Ghiradella, and D.M. Ruden 2003, Neurotoxicology 24: 435-442; Hoftman, G.S., R.Y. Leavitt, T.A. Fleister, J.R. Minor, and A.S. Fauci 1990, The Amer. J. Med. 89: 4, 403-410; Lakhotia, S.C., and A.K. Singh 1989, J. Genet. 68: 129-138; Lis, J.T., J.A. Simon, and C.A. Sutton Cell 35: 403-413; Mukhopadhyay, I., A. Nazir, K. Mahmood, D.K. Saxena, M. Das, S.K. Khanna, and D.K. Chbowdhuri 2002, Cell Biol. Toxicol. 18: 1-11; Mukhopadhyay, I., D.K. Saxena, and D.K. Chowdhuri 2003, Environ. Health Persp. 3: 1926-1932; Nazir, A., I. Mukhopadhyay, D.K. Saxena, M. Siddiqui, and D.K. Chowdhuri 2003, J. Biochem. Mol. Toxicol. 17: 98-107; Nover, L., 1994, *Heat Shock Response of Eukaryotic Cells*, Berlin, Springer Verlag; Nover, L., 1991, *The Heat Shock Response*, Boca Raton, FL: CRC Press, pp.5-344; Ren, S., T.F. Kalhorn, G.B. McDonald, C. Anasetti, F.R. Appelbaun, and J.T. Slattory 1998, Clin. Pharmacol. Therap. 64: 289-301; Scott, D.G.I., and P. Bacon 1984, The Amer. J. Med. 76: 3, 377-384; Shanafelt, T.D., T. Lin, S.M. Geyer, C.S. Zent, N. Leung, B. Kebat, D. Bowen, M.R. Grever, J.C. Byrd, and N.E. Kay 2007, Cancer 109: 11, 2291-2298; Young, S.D., M. Whissell, J.C.S. Noble, P.O. Cano, P.G. Lopez, and C.J. Germond 2006, Clin. Can. Res. 12: 3092-3098.

Inconsistent associations between recombination rate and codon bias across *Drosophila* species.

Varney, Robin L., and Mohamed A.F. Noor. Biology Department, Duke University, Durham, NC USA; noor@duke.edu.

Abstract

A positive association between recombination rate and codon bias has been observed at broad and fine scales in *Drosophila* species. However, this relationship is complicated by other genomic features that correlate with codon bias. No prior studies have evaluated the relationship between recombination rate and codon bias across multiple species within this genus. Utilizing published recombination maps along with complete genome sequences, we contrasted recombination rate to codon bias and intronic GC across four *Drosophila* species. We did not observe a consistent significant correlation between recombination rate and codon bias across the species examined. Indeed, we did not observe even the previously reported trend of higher codon bias in regions of high recombination, though we saw some evidence this pattern may be affected by differences between centromeric/telomeric regions and central regions. More fine-scale recombination data from more *Drosophila* species is necessary for a comprehensive picture of the relationship between recombination rate and codon bias.

Introduction

Rates of recombination vary across *Drosophila* genomes at broad and fine scales and have been shown to be correlated with codon usage bias (Cirulli *et al*., 2007; Comeron *et al*., 1999; Hey and Kliman, 2002; Kliman and Hey, 1993; Marais *et al*., 2001; Singh *et al*., 2005; Stevison and Noor, 2010). Specifically, regions of high recombination tend to have disproportionately high usage of G- or C-ending codons relative to regions of low recombination. The relationship of recombination rate to codon bias is complicated by other features that also correlate with codon bias (*e.g.*, expression level) but may nonetheless indicate the intensity of natural selection on translation efficiency (Duret and Mouchiroud, 1999). However, biased gene conversion (BGC), a neutral process, can also occur as a result of high recombination rates and result in strong codon bias, but this force would apply equally in codons as in noncoding (*e.g.*, intronic) regions. Analysis of recombination rate variation and codon bias across several species of *Drosophila* may provide a more comprehensive picture of this relationship.

Several previous studies have documented a positive correlation between recombination rate and codon bias on a genome-wide scale in *Drosophila melanogaster* (Comeron *et al*., 1999; Hey and Kliman, 2002; Kliman and Hey, 1993; Marais *et al*., 2001, 2003; Singh *et al*., 2005). The degree of association between codon bias and recombination varies over the range of recombination, and a much stronger relationship between codon bias and recombination has been observed in genes with the lowest levels of recombination (Comeron *et al*., 1999; Hey and Kliman, 2002; Kliman and Hey, 1993). Further, a strong negative correlation has been documented between recombination rate and codon bias on the X chromosome (Singh *et al*., 2005).

