DOI: 10.1534/genetics.107.085803

Multilocus Patterns of Polymorphism and Selection Across the X Chromosome of *Caenorhabditis remanei*

Asher D. Cutter[1]

Department of Ecology and Evolutionary Biology and Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Ontario M5S 3G5, Canada

Manuscript received December 13, 2007
Accepted for publication January 10, 2008

ABSTRACT

Natural selection and neutral processes such as demography, mutation, and gene conversion all contribute to patterns of polymorphism within genomes. Identifying the relative importance of these varied components in evolution provides the principal challenge for population genetics. To address this issue in the nematode *Caenorhabditis remanei*, I sampled nucleotide polymorphism at 40 loci across the X chromosome. The site-frequency spectrum for these loci provides no evidence for population size change, and one locus presents a candidate for linkage to a target of balancing selection. Selection for codon usage bias leads to the non-neutrality of synonymous sites, and despite its weak magnitude of effect ($N_e s \sim 0.1$), is responsible for profound patterns of diversity and divergence in the *C. remanei* genome. Although gene conversion is evident for many loci, biased gene conversion is not identified as a significant evolutionary process in this sample. No consistent association is observed between synonymous-site diversity and linkage-disequilibrium-based estimators of the population recombination parameter, despite theoretical predictions about background selection or widespread genetic hitchhiking, but genetic map-based estimates of recombination are needed to rigorously test for a diversity–recombination relationship. Coalescent simulations also illustrate how a spurious correlation between diversity and linkage-disequilibrium-based estimators of recombination can occur, due in part to the presence of unbiased gene conversion. These results illustrate the influence that subtle natural selection can exert on polymorphism and divergence, in the form of codon usage bias, and demonstrate the potential of *C. remanei* for detecting natural selection from genomic scans of polymorphism.

NEUTRAL and selective processes interact to shape polymorphism and divergence across genomes, yet it continues to be a difficult problem to derive a robust understanding of the relative influence of the different component forces that contribute to evolution. Demographic change in populations is a neutral process that affects the entire genome, so population samples for a large number of loci are required to accurately infer the general features of demographic history that are recorded in patterns of DNA sequence polymorphism. Selection, on the other hand, acts locally on specific targets, so a reasonable characterization of the background patterns molded in the genome by global forces (like population size change) is necessary to accurately detect signatures of selection from molecular variation relative to this background. Here, I characterize nucleotide polymorphism and divergence across the X chromosome of the nematode *Caenorhabditis remanei*, an obligately outbreeding relative of the classic model organism *C. elegans*, to elucidate the contributions of selective and neutral processes in the evolutionary history of this organism.

Due to recent completion of a genome-sequencing effort and its close relationship with *C. elegans*, *C. remanei* is experiencing new interest as a focal taxon for comparative research in evolutionary genetics and genomics (Haag *et al.* 2007). The high nucleotide diversity previously reported for this species provides ample data for population genetic inference from even short stretches of DNA sequence (Graustein *et al.* 2002; Jovelin *et al.* 2003; Haag and Ackerman 2005; Cutter *et al.* 2006a). Analysis of a sample in a previous study suggested that the effective population size of *C. remanei* is large ($N_e \sim 1.6 \times 10^6$) and that its demographic history as reflected in sequence polymorphisms may be relatively uncomplicated (Cutter *et al.* 2006a), which should facilitate attempts to infer the action of natural selection from patterns of polymorphism. Beyond this, however, the structure and dynamics of *C. remanei* populations are unknown. An understanding of how natural selection shapes the *C. remanei* genome will provide important evolutionary insights from a species that represents the ancestral obligately outcrossing breeding system in the genus, and therefore a contrast to the derived condition

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. EU480769–EU481404.

[1]*Address for correspondence:* Department of Ecology and Evolutionary Biology, University of Toronto, 25 Harbord St., Toronto, ON M5S 3G5, Canada. E-mail: asher.cutter@utoronto.ca

of hermaphroditism in *C. elegans* and *C. briggsae* (Cho *et al.* 2004; Kiontke *et al.* 2004).

The large effective population size and obligate outcrossing of *C. remanei*, in contrast to its self-fertile relatives (Sivasundar and Hey 2003; Barrière and Félix 2005; Cutter 2006; Cutter *et al.* 2006b), imply that even very weak selection may yield an evolutionary response and that the response to selection may be tempered by the local recombination environment. Selection among alternative synonymous codons for translational efficiency and/or accuracy is expected to be due to very small selection differentials (Li 1987; Bulmer 1991), yet is a notable force shaping coding sequences within a variety of taxa (Duret 2002; Merkl 2003), including nematodes (Stenico *et al.* 1994; Duret and Mouchiroud 1999; Cutter and Charlesworth 2006; Cutter *et al.* 2006c). Directional selection in general eliminates neutral genetic variation in the neighborhood of a selected target (Maynard Smith and Haigh 1974; Charlesworth *et al.* 1993). However, the size of the region that is subject to reduced variation will depend on the number of recombination events that occur by the time that the selected variant becomes fixed in (or eliminated from) the population, resulting in a prediction of reduced neutral genetic diversity in regions that experience low levels of recombination (Wiehe and Stephan 1993; Hudson and Kaplan 1995). However, some neutral processes might result in a similar pattern (Lercher and Hurst 2002; Hellmann *et al.* 2003), making empirical support mixed for selective explanations of diversity–recombination correlations. Thus, it is of considerable interest to understand the relative importance of selective and neutral forces in shaping patterns of polymorphism across genomes.

Here, I survey nucleotide polymorphism for 40 loci across the X chromosome in a sample of *C. remanei* to (i) infer the potential role of demographic history in shaping patterns of genetic variation, (ii) quantify weak selection on codon usage, and (iii) evaluate the potential for selection at linked sites to alter diversity levels in genomic regions experiencing different recombination environments.

