five loci differ between the two species because of chromosomal structural changes. The *f, Sh*, and *bnb* loci are located in a region closer to the centromere in *D. melanogaster* than in *D. yakuba*, resulting in slightly higher ratios of the *yakuba*-to-*melanogaster* map lengths (3.5 and 2.8) than the *AnnX-sol* and *l(2)gl-dpp* intervals (see Fig. 3). In contrast, *snf* and *dx* are located in a region closer to the telomere in *D. yakuba* than in *D. melanogaster*, and this interval showed a relatively shorter *yakuba* map length. Thus, although the *yakuba* X-telomeric region has a remarkably higher crossover frequency than that of *melanogaster*, the regional heterogeneity in crossover frequencies of the two species in the other regions can be explained by the centromere and telomere effects. That is to say, proximity to the centromere and telomere reduces crossover frequencies (Mather 1939).

Two additional loci, *v* and *r*, were also studied. The results showed that the *r* locus was distally displaced between the *snf* and *v* loci in *D. yakuba* (fig. 3). This was inconsistent with the previous study (Lemeunier and Ashburner 1976), which reported no structural change in a 12D–18D region between the two species (*r* at 15A and *f* at 15F in *D. melanogaster*; thus, *r* should be next to *f* based on their results). Taken together, pairwise map distances indicated the following distal-to-proximal order of genes: *y, su(w*), *sta, dx, snf, r, v, f, Sh, bnb, AnnX, and sol*. The estimated distances were 3.6 cm for the *sta-dx* interval, 30.0 cm for *snf-r*, 5.2 cm for *r-v*, 30.0 cm for *v-f*, and 10.9 cm for *bnb-AnnX*, giving 100.6 cm as an estimate for a *yakuba* map length between *y* and *sol*. The ratio of the *yakuba*-to-*melanogaster* map lengths in the *y-sol* interval, which covers most of the X chromosome, was 1.5. This is very close to the ratios of map distances for the *AnnX-sol* and *l(2)gl-dpp* intervals, which supports the hypothesis that the relatively higher ratios in *f-bnb* and the lower ratio in *snf-dx* are attributable to their position effects.

Takano-Shimizu (1999) has shown that the magnitude of crossover frequency in the X-telomeric *y-su(w*)-*sta* region is >10-fold higher in *D. yakuba* than in *D. melanogaster*. The present study further revealed that a difference in crossover frequencies between these two species depended on regions but that the ratio of the *yakuba* map length to the *melanogaster* one ranged only from 1 to 3 in the other regions studied, with the overall ratio being 1.5 (the *y-sol* and *l(2)gl-dpp* regions). In conclusion, the findings suggested that the X-telomeric regions of *D. yakuba* specifically had much higher crossover frequencies than the *D. melanogaster* one.

Equal Crossover Frequencies in *D. yakuba* and *D. erecta*

In order to infer on which branches changes in crossover frequencies have occurred, crossover frequencies in the X-telomeric *y-sta* region were studied simultaneously for *D. yakuba* and *D. erecta*. *Drosophila yakuba* is likely to be more closely related to *D. melanogaster* than to *D. erecta* (fig. 1). The four different *F*<sub>1</sub> females of *D. yakuba* gave similar crossover frequencies: 0.036 (12/337) and 0.037 (13/347) from Takano-
In the D. yakuba lineage, significantly GC-biased silent substitutions in the coding regions ($G' = 13.8$, $P < 0.001$ for $20$ A/T→G/C and $3$ G/C→A/T substitutions), but not in the noncoding regions ($G' = 0.2$, $P > 0.5$ for $9$ A/T→G/C and $11$ G/C→A/T substitutions). Based on this finding, I previously assumed that the ancestral population of $D$. melanogaster and $D$. yakuba had intermediate crossover frequencies between those of the extant two species and that an increase in crossover frequencies was responsible for the biased substitutions toward major codons (all the major codons are G- or C-ending codons) at the $y$ locus. The finding of the local mutation effects is in line with the results of significant heterogeneities in G+C content among $140$–$340$-kb fragments (Carulli et al. 1993) and in intron G+C content among genes in Drosophila (Kliman and Hey 1993, 1994). Regional variation in G+C content and silent substitution rates along chromosomes has been well documented in mammals (Wolfe, Sharp, and Li 1989; Bernardi 1995; Matassi, Sharp, and Gautier 1999). The so-called isochore structure of mammalian chromosomes is a remarkable example. Although a recent finding on within-species variation patterns at silent sites in the MHC genes supports the hypothesis that selection is acting on the G+C content in the region (Eyre-Walker 1999), local mutational pressures may also contribute to the maintenance of the isochores (Filipski 1987; Sueoka 1988; Wolfe, Sharp, and Li 1989). One possible explanation for region-dependent mutation patterns is replication-timing effects. Indeed, replication timing differs between GC-poor and GC-rich regions precisely at one boundary of two domains (Tenzen et al. 1997). Studies of flies may yield similar findings with regard to the relationship between replication timing and substitution patterns.