Examination of the association between recombination rate and codon bias has been conducted on a much smaller spatial scale in *D. pseudoobscura* and *D. persimilis* (Cirulli *et al*., 2007; Stevison and Noor, 2010). Cirulli *et al*. (2007) detected a significant positive association between crossover rate and codon bias across 126 full or partial gene CDSs within a 2-megabase (Mb) interval on the XL chromosome arm of *D. pseudoobscura*.

Stevison and Noor (2010) conducted a fine-scale analysis of recombination rate variation on the right arm of the X chromosome (XR) and the entire second chromosome of *D. persimilis* and observed a significant positive correlation between recombination rate and codon bias on chromosome 2 but not on XR. As more frequently used codons in *Drosophila* genomes end in G or C, codon bias is often measured by the GC-content of third-base positions in codons (Akashi and Schaeffer, 1997; Bachtrog, 2007). Stevison and Noor (2010) also examined GC content at fourfold degenerate codons (GC4), which should be less constrained than non-degenerate codons, and GC content in small introns (<100 bp), which appear to be under less evolutionary constraint than longer introns (de Procé *et al*., 2009; Parsch, 2003). They observed strong correlations between recombination rate and GC4 and GC content in small introns in chromosome 2 (Stevison and Noor, 2010). For their analysis, they excluded 5 Mb upstream and downstream from the centromere and telomere—which undergo little crossing over and can potentially disrupt correlations between recombination rate and other genomic features—and one interval of especially high recombination rate (which may have been anomalously inflated) from each chromosome (Stevison and Noor, 2010).

These previous studies have examined the relationship between recombination rate and codon bias in individual *Drosophila* species, but none have evaluated the relationship across multiple species within a genus. The publication of broad and fine scale recombination maps along with complete genome sequences of several *Drosophila* species allow for the evaluation of meiotic recombination and codon bias across this genus. Broad-scale recombination linkage maps have been generated for *D. virilis* and *D. mojavensis* by Huttunen *et al*. (2004) and Staten *et al*. (2004), respectively, using microsatellite markers. More recently, fine-scale recombination maps have been generated for *D. pseudoobscura* and *D. persimilis* by Kulathinal *et al*. (2008) and Stevison and Noor (2010), respectively, using SNP genotyping markers. Complete genome sequences for each of these four species have been made available through the publication of the *Drosophila* 12 genomes sequencing project (Clark *et al*., 2007). In an effort to understand better the relationship between meiotic recombination and codon bias in *Drosophila*, we utilized these data to contrast recombination rate to GC4 and intronic GC across four *Drosophila* species.

Methods

Recombination linkage maps of microsatellite markers for multiple chromosomes by Huttunen *et al.* (2004) and Staten *et al.* (2004) provide Kosambi recombinational distances between pairs of markers in centiMorgans for *D. virilis* and *D. mojavensis*, respectively. Microsatellite primer sequences obtained from Huttunen and Schlötterer (2002) and Staten *et al.* (2004) were BLASTed against the caf1 *D. virilis* and *D. mojavensis* genome scaffolds, respectively. The physical distance between adjacent markers was determined and recombination rates were calculated in centiMorgans per megabase (cM/Mb) for pairs of markers. For *D. pseudoobscura* (Kulathinal *et al.*, 2008) and *D. persimilis* (Stevison and Noor, 2010), the start/end position for each SNP genotyping marker along the scaffolds and Kosambi recombination distances between pairs of markers in centiMorgans were obtained from the Dryad data repository (doi:10.5061/dryad.484; doi:10.5061/dryad.1877), and used to calculate recombination rate in cM/Mb.

Chromosome and scaffold sequences containing the markers used for recombination rate calculations were downloaded from GenBank. GC content at fourfold degenerate codons have been shown to be less constrained than non-degenerate codons (de Procé *et al.*, 2009). Parsch *et al.* (2010) showed bases 8 – 30 of introns ≤65 bp in length are under the least selective constraint and can be used as a reference for unconstrained evolution. A series of custom Perl scripts were used to annotate the GenBank files and calculate GC and AT content at third positions in codons for fourfold degenerate codons (GC4 and AT4, respectively) and GC and AT content for bases 8 – 30 of introns ≤65 bp in length (GCi_65 and ATi_65, respectively). The percent of GC4 and GCi_65 were calculated for each recombination interval and regressed against recombination rate in Excel.

Broad-scale microsatellite linkage map data were available for chromosomes 2, 3, 4, and 5 of *D. virilis* and chromosomes 2, 3, 4, 5, and X of *D. mojavensis* (Huttunen *et al.*, 2004; Staten *et al.*, 2004). Data from *D. virilis* and *D. mojavensis* autosomal chromosomes were combined into a single group for each species for regression analyses. Fine-scale SNP genotype data were available for chromosome 2 of *D. pseudoobscura* and chromosomes X and 2 of *D. persimilis* (Kulathinal *et al.*, 2008; Stevison and Noor, 2010). One *D. mojavensis* recombination interval had an especially high recombination rate and was excluded from our regression analysis. Kosambi recombinational distances between microsatellite and SNP genotyping markers varied widely. To obtain similar Kosambi recombinational distances between markers for each of the four species, short recombination intervals of *D. pseudoobscura* and *D. persimilis* were grouped together for further regression analyses.