MATERIALS AND METHODS

Molecular methods: Genomic DNA was amplified from *C. remanei* individuals by directly adding single males to QIAGEN (Valencia, CA) Repli-g mini kit reactions. Diluted aliquots of the resulting whole-genome-amplification DNA samples were then used as template for locus-specific amplification by PCR. I designed PCR primers with Primer3 (http://primer3.sourceforge.net) for putatively X-linked loci with predicted long exons, using contigs that had strong synteny to the *C. elegans* X chromosome from the preliminary assembly of the *C. remanei* genome sequence produced by the Washington University School of Medicine Genome Sequencing Center (supplemental Table 1). The X chromosome experiences the greatest interspecific synteny in Caenorhabditis (Stein *et al.* 2003; Hillier *et al.* 2007). Both strands of the resulting PCR products were then sequenced directly on an ABI 3730 through the University of Arizona's Genomic Analysis and Technology Core sequencing service. Five of 45 primer sets yielded obvious double peaks in all or some sequence traces, indicating the presence of heterozygotes and therefore linkage to autosomes; in this study, I focus only on the remaining putatively X-linked loci (supplemental Table 1). Sequences for each of 40 loci were obtained for 16 *C. remanei* individuals (supplemental Table 1): 14 were derived from isofemale lines collected by S. E. Baird in the Wright State University forest in Dayton, Ohio (PB207, PB210, PB211, PB213, PB215, PB219, PB242, PB243, PB244, PB247, PB249, PB252, PB253, PB275), as well as strains PB4641 (Brooklyn, NY) and SB146 (Freiburg, Germany) from the Caenorhabditis Genetics Center (Cutter *et al.* 2006a). All analyses of polymorphism focus on the sample of strains from Ohio, unless stated otherwise. All loci include coding sequence only, and average 742 bp in length after trimming primer and ambiguous bases. Two loci have slightly smaller sample sizes because PCR was unsuccessful for one (*Cre-lam-2*) or three (*Cre-mbk-1*) strains. Assuming that the *C. remanei* X chromosome is of a size similar to the *C. elegans* X chromosome (~20 Mb, ~50 cM), these loci are expected to be spaced ~2/Mb and 1/cM.

Sequence analysis: Sequence trace editing and alignment were performed using Sequencher v. 4.7 and BioEdit v. 7.05.3 to confirm sequence quality and to remove primer sequences. I used DnaSP v. 4.10.9 to calculate diversity from pairwise differences (π) and from the number of segregating sites (θ) (Watterson 1975; Nei and Li 1979) and to conduct tests of neutrality [Tajima's (1989) *D*, Fu and Li's (1993) *D** and *F**], each computed separately for synonymous (denoted with subscript "s") and nonsynonymous sites (denoted with subscript "a"). SITES was used to calculate Tajima's *D* for the restricted set of sites corresponding to preferred–preferred and unpreferred–unpreferred polymorphisms (Hey and Wakeley 1997). All correlations use the nonparametric Spearman's rank correlation, r_{SRC}.

The population recombination parameter ($\rho = 4N_e r$; effective population size N_e, recombination rate r) was inferred with LDhat (ρ_{LDhat}; McVean *et al.* 2002; G. McVean, http://www.stats.ox.ac.uk/~mcvean/LDhat/) and maxhap (ρ_{H01} and ρ_{H01+f}, which accounts for gene conversion, f; R. Hudson, http://home.uchicago.edu/~rhudson1) based on Hudson's (2001) composite-likelihood method. LDhat uses a finite-sites mutational model, unlike the infinite-sites model used by maxhap, which results in slightly different estimates of ρ despite an otherwise identical composite-likelihood methodology. Values of ρ_{LDhat} and ρ_{H01} were strongly correlated ($r_{SRC} = 0.87$, $P < 0.0001$; Ohio sample only). With maxhap, the ratio of gene conversion to recombination rate (f) was inferred jointly with ρ_{H01+f}. Although gene conversion tract lengths are unknown in this species, transgene-mediated gene conversion in *C. elegans* following double-strand break repair of transposon excision yields tracts at least 191 bp long (Plasterk and Groenen 1992) and reports of a few hundred base pairs are typical of other taxa (Hilliker *et al.* 1994; Jeffreys and May 2004); a gene conversion tract length of 400 bp is assumed in the calculations presented here. Estimates of ρ_{H01+f} made with other tract lengths (50, 200, 1000) did not dramatically alter the values (not shown), although shorter tract lengths yielded somewhat higher estimates of f and slightly lower estimates of ρ_{H01+f}. Values for ρ_{H01} could not be obtained for two loci (*Cre-myo-2* exon 8, *Cre-let-2*), which were excluded from corresponding analyses.

The *C. elegans* and *C. briggsae* orthologs of the *C. remanei* loci were inferred by identification in the TreeFam database on the WormBase website and by manual reciprocal-best-hits Blast. Alignments of the coding sequences that correspond to the

regions assayed for polymorphism were made by eye or with the aid of ClustalW in BioEdit. I then used the program Kaks_calculator (Li *et al.* 2006) to compute pairwise rates of synonymous (d_S) and nonsynonymous (d_N) site divergence with the Goldman–Yang method (Goldman and Yang 1994) for the coding regions containing *C. remanei* polymorphism data, where a consensus sequence was used for *C. remanei*. Because *C. remanei* shares a most recent common ancestor with *C. briggsae* (Kiontke *et al.* 2004), *C. briggsae–C. remanei* measures of divergence are used unless stated otherwise. The *C. remanei* consensus sequences were also used for computation of the codon bias statistics F_{op} (frequency of optimal codons; Stenico *et al.* 1994) and ENC (effective number of codons; Ikemura 1985) as well as G + C base content, using the program CodonW (J. Peden, http://codonw.sourceforge.net) and applying the *C. elegans* optimal codon table (Stenico *et al.* 1994). The resulting F_{op} and GC-content values were then used to generate corrected estimates of *C. briggsae–C. remanei* d_S (d_S') via multiple regression that removes the correlation between synonymous divergence, codon bias, base composition, and their interaction.