Effects of Changes in GC/AT Biases on Substitution Rates

The $y$, $l'sc$, and $ase$ loci showed both significantly GC-biased silent substitutions and significantly higher substitution rates in $D$. orena than in $D$. erecta. These coincidences suggest that a change or changes in mutation matrix led simultaneously to the GC-biased substitutions and the higher substitution rates. However, because we do not know whether the population is in equilibrium or not, the numbers of A/T→G/C and G/C→A/T substitutions and their ratios in the $orena$ lineage do not provide a direct estimate of the mutational bias or of the steady-state substitution rate.

Discussion

Mutational Biases as a Cause of Between-Loci Heterogeneity in GC/AT Substitution Biases

This study clearly revealed that the patterns and degrees of GC/AT substitution biases depend on locus (or chromosomal region) and frequently change even among closely related species of Drosophila. The consistent silent-substitution patterns in the coding and noncoding regions were taken as evidence that the biased substitutions were due to local mutational biases, assuming that the noncoding regions were not under the same selection pressure as the coding regions and that the similar substitution biases in the two regions reflected their mutation pressures. However, most of the studied noncoding regions were introns, and selection on these untranslated regions and on their translated regions might be associated, for instance, through pre-mRNA stability related to gene expression levels. Only the $y$ and $sta$ loci include their 5′ flanking untranslated regions (the $l'sc$ region studied also contains its 5′ untranslated region, but the number of sites compared was only 51). The numbers of substitutions in the transcribed regions of $D$. orena were $15$ A/T→G/C and $1$ G/C→A/T substitutions at $y$, and $0$ A/T→G/C and $6$ G/C→A/T substitutions at $sta$, showing a clear difference in the GC/AT substitution biases between the two genes ($P$ in Fisher’s exact test $< 0.0001$). This substitution pattern is in very good agreement with the synonymous substitution patterns in their coding regions ($38$ A/T→G/C and $1$ G/C→A/T substitution at $y$, $0$ A/T→G/C and $6$ G/C→A/T substitutions at $sta$). Thus, at least in these cases, we could not attribute the biased substitutions to natural selection acting on gene products, and the observed variations were better explained by local mutation effects.