Results

Table 1 presents regression analyses of recombination rates with %GC4 and %GCi_65 for four species of *Drosophila* at different scales. Data from *D. mojavensis* and *D. persimilis* X chromosome regression analyses are not shown. Analyses of initial recombination intervals reveal a significant correlation between recombination rate and %GCi_65 only in *D. pseudoobscura*, a marginally significant correlation between recombination rate and both %GC4 and %GCi_65 in *D. virilis*, and a marginally significant correlation between recombination rate and %GC4 only in *D. mojavensis* and *D. pseudoobscura*. When recombination intervals of *D. pseudoobscura* and *D. persimilis* were combined to make them more similar in size to *D. virilis* and *D. mojavensis* recombination intervals, no significant correlation was observed between recombination rate and %GC4 or %GCi_65 in either species.

Table 1. Results of regression analysis of recombination rate and %GC4 and %GCi_65 across four *Drosophila* species.

				%GC4 regressed against cM/Mb			%GCi_65 regressed against cM/Mb		
	N	cM range	cM/Mb range	r	Coefficients	*P* value	r	Coefficients	*P* value
D. virilis (Chr 2,3,4,5)	15	10.8 - 43.4	3.1 - 23	0.441	68.486	0.100	0.425	109.745	0.114
D. virilis ≤25 cM	9	10.8 - 24.9	3.1 - 23	0.443	64.085	0.232	0.396	94.221	0.292
D. virilis >26 cM	6	25.4 - 43.4	3.4 - 9.2	0.019	2.255	0.972	0.733	-269.330	0.097
D. mojavensis (Chr 2,3,4,5)	12	1.2 - 37.1	1.4 - 11.7	0.509	52.103	0.091	0.259	31.033	0.417
D. mojavensis ≤10 cM	9	1.2 - 9.9	1.4 - 11.7	0.670	60.031	**0.048**	0.347	38.176	0.360
D. mojavensis >10 cM	3	11.7 - 37.1	1.8 - 11.1	0.846	-475.387	0.358	0.875	-795.930	0.321
D. pseudoobscura (Chr 2)	49	0 - 9.8	0 - 15	0.271	5.399	0.060	0.420	15.462	**0.003**
D. pseudoobscura (intervals combined)	10	5.1 - 17	2.8 - 5.3	0.458	-6.962	0.183	0.112	-4.432	0.758
D. pseudoobscura (intervals combined) ≤10 cM	5	5.1 - 9.5	2.9 - 5.3	0.773	-9.744	0.125	0.801	-30.859	0.103
D. pseudoobscura (intervals combined) >10 cM	5	12.1 - 17	2.8 - 4.8	0.801	29.549	0.103	0.873	36.006	**0.053**
D. persimilis (Chr 2)	129	0 - 25.2	0 - 30	0.104	4.429	0.243	0.035	1.723	0.697
D. persimilis (intervals combined)	13	1 - 30.4	0.4 - 8.3	0.049	-2.404	0.874	0.215	22.713	0.481
D. persimilis (intervals combined) ≤10 cM	8	1 - 8.2	0.4 - 3.6	0.492	-9.368	0.215	0.550	-24.718	0.158
D. persimilis (intervals combined) >10 cM	5	10.6 - 30.4	4.9 - 8.3	0.050	-3.287	0.936	0.299	45.721	0.625

For each species, data were sorted by recombination interval size and separated into small and large intervals to examine correlation between recombination rate and %GC4 and %GCi_65 using similar sized intervals. For *D. mojavensis*, *D. pseudoobscura*, and *D. persimilis*, small recombination intervals were ≤10 cM, and large recombination intervals were >10 cM. The available recombination intervals for *D. virilis* were larger and were separated into small intervals ≤25 cM and large intervals >26 cM. Subsequent regression analyses revealed a significant correlation between recombination rate and %GC4 in *D. mojavensis* ≤10 cM and between recombination rate and %GCi_65 in *D. pseudoobscura* >10 cM; marginally significant correlation was observed between recombination rate and %GC4 and %GCi_65 in D. psedoobscura ≤10 cM, between recombination rate and %GC4 in *D. pseudoobscura* >10 cM, and between recombination rate and %GCi_65 in *D. virilis* >26 cM.