C. remanei synonymous-site polymorphisms were characterized as encoding preferred–unpreferred variants (PU) if one variant yielded a preferred codon in the gene sequence and the other segregating variant produced an unpreferred codon, according to the *C. elegans* optimal codon table, as in Cutter and Charlesworth (2006). Similarly, unpreferred–unpreferred (UU), preferred–preferred (PP), and GC-AT variant sites were identified (211 total sites in 36 loci formed GC-AT variants at UU or PP sites). Using the frequencies of the preferred variant at PU sites, I applied a maximum-likelihood procedure (Cutter and Charlesworth 2006) based on the model of McVean and Charlesworth (1999) to infer the selection intensity ($\gamma = 4N_e s$, selection coefficient s) acting on preferred codons for each locus individually and for all loci considered together (a total of 529 PU sites in 40 loci). All calculations are based on polymorphism in the Ohio sample of strains. The implementation used here corrects an error in normalization of the likelihoods (B. Charlesworth, personal communication) that was discovered in the program used in Cutter and Charlesworth (2006); however, the effect is small, and the error was conservative with respect to detecting selection on codon usage. The statistical significance of the maximum-likelihood estimates for γ were inferred from overlap with $\gamma = 0$ of the $2\ln L$ interval. Correlations involving per-locus maximum-likelihood estimates of γ for the Ohio sample exclude *Cre-let-2*.

Simulations: Coalescent simulations using Hudson's (2002) program ms were implemented to test the distribution of Tajima's *D* for consistency with expectations under neutrality. Specifically, 10,000 replicates of 9 or 35 loci (see below) were generated using corresponding observed values of the number of segregating synonymous sites and sample size for the Ohio strains only (Table 1; supplemental Table 2) from which Tajima's *D* was calculated to compute the expected mean and variance in *D* for the loci under neutrality (using E. Stahl's program samplestats, http://molpopgen.org/software). The observed mean and variance in *D* was then compared to the simulated distributions. Different loci were assumed to be unlinked, and intralocus recombination was either excluded or included using observed population recombination parameter values of ρ_{LDhat} for corresponding loci. To reduce the potential for selection on synonymous sites to compromise the analysis, I restricted the empirical data set in each of two ways (in both cases also excluding one locus with evidence of selection). First, I considered polymorphic synonymous sites for the 9 loci that exhibited low overall codon bias ($F_{op} < 0.5$). Second, I considered only sites that defined preferred–preferred or unpreferred–unpreferred polymorphisms for the 35 loci containing such variants, because these variant sites are unlikely to experience direct selection.

I also performed a series of coalescent simulations to test for correlations between θ and Hudson's (2001) estimator of ρ (ρ_{H01}). Sets of 1000 neutral genealogies were simulated for each combination of fixed values of θ_{sim} (1, 5, 10, or 20), ρ_{sim} (1, 4, 10, or 16), f_{sim} (0, 1/2, 2, 4, or 8), and gene conversion tract length (100, 400, or 1600 bp). Gene conversion was incorporated in the simulation of neutral genealogies with the parameter f, the ratio of gene conversion to recombination rate (Wiuf and Hein 2000; Hudson 2002). Then, θ_s, ρ_{H01}, and ρ_{H01+f} were estimated from the resulting simulated neutral genealogies (using samplestats and R. Hudson's programs exhap and maxhap, http://home.uchicago.edu/~rhudson1). Estimates of θ and ρ (ρ_{H01}, ρ_{H01+f}) for a set of input values therefore varied solely due to stochasticity in the coalescent process and were computed along with Spearman-rank correlations between the θ and ρ estimates across each of the 1000 replicates. The resulting correlation coefficients were then evaluated in an ANOVA model as a function of the input values to determine the influence of each parameter, their second-order polynomials, and their first-order interactions on the correlation between estimated θ and ρ. A recursive partition analysis (JMP 5.0.1) was also applied to derive a heuristic interpretation of the effects of the dominant explanatory variables. This simulation regime tests for whether stochasticity in the coalescent process alone can generate correlations between estimates of θ and ρ, even when there is no underlying variation among loci in these two parameters.

RESULTS

Multilocus patterns of polymorphism and divergence: The 40 loci sampled here for 16 individuals, putatively linked to the *C. remanei* X chromosome, provided 953 polymorphic sites and yielded average nucleotide diversities at synonymous sites (π_s) of 3.6% and at nonsynonymous sites (π_a) of 0.1%. For the subsample of 14 strains from Ohio, the diversity values are comparable (Table 1). These mean values for 29.7 kb of sequence per sampled chromosome are somewhat lower than the averages previously reported for 9 loci (Cutter *et al.* 2006a), although the range is nearly identical. If a 4/3 correction for diversity estimates is appropriate for these loci, due to a reduction in effective population size caused by hemizygosity of the X chromosome in males, then the numerical average diversity value previously reported ($\pi_s = 4.7\%$) will hold, but additional data from autosomal loci are necessary to confirm this possibility. One locus demonstrated particularly high polymorphism (*Cre-D1005.1*; $\pi_s = 0.128$), confirming the high diversity observed previously for an adjacent portion of sequence in this gene (Cutter *et al.* 2006a).