The finding of the local mutation effects is in line with the results of significant heterogeneities in G+C content among $140$–$340$-kb fragments (Carulli et al. 1993) and in intron G+C content among genes in Drosophila (Kliman and Hey 1993, 1994). Regional variation in G+C content and silent substitution rates along chromosomes has been well documented in mammals (Wolfe, Sharp, and Li 1989; Bernardi 1995; Matassi, Sharp, and Gautier 1999). The so-called isochore structure of mammalian chromosomes is a remarkable example. Although a recent finding on within-species variation patterns at silent sites in the MHC genes supports the hypothesis that selection is acting on the G+C content in the region (Eyre-Walker 1999), local mutational pressures may also contribute to the maintenance of the isochores (Filipski 1987; Sueoka 1988; Wolfe, Sharp, and Li 1989). One possible explanation for region-dependent mutation patterns is replication-timing effects. Indeed, replication timing differs between GC-poor and GC-rich regions precisely at one boundary of two domains (Tenzen et al. 1997). Studies of flies may yield similar findings with regard to the relationship between replication timing and substitution patterns.
Sueoka (1993) has shown that changes in mutational biases can be followed under some conditions by transient increases in substitution rates. This is true even if the steady-state substitution rates following the changes are equal to those preceding the changes. Using Sueoka’s deterministic mutation model, let us see this point in more concrete terms. For simplicity, consider only A/T→G/C and G/C→A/T neutral mutations (neglecting G→C and A→T mutations), and assume that all mutations are destined to be fixed in a population. Let \( v \) be the generation rate of mutating from an A or T nucleotide to a G or C nucleotide, and let \( u \) be that of mutating from a G or C to an A or T. Because the fixation probability of a neutral mutation is its frequency, \( v \) and \( u \) may be considered the actual mutation rates divided by two times the population size. The steady-state per-site substitution rate (the total number of mutations per site), \( k \), is given by

\[
k = v(1 - \hat{P}) + u\hat{P},
\]

where \( \hat{P} \) is the equilibrium G+C content and is equal to \( v/(u + v) \) (directional mutation pressure, or simply mutational bias in this article) (Sueoka 1962). Let us now consider a case in which the degree of mutational bias is suddenly changed from \( \alpha = v_0/(u_0 + v_0) \) to \( \beta = v/(u + v) \) at time \( t = 0 \). When the G+C content has gone from \( \alpha \) to a transient value \( P \) after \( t \) generations, the number of per-site substitutions accumulated during that period is

\[
M(P) = 2\beta(1 - \beta) \times \ln\left(\frac{\beta - \alpha}{\beta - P}\right) + (2\beta - 1)(P - \alpha) \tag{2}
\]

where \( \alpha < P < \beta \) for \( \alpha < \beta \) and \( \beta < P < \alpha \) for \( \beta < \alpha \). The number of substitutions in the same time interval under a new equilibrium condition (mutational bias = \( \beta \)) is given by

\[
M_e = 2\beta(1 - \beta) \times \ln\left(\frac{\beta - \alpha}{\beta - P}\right) \tag{3}
\]

Let \( r(P) \) be the ratio of the above two quantities, namely,

\[
r(P) = \frac{2\beta(1 - \beta) \times \ln\left(\frac{\beta - \alpha}{\beta - P}\right) + (2\beta - 1)(P - \alpha)}{2\beta(1 - \beta) \times \ln\left(\frac{\beta - \alpha}{\beta - P}\right)}.
\]

Importantly, this depends on \( \alpha \) and \( \beta \), but not on the individual mutation rates \( u \) and \( v \) (\( u_0 \) and \( v_0 \)). This ratio is always 1 when \( \beta = 0.5 \) (no mutational bias) and \( > 1 \) for \( \alpha < \beta \) and \( \beta > 0.5 \) and for \( \beta < \alpha \) and \( \beta < 0.5 \). Figure 4 illustrates \( r(P) \) as a function of \( P \) for \( \alpha = 0.45 \) and \( \beta = 0.8, 0.9, \) and 0.95. For example, when \( \alpha = 0.45 \) and \( \beta = 0.9 \), we obtain \( r(0.5) = 2.9 \). This well explains the actual data at \( y \), \( l^{sc} \), and \( \text{ase} \) of \( D. orena \) and \( D. erecta \): the average G+C content in the noncoding regions and the third positions of twofold-degenerate codons in the \( y \), \( l^{sc} \), and \( \text{ase} \) sequences of \( D. erecta = 45.7\% \), that of \( D. orena = 49.4\% \), and the ratio of the numbers of total substitutions on the \( \text{orena} \) and \( \text{erecta} \) lineages is \( 109/35 = 3.1 \). The condition \( \alpha = 0.45 \) and \( \beta = 0.9 \), however, does not necessarily require a change in steady-state substitution rates before and after the change in mutational biases. For example, when \( \alpha = 0.45 \), \( \beta = 0.9 \), and \( u = (1/2)u_0 \), the new equilibrium substitution rate, \( 2uv/(u + v) \), is equal to the original substitution rate, \( 2uv_0/(u_0 + v_0) \).