Discussion

We did not observe a consistent significant correlation between recombination rate and codon bias across the four species of *Drosophila* examined. Overall, broader scale recombination intervals showed more significant correlations between recombination rate and %GC4 and %GCi_65 than fine-scale recombination intervals. However, when the fine-scale recombination intervals were grouped together to make larger intervals and reanalyzed, the relationships between recombination rate and %GC4 and %GCi_65 were not significant. Further separation of the data into small and large recombination intervals did not provide any more insight into the relationship between recombination rate and codon bias among *Drosophila* species.

Unlike previous studies, we did not observe a trend of higher codon bias in regions of high recombination. Recombination rate and codon bias were not significantly correlated on a genome-wide scale in *D. virilis* or *D. mojavensis* in our study, as has been observed in *D. melanogaster*. Although the fine-scale regression analysis of only the second chromosome for *D. pseudoobscura* revealed significant correlation between recombination rate and intronic GC and marginally significant correlation between recombination rate and codon bias, analysis of the second chromosome of *D. persimilis* did not reveal significant correlations between recombination rate and codon bias or intronic GC when all data points were examined.

Our results initially appear inconsistent with those reported by Stevison and Noor (2010) due to differences in the data sets used in each study. However, when we reanalyzed the data using their specified parameters—excluding 5 Mb upstream and downstream from the centromere and telomere and high recombination rates from the analysis—for *D. persimilis*, we observed the same significant correlation between recombination rate and codon bias and intronic GC (results not shown) that they reported. Clearly, excluding centromeric and telomeric genomic regions that undergo little crossing over affects the results of studies examining the correlation between recombination rate and codon bias.

Our study was limited by the scarcity of fine-scale recombination data available for analysis in these *Drosophila* species. Further analyses of fine-scale recombination rates at multiple chromosomes in more *Drosophila* species are necessary to develop a comprehensive picture of the association of recombination rate and codon bias in this genus.

Acknowledgments: This work was funded by National Institutes of Health grants GM076051 and GM086445, including an ARRA supplement. The authors thank Shannon McDermott and Caiti Smukowski for comments on the manuscript.

References: Akashi, H., and S.W. Schaeffer 1997, Genetics 146: 295-307; Bachtrog, D., 2007, J. Mol. Evol. 64: 586-590; Cirulli, E.T., R.M. Kliman, and M.A.F. Noor 2007, J. Mol. Evol. 64: 129-135; Clark, A.G., M.B. Eisen, D.R. Smith, C.M. Bergman, *et al.* 2007, Nature 450: 203-218; Comeron, J.M., M. Kreitman, and M. Aguade 1999, Genetics 151: 239-249; de Procé, S.M., D.L. Halligan, P.D. Keightly, and B. Charlesworth 2009, J. Mol. Evol. 69: 601-611; Duret, L., and D. Mouchiroud 1999, Proc. Natl. Acad. Sci. USA 96: 4482-4487; Hey, J., and R.M. Kliman 2002, Genetics 160: 595-608; Huttunen, S., J. Aspi, A. Hoikkala, and C. Schlötterer 2004, Heredity 92: 263-269; Huttunen, S., and C. Schlötterer 2002, Molec. Ecol. Notes 2: 593-597; Kliman, R.M., and J. Hey 1993, Mol. Biol. Evol. 10: 1239-1258; Kulathinal, R.J., S.M. Bennett, C.L. Fitzpatrick, and M.A.F. Noor 2008, Proc. Natl. Acad. Sci. USA 105: 10051-10056; Marais, G., D. Mouchiroud, and L. Duret 2001, Proc. Natl. Acad. Sci. USA 98: 5688-5692; Marais, G., D. Mouchiroud, and L. Duret 2003, Genet. Res. 81: 79-87; Parsch, J., 2003, Genetics 165: 1843-1851; Parsch, J., S. Novozhilov, S.S. Saminadin-Peter, K.M. Wong, and P. Andolfatto 2010, Mol. Biol. Evol. 27: 1226-1234; Singh, N.D., J.C. Davis, and D.A. Petrov 2005, J. Mol. Evol. 61: 315-324; Staten, R., S. Dixon Schully, and M.A.F. Noor 2004, BMC Genetics 5:12; Stevison, L., and M.A.F. Noor 2010, J. Mol. Evol. 71: 332-345.

Drosophilids of Perumalai hills of Kodaikanal (Tamilnadu State, India).

Harini, B.P., and D.S.S. Pranesh. Department of Zoology, Drosophila Culture Laboratory, Bangalore University, Jnana Bharathi Campus, Bangalore 560 056, India, Karnataka.

Significant progress has been made in the field of taxonomy and systematics of the family Drosophilidae (Diptera) in India. The *Drosophila* species are observed in essentially any environment, from the sea level to considerable altitudes, and in temperate as well as in equatorial zones (Throckmorton, 1975). However, Lachaise (1979) suggests that these species are subject to restrictions as regards the habitats they live in. Many factors affect the ability of a species to survive