As reported previously (Cutter *et al.* 2006a), linkage disequilibrium decays rapidly with distance in *C. remanei*. For these loci, the r^2 measure of linkage disequilibrium averages 0.19 within loci (40 loci) and 0.031 between loci (38 loci with complete data for Ohio sample), although 4 loci had intralocus r^2 values >0.4 (supplemental Table 2).

TABLE 1

Summary statistics for *C. remanei* diversity and the site-frequency spectrum, based on the Ohio sample, and for *C. remanei–C. briggsae* divergence

Locus	n	Sites	S_s	S_a	θ_s	θ_a	Tajima's D	d_S	d_S'	d_N	d_N/d_S'
Cre-C25F6.3	14	628	6	0	0.0124	0.00000	−1.29	0.792	0.888	0.0461	0.0518
Cre-dpy-8	14	659	11	0	0.0202	0.00000	1.57	1.129	1.094	0.0261	0.0239
Cre-C34F6.1	14	736	6	0	0.0117	0.00000	−1.50	1.112	0.847	0.0319	0.0377
Cre-col-175	14	650	13	4	0.0239	0.00264	−0.36	0.619	0.881	0.0561	0.0638
Cre-vit-2	14	815	17	7	0.0284	0.00352	−0.16	0.392	2.038	0.0930	0.0456
Cre-lam-2	13	859	19	1	0.0327	0.00048	−0.28	0.809	0.946	0.0563	0.0595
Cre-ncr-1	14	589	4	0	0.0094	0.00000	−0.24	2.797	2.354	0.0333	0.0141
Cre-slo-2	14	657	12	0	0.0251	0.00000	0.67	2.074	1.212	0.0124	0.0103
Cre-col-184	14	703	25	2	0.0411	0.00124	−0.71	0.458	0.705	0.0672	0.0953
Cre-pgp-12	14	705	42	9	0.0807	0.00588	−0.63	1.691	0.979	0.0464	0.0474
Cre-pgp-13	14	608	43	3	0.0928	0.00205	−1.41	1.661	1.154	0.0444	0.0385
Cre-pcca-1	14	610	20	3	0.0427	0.00206	−0.97	0.701	1.044	0.0408	0.0391
Cre-nmy-1	14	834	14	1	0.0247	0.00048	−0.52	1.216	0.975	0.0238	0.0244
Cre-glit-1	14	765	26	5	0.0467	0.00268	−0.41	4.763	3.778	0.0146	0.0039
Cre-asp-3	14	675	18	1	0.0368	0.00060	−1.30	0.443	1.628	0.0282	0.0173
Cre-spc-1	14	803	8	2	0.0150	0.00099	−1.88*	0.729	1.503	0.0134	0.0089
Cre-K10C2.1	14	616	21	7	0.0459	0.00470	−0.38	0.851	1.254	0.0469	0.0374
Cre-mbk-1	11	723	31	0	0.0614	0.00000	0.26	1.318	0.274	0.0125	0.0455
Cre-col-41	14	758	17	0	0.0276	0.00000	0.37	0.521	1.048	0.0416	0.0397
Cre-myo-2 exon 8	14	832	3	1	0.0057	0.00047	−1.28	0.314	1.150	0.0102	0.0089
Cre-T21B6.3	14	809	13	1	0.0215	0.00051	−0.35	0.512	0.220	0.0195	0.0883
Cre-lfi-1	14	679	28	3	0.0629	0.00175	0.52	0.697	0.660	0.0076	0.0116
Cre-col-19	14	627	12	2	0.0222	0.00138	−0.09	0.747	1.799	0.0695	0.0386
Cre-cht-1	14	778	34	5	0.0594	0.00263	−1.43	0.912	0.999	0.0123	0.0123
Cre-spc-1	14	748	6	0	0.0121	0.00000	−0.41	0.598	1.251	0.0111	0.0089
Cre-myo-2 exon 7	14	793	3	0	0.0056	0.00000	0.26	0.268	1.017	0.0127	0.0125
Cre-pgp-14	14	801	31	6	0.0525	0.00307	−0.72	1.082	1.439	0.0282	0.0196
Cre-Y102A11A.8	14	741	35	1	0.0654	0.00055	0.57	1.438	0.952	0.0829	0.0871
Cre-ifa-1	14	778	16	0	0.0303	0.00000	−0.05	0.684	1.412	0.0140	0.0099
Cre-dgn-1	14	766	17	2	0.0316	0.00106	1.72	1.109	1.893	0.0301	0.0159
Cre-F42D1.2.1	14	834	24	1	0.0382	0.00050	1.22	0.652	1.012	0.0291	0.0288
Cre-sym-4	14	705	33	3	0.0663	0.00173	−0.04	4.507	3.094	0.0289	0.0093
Cre-D1005.1	14	713	68	2	0.1227	0.00117	0.20	1.702	1.465	0.0381	0.0260
Cre-W07E11.1	14	820	5	1	0.0080	0.00050	−0.24	0.976	1.155	0.0457	0.0395
Cre-mrp-4	14	786	28	1	0.0468	0.00053	−1.09	1.661	1.276	0.0293	0.0229
Cre-F47A4.5	14	758	14	3	0.0248	0.00163	2.31*	2.246	1.256	0.0306	0.0243
Cre-let-2	14	757	3	0	0.0045	0.00000	−1.67	0.717	1.281	0.0165	0.0129
Cre-alg-1	14	807	31	0	0.0496	0.00000	−0.19	1.168	0.835	0.0012	0.0014
Cre-apa-2	14	726	26	0	0.0469	0.00000	−0.72	1.919	1.614	0.0040	0.0024
Cre-E01G6.1	14	1113	26	3	0.0314	0.00111	0.30	1.156	0.702	0.0398	0.0566
Average	13.9	744.1	20.2	2.0	0.0372	0.00115	−0.26	1.229	1.277	0.0324	0.0310

S_s, number of segregating synonymous sites; S_a, number of segregating replacement sites. *$P < 0.05$.