In conclusion, the results suggest big changes in mutational bias (likely from \( v/(u + v) = 0.45 \) to 0.9 or so) have occurred at the \( y \), \( l^{sc} \), and \( \text{ase} \) loci in the \( \text{orena} \) lineage, but an estimation of the new steady-state substitution rate in the \( \text{orena} \) lineage would be impossible with the presently available data. The substitution rates may be only transiently higher and may be gradually approaching the same level as that in the \( \text{erecta} \) lineage.

**Region-Dependent Regulation of Crossover Frequencies**

At present, I cannot be sure about the molecular basis for the >10-fold reduction in crossover frequencies in the telomeric region of the \( \text{melanogaster} \) X chromosome; the decreases in crossover frequencies in the two consecutive regions, the y-su(w^a) and su(w^a)-sta regions (Takano-Shimizu 1999), suggested a change with a regionwide effect. In Drosophila, Hawley (1980) proposed that there are four meiotic-chromosome-pairing or chromosome-synapsis-initiation sites crucial for normal levels of crossover frequencies on the X chromosome and that each site affects only the interval surrounding the site. Indeed, the low level of crossover frequencies in the telomeric region of the X chromosome may be attributed to the discontinuity of the synaptonemal complex in this region (Carpenter 1979). The present results might be explained by Hawley’s regional
regulation hypothesis: loss of a meiotic-chromosome-paring or synaptonemal complex-promoting site in the telomeric region in *D. melanogaster* may be one cause of the severely reduced levels of crossover frequencies in this region. Previous Drosophila studies revealed similarities and differences in the telomere structures among the closely related species (Young et al. 1983; Danilevskaya et al. 1998), but so far none of them have been able to explain the great difference in crossover frequencies. This between-species variation in crossover frequencies is good material for further molecular study on recombination rate determinants in Drosophila.

Three Possible Causes of Episodic Synonymous Substitution Rates

Significant locus-lineage interaction exists in the synonymous substitution rates among the Drosophila lineages (Takano 1998; Zeng et al. 1998). I summarize here three possible causes for the irregular synonymous substitutions: effective population size, recombination rate, and mutation. Because synonymous changes in coding regions are presumably under weak selection in Drosophila (Shields et al. 1988; Sharp and Li 1989; Akashi 1994, 1995; Moriyama and Powell 1997), changes in selection intensity acting directly on the genes in question can also inflate rate variation at synonymous sites. However, no direct evidence for locus-specific changes in selection intensity acting on synonymous changes among closely related Drosophila species has been reported so far. As mentioned in this article, the *yakuba* lineage showed significantly GC-biased silent substitutions in the coding but not in the noncoding regions, specifically at the y locus. This might be an example of such locus-specific changes in selection intensity on synonymous changes.

Akashi (1996) has suggested that a reduction in effective population size is responsible for lower codon biases and faster replacement and synonymous substitution rates in *D. melanogaster* than in *D. simulans*. Changes in effective population sizes have generally been considered a genomewide effect, but this is not always the case. Indeed, faster synonymous substitution rates in *D. melanogaster* were observed in many genes; however, no difference in synonymous substitution rates between the two species was observed in the four *achaete-scute* complex (AS-C) genes and the *ci* gene (Takano 1998). The AS-C and *ci* genes are located in regions of severely reduced crossover frequencies, AS-C on the tip of the X chromosome and *ci* on the fourth chromosome (Lindsley and Sandler 1977). A lack of between-species differences in substitution rates at these loci is most likely attributable to the low efficacy of natural selection as the result of the very low crossover frequencies, regardless of their effective population sizes (Takano 1998). Possible effects of recombination frequencies on the efficacy of selection are described below. Thus, combined with regional variations in crossover frequencies along chromosomes, fluctuations in effective population sizes can contribute to the heterogeneous synonymous substitution rates among genes.