This polymorphism data set for a large collection of loci along the X chromosome is useful for testing for any influence of demographic perturbations in shaping genetic diversity in the sample. Demographic processes such as population size change or migration should skew the site-frequency spectrum for loci across the genome, whereas selection is likely to skew allele frequencies only in the vicinity of a target of selection. Across loci for the Ohio sample, synonymous sites yielded an average value of −0.259 for Tajima's D (Tajima 1989), a statistic that quantifies skew in the frequency spectrum of variant sites on the basis of alternative measures of nucleotide diversity (Table 1; Figure 1; mean D for the full sample = −0.322). For two loci individually, D differs significantly from the standard neutral expectation (*Cre-F47A4.5*, *Cre-spc-1*; Table 1) although the background skew makes the test for *Cre-spc-1* nonconservative; three other loci deviate significantly for the related metric Fu and Li's D^* (*Cre-K10C2.1*, *Cre-col-19*, *Cre-sym-4*; $D^* \sim 1.5$ for all three loci; supplemental Table 2). Only *Cre-F47A4.5* shows significance for more than one test of neutrality ($D = 2.31$, $P < 0.05$; $D^* = 1.49$, $P < 0.02$; $F^* = 1.96$, $P < 0.02$). *C. elegans spc-1* encodes a spectrin, which is involved in body morphogenesis (Norman and Moerman 2002),

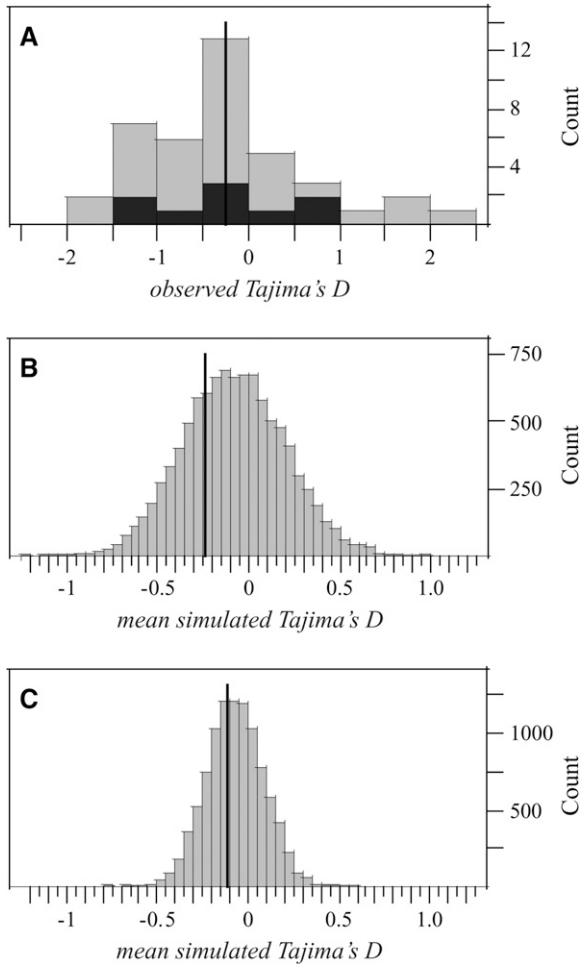

FIGURE 1.—Observed and 10,000 simulated distribution means of Tajima's *D* for the Ohio sample without recombination. In A, the 9 low-codon-bias loci used in simulations are solid bars; the vertical line indicates the observed mean $D = -0.230$. (B) Distribution of mean *D* based on all segregating synonymous sites for 9 loci with low codon bias ($F_{op} < 0.5$). The vertical line indicates the observed mean $D = -0.230$. (C) Distribution of mean *D* based on UU and PP polymorphic sites from only 35 loci. The vertical line indicates the observed mean $D = -0.113$.

and *col-19* encodes a collagen protein, which is associated with the cuticle (THEIN *et al.* 2003). The function of the *C. elegans* orthologs of the other genes is largely unknown.

To determine whether the distribution of *D*-values across the X chromosome is consistent with the distribution expected under a neutral scenario at mutation-drift equilibrium, I conducted coalescent simulations conditioned on the observed number of segregating synonymous sites and the sample size for each locus. Weak selection on synonymous sites can lead to negative *D* of a magnitude relevant to this study (MCVEAN and CHARLESWORTH 2000), which would complicate interpretations of a negative skew in the site-frequency spectrum. Selection on synonymous sites was detected in these data (see below; 70% of polymorphic synonymous sites form a PU site that is potentially subject to selection), although *D* does not correlate significantly with measures of codon bias ($D \times F_{op}$: $r_{SRC} = -0.12$, $P = 0.47$; D × γ: $r_{SRC} = -0.23$, $P = 0.15$; Ohio samples only; see below for details on codon bias). Consequently, I analyzed the data in two ways to limit the potential for selection at synonymous sites to impact inferences about demography.

First, I limited the tests for an overall departure in *D* to a set of nine loci with low codon bias ($F_{op} < 0.5$) and excluded *Cre-F47A4.5* due to the potential impact of selection on this locus; I also limited these analyses to the 14 strains from Ohio. This restricted sample exhibited mean $D = -0.230$. Neutral simulations based on the observed number of segregating sites and sample size, without recombination, indicate that a mean $D \leq -0.230$ is expected 31.9% of the time, suggesting no deviation from equilibrium in the Ohio sample on a timescale that can be detected with single nucleotide polymorphism (Figure 1). The variance in *D* (var[*D*]) also conforms well to the neutral expectation (25.2% of simulations showed a variance in *D* less than the observed 0.527). Two caveats are that *D* exhibits notoriously low power and is conservative under the assumption of no recombination (WALL 1999). When the simulations for the Ohio sample are repeated, including observed estimates of ρ_{LDhat}, virtually identical results are obtained (23.2% of simulations have $D \leq -0.230$).

Second, I calculated Tajima's *D* only for those synonymous sites that corresponded to preferred–preferred (PP) or unpreferred–unpreferred (UU) polymorphisms. Selection for codon bias should be negligible at such sites. Among the 35 loci in the Ohio sample that contained PP or UU polymorphisms (again excluding *Cre-F47A4.5*), Tajima's *D* averaged −0.113. Neutral coalescent simulations without recombination, conditioned on the number of segregating sites and the size of the sample, indicate that $D \leq -0.113$ is expected in 39.3% of cases (14.7% of simulations have var[*D*] less than or equal to the observed 0.702), which is indicative of no deviation from equilibrium (Figure 1). Simulations with recombination corroborate this result (31.5% of simulations with $D \leq -0.113$, 36.5% of simulations with var[*D*] ≤ 0.702). This approach has the advantage of considering most of the loci (35 of 40), but is limited by using a restricted set of polymorphic sites per locus. This contrasts with the first approach, which is limited in the number of loci considered (9 of 40), but which evaluates all polymorphic synonymous sites. The results of both approaches support the same conclusion that there is no evidence for a change in population size in the site-frequency spectrum for sites with little codon bias. The particularly high value of *D* for the gene of unknown function *Cre-F47A4.5* (2.33 entire sample, 2.31 Ohio only, both $P < 0.05$) therefore suggests that it can be

TABLE 2

Per-locus summaries of codon bias and of per-base-pair estimates of the population recombination parameter and gene conversion based on the Ohio sample

Locus	F_{op}	ENC	GC	γ_{PU} (2ln$L$ interval)	$\gamma_{UUPP\text{-}GC/AT}$	ρ_{LDhat}	ρ_{H01}	ρ_{H01+f}	f
Cre-C25F6.3	0.606	37.39	0.47	−0.16 (−2.5, 2.08)	< −10	0.00044	0.00062	0.00062	0
Cre-dpy-8	0.670	40.16	0.52	−0.13 (−1.74, 1.45)	0.90	0.00182	0.00222	0.00222	0
Cre-C34F6.1	0.513	42.39	0.45	0.49 (−1.69, 2.98)	ND	0.00000	0.00000	0.00000	0
Cre-col-175	0.743	37.16	0.55	0.67 (−1.18, 2.83)	1.16	0.00922	0.01614	0.00058	13.5
Cre-vit-2	0.891	25.53	0.53	2.19 (0.52, 4.75)	−0.10	0.01241	0.01778	0.00047	18.5
Cre-lam-2	0.603	37.22	0.47	0.26 (−1.01, 1.58)	0.24	0.01958	0.02612	0.02612	0
Cre-ncr-1	0.473	45.97	0.44	−0.64 (−5.5, 3.04)	−0.64	0.00000	0.00000	0.00000	0
Cre-slo-2	0.347	50.43	0.42	−0.52 (−2.33, 1.11)	1.23	0.00107	0.00000	0.00000	0
Cre-col-184	0.793	33.32	0.60	0.25 (−0.88, 1.43)	−1.45	0.00882	0.01789	0.01789	0
Cre-pgp-12	0.374	48.43	0.42	0.68 (−0.24, 1.68)	−0.15	0.02527	0.03953	0.00208	9.0
Cre-pgp-13	0.447	47.54	0.44	0.64 (−0.27, 1.62)	−1.35	0.00000	0.00300	0.00150	0.5
Cre-pcca-1	0.677	38.17	0.49	0.86 (−0.55, 2.49)	0.00	0.02155	0.03942	0.00232	8.0
Cre-nmy-1	0.535	48.07	0.46	0.37 (−1.37, 2.24)	−0.67	0.04084	0.08190	0.00292	13.5
Cre-glit-1	0.410	52.04	0.45	0.11 (−1.15, 1.39)	−2.78	0.03572	0.05000	0.01668	1.0
Cre-asp-3	0.816	28.80	0.52	0.81 (−0.59, 2.42)	−1.35	0.00000	0.00000	0.00000	0
Cre-spc-1	0.702	38.62	0.47	−1.27 (−3.58, 0.5)	ND	0.00000	0.00000	0.00000	0
Cre-K10C2.1	0.636	39.41	0.46	1.18 (−0.17, 2.82)	0.73	0.00000	0.00000	0.00000	0
Cre-mbk-1	0.413	56.65	0.46	0.42 (−0.88, 1.8)	−1.86	0.01079	0.01816	0.00052	17.0
Cre-col-41	0.790	34.79	0.57	1.16 (−0.37, 3.07)	−0.49	0.00860	0.01649	0.00046	17.5
Cre-myo-2 exon 8	0.767	34.44	0.51	4.86 (−0.14, >10)	ND	0.00000	ND	ND	ND
Cre-T21B6.3	0.632	34.41	0.52	0.37 (−1.16, 2.02)	3.16	0.00392	0.00448	0.00448	0
Cre-lfi-1	0.557	41.53	0.45	−0.23 (−1.53, 1.04)	−0.14	0.01194	0.01904	0.00060	15.5
Cre-col-19	0.877	33.95	0.60	0.77 (−0.77, 2.55)	1.43	0.00000	0.00000	0.00000	0
Cre-cht-1	0.594	37.19	0.46	0.15 (−0.77, 1.09)	1.49	0.00000	0.00197	0.00033	2.5
Cre-spc-1	0.697	33.86	0.48	1.06 (−1.16, 4.05)	< −10	0.00000	0.03151	0.03151	0
Cre-myo-2 exon 7	0.726	35.86	0.49	0.64 (−3.04, 5.5)	−1.43	0.00655	0.03276	0.03276	0
Cre-pgp-14	0.571	43.70	0.43	0.78 (−0.37, 2.07)	−0.23	0.00125	0.00648	0.00030	10.5
Cre-Y102A11A.8	0.376	43.09	0.41	0.49 (−0.5, 1.55)	−0.13	0.00778	0.01200	0.00036	16.0
Cre-ifa-1	0.723	33.60	0.49	1.45 (−0.2, 3.67)	−1.11	0.00334	0.00713	0.00022	16.0
Cre-dgn-1	0.740	37.11	0.50	1.04 (−0.36, 2.73)	−0.14	0.00613	0.00972	0.00972	0
Cre-F42D1.2.1	0.688	38.85	0.50	−0.09 (−1.26, 1.06)	−0.74	0.02375	0.03029	0.00101	14.5
Cre-sym-4	0.408	55.33	0.48	0.21 (−0.93, 1.38)	−1.57	0.00399	0.00505	0.00030	8.0
Cre-D1005.1	0.577	45.71	0.48	0.27 (−0.5, 1.06)	0.19	0.01839	0.02641	0.00075	17.0
Cre-W07E11.1	0.621	41.29	0.47	2.69 (−0.74, >10)	1.40	0.02397	0.04769	0.04769	0
Cre-mrp-4	0.393	44.88	0.41	0.63 (−0.77, 2.18)	0.00	0.00357	0.00612	0.00020	14.5
Cre-F47A4.5	0.414	50.41	0.45	0.14 (−1.54, 1.88)	−0.32	0.00626	0.00748	0.00748	0
Cre-let-2	0.810	36.38	0.58	>10 (0.19, >10)	ND	0.03516	ND	ND	ND
Cre-alg-1	0.585	44.87	0.49	0.52 (−0.65, 1.78)	−0.55	0.00950	0.01549	0.01549	0
Cre-apa-2	0.551	48.54	0.47	−0.30 (−1.61, 0.94)	0.37	0.00838	0.01298	0.01298	0
Cre-E01G6.1	0.522	41.43	0.47	−0.04 (−1.08, 1)	−2.53	0.00131	0.00225	0.00225	0
Average	0.607	40.96	0.483	0.44 (0.23, 0.65)[a]	−0.37 (−0.71, −0.04)[a]	0.00928	0.01600	0.00639	5.6