Second, local changes in recombination rates may also produce irregular synonymous substitution patterns (Charlesworth 1994; Comeron, Kreitman, and Aguadé 1999; Takano-Shimizu 1999). This is because recombination frequencies affect the effective population sizes and the effectiveness of natural selection in several ways (although the hitchhiking and background selection effects may be viewed as special cases of the Hill-Robertson effect; Comeron, Kreitman, and Aguadé 1999). Selective sweeps of linked advantageous mutations and rapid elimination of deleterious mutations reduce the effective number of gametes that have reasonable probabilities of fixations in future generations. The former process is known as hitchhiking (Maynard Smith and Haigh 1974), and the latter is known as background selection (Charlesworth, Morgan, and Charlesworth 1993). Moreover, interference between selective loci also increases stochastic variances of the amount of change in mutant frequency per generation, the so-called Hill-Robertson effect (Hill and Robertson 1966; Felsenstein 1974). All of these effects result in reduced efficacy of natural selection, followed by accelerated substitution rates of slightly deleterious mutations (Birky and Walsh 1988; Charlesworth 1994; McVean and Charlesworth 1999, 2000; Stephan, Charlesworth, and McVean 1999).

Indeed, Munté, Aguadé, and Segarra (1997) have suggested that a change in crossover frequency due to chromosome rearrangements affected the degree of codon bias and synonymous substitution rates between *D. melanogaster* and *D. subobscura*. The codon bias for *y* and *scute* (one member of the AS-C complex) is much higher in *D. subobscura* than in *D. melanogaster*, and their synonymous substitution rates are the highest among the 18 genes compared. Munté, Aguadé, and Segarra (1997) assumed from their chromosomal locations that the *y* and AS-C genes were in regions of normal crossover frequencies in *D. subobscura* but in regions of very reduced crossover frequencies in *D. melanogaster*. The actual story, however, is not so simple, because a severe reduction of crossover frequency in the telomeric region of the X chromosome occurred in the ancestral population of *D. melanogaster* and *D. simulans* after the separation from the *electa* and *yakuba* lineages. This suggests that the *y* and AS-C regions of *D. melanogaster* suffered a two-step reduction in crossover frequencies, which resulted in the reduction of the efficacy of selection and then in the AT-biased synonymous substitutions in *D. melanogaster* (Takano-Shimizu 1999). This second reduction in crossover frequencies may explain the higher replacement and synonymous substitution rates in the *melanogaster* lineage than in the *yakuba* lineage.

Third, the present findings strongly suggest that local changes in mutational biases and total mutation rates have greatly contributed to DNA sequence evolution. Changes in mutational biases can lead to transient increases in substitution rates even if the steady-state substitution rates following the changes are equal to those preceding the changes. Evidence of between-lineages evolutionary rate variation has been recently reported in *Insecta* (Friedrich and Tautz 1997) and in Drosophila
(Rodríguez-Trellis, Tarrio, and Ayala 1999). Both studies suggested changes in mutation rates as the cause. What is more, this study revealed striking between-loci variability as well as between-lineages variation in substitution patterns in the very closely related species of Drosophila. These variations are best explained by local mutation effects, which, in turn, are very likely to be responsible for the higher substitution rates of the y, l’,sc, and ase genes in the orena lineage as compared with the erecta lineage. It is conceivable that changes also occurred in other regions and in other genomes. Mutation has direct effects on substitution rates, and its effects seem to be larger than population size and recombination rate effects.

In conclusion, although an exact evaluation of the relative contribution of each factor is difficult, effective population size, recombination, and mutation all contribute to the synonymous-substitution-rate variation, but mutation probably has the largest effect.

Acknowledgments

I thank Y. Ishii for technical assistance and L. Gilner for improving the manuscript. I am grateful to T. Ohta for her encouragement and comments and to two anonymous reviewers for their criticisms and suggestions. I also thank the National Drosophila Species Resource Center, C. C. Laurie, and N. Inomata for fly stocks. This work was supported by the Ministry of Education, Science, Sports and Culture of Japan.

LITERATURE CITED