ND, no data.
[a] Combined maximum-likelihood estimates across all loci.

considered as a candidate for linkage to a target of balancing selection.

In calculations of pairwise divergence between homologous sequences in *C. remanei*, *C. briggsae*, and *C. elegans*, no significant differences were observed in mean d_N, d_S', or d_N/d_S' (supplemental Table 3). This accords with the findings of Cutter and Payseur (2003a). Focusing on the *C. remanei*–*C. briggsae* sequence comparison, divergence at nonsynonymous sites correlates positively with polymorphism at nonsynonymous sites in the Ohio sample ($d_N \times \pi_a$: $r_{SRC} = 0.41$, $P = 0.0086$; $d_N \times \theta_a$: $r_{SRC} = 0.41$, $P = 0.0087$), indicating that both diversity and divergence at nonsynonymous sites reflect a similar selective regime. Divergence at synonymous sites correlates positively with diversity at synonymous sites in the Ohio sample ($d_S \times \pi_s$: $r_{SRC} = 0.48$, $P = 0.0019$; $d_S \times \theta_s$: $r_{SRC} = 0.42$, $P = 0.0065$), but only when it is uncorrected for its association with codon bias and GC content ($P \geq 0.8$ for $d_S' \times \theta_s$). This supports the observation that synonymous sites are not strictly neutral (see below). Nonsynonymous and synonymous divergence, however, are not associated ($d_N \times d_S$: $r_{SRC} = -0.21$, $P = 0.18$; $d_N \times d_S'$:

$r_{SRC} = -0.02$, $P = 0.9$). The average nonsynonymous/synonymous ratio of polymorphism ($\theta_a/\theta_s = 0.034$) is similar to that of divergence ($d_N/d_S' = 0.031$), although the two are not correlated across loci ($P = 0.2$). Unfortunately, the saturated divergence (*i.e.*, mean $d_S > 1$) at synonymous sites between the *C. elegans*–*C. briggsae*–*C. remanei* species pairs rules out the application of divergence-corrected tests of selection, such as HKA and McDonald–Kreitman tests (Hudson *et al.* 1987; McDonald and Kreitman 1991).

Weak selection on codon usage: Selection on alternative degenerate codons is responsible for codon usage bias in *C. remanei*, as evidenced by several patterns. First, I quantify the selection intensity ($\gamma = 4N_e s$) on preferred codons to be significantly greater than zero overall among the 40 new loci assessed here for the Ohio sample ($\gamma = 0.44$, $2\ln L$ range = 0.23–0.65) and for two loci individually (*Cre-vit-2*, *Cre-let-2*; Table 2, supplemental Table 4). Similar results were obtained for the full sample (not shown). To the extent that Hill–Robertson interference (Hill and Robertson 1966) affects allele frequencies at these weakly selected sites, the strength of selection on preferred codons may be underestimated (McVean and Charlesworth 2000). Second, codon usage bias and synonymous site divergence are strongly negatively correlated in this sample of loci ($F_{op} \times d_S$ $r_{SRC} = -0.85$, $P < 0.0001$; Figure 2), as predicted for the action of selection at synonymous sites (Sharp and Li 1987). Although I have applied the *C. elegans* optimal codon table to *C. remanei*, this is a reasonable approach because *C. briggsae* uses an identical optimal codon table to *C. elegans* (Stein *et al.* 2003) and *C. elegans* is as divergent from *C. remanei* as it is from *C. briggsae*. In addition, codon bias is strongly correlated for these loci among all three species (r_{SRC} of F_{op} for *C. elegans*–*C. briggsae* = 0.94, for *C. elegans*–*C. remanei* = 0.96, and for *C. briggsae*–*C. remanei* = 0.95, all $P < 0.0001$). Consistent with the association between accumulated codon bias and divergence, contemporary selection on codon usage measured with γ correlates negatively with synonymous site divergence ($\gamma \times d_S$ $r_{SRC} = -0.45$, $P = 0.0039$). Furthermore, the per-locus contemporary estimates of selection on codon usage (γ) that are based on preferred–unpreferred codon polymorphisms correlate positively with codon bias that has accumulated over time and is reflected in the overall frequency of optimal codons in the sequences ($\gamma \times F_{op}$ $r_{SRC} = 0.46$, $P = 0.0034$; Figure 2). Finally, loci with strong codon bias exhibit reduced synonymous site diversity ($F_{op} \times \pi_s$: $r_{SRC} = -0.49$, $P = 0.0013$; $F_{op} \times \theta_s$: $r_{SRC} = -0.47$, $P = 0.0021$). All of these patterns are consistent with selection differentiating between alternative synonymous codons in *C. remanei*.

Because a positive correlation was observed between GC content and codon bias (GC × F_{op}: $r_{SRC} = 0.89$, $P < 0.0001$; GC × γ: $r_{SRC} = 0.37$, $P = 0.057$), the concern arises that biased gene conversion might be responsible for skewed frequencies of preferred variants (Marais 2003; Galtier *et al.* 2006). This is plausible because most preferred codons in Caenorhabditis are rich in guanine and cytosine (Stenico *et al.* 1994; Duret and Mouchiroud 1999). Consequently, I calculated γ for G or C variants at polymorphic sites that defined a GC-AT polymorphism for only those sites at which the corresponding alternatively encoded codons did not shift between a preferred and unpreferred designation (*i.e.*, for unpreferred–unpreferred and preferred–preferred variant sites, for which the selection differential is likely to be negligible). The maximum-likelihood estimate of γ on these G or C variants was −0.37 ($2\ln L$ interval: −0.71 to −0.04), a value that is of opposite sign to that expected for biased gene conversion [for the full data set, $\gamma = -0.30$ (−0.61–0.01)]. This result implies that biased gene conversion is not the driving force behind the skewed variant frequencies for preferred codons, consistent with previous results in this species (Cutter and Charlesworth 2006).

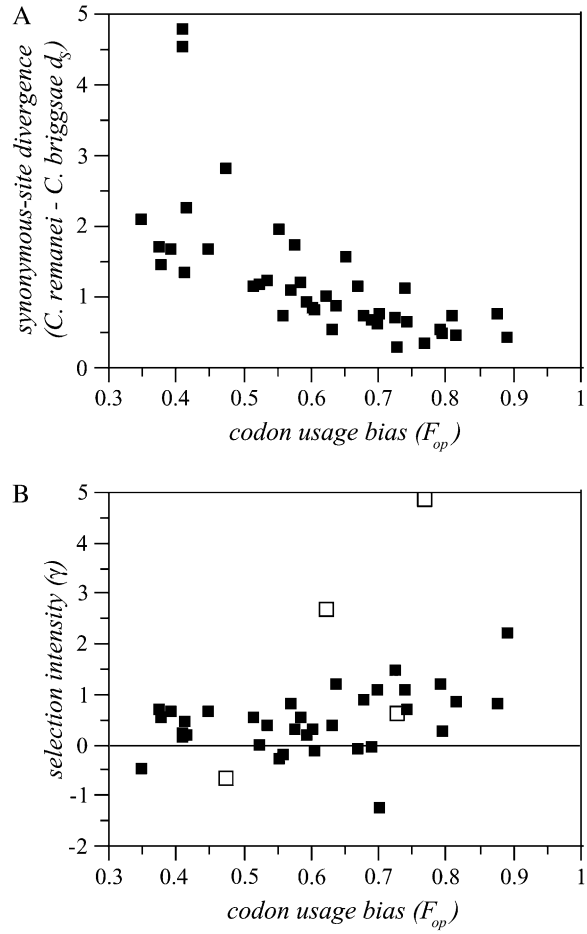

Figure 2.—Selection on codon usage. (A) Synonymous-site divergence between *C. remanei* and *C. briggsae* (d_S) is negatively correlated with the frequency of optimal codons ($r_{SRC} = -0.85$, $P < 0.0001$). (B) Contemporary selection on preferred codons ($\gamma = 4N_e s$) is stronger in loci with greater overall codon bias ($r_{SRC} = 0.46$, $P = 0.0034$; Ohio sample only, *Cre-let-2* excluded). Open squares correspond to loci with fewer than five PU sites used in maximum-likelihood estimation of γ.

Selection at linked sites: Empirical recombination rates are not available for *C. remanei*, so inferences about recombination currently must rely on inverse measures of linkage disequilibrium used to estimate the population recombination parameter, ρ. Overall, the ratio ρ/θ gives a measure of the number of recombination events per mutation; at neutral equilibrium, $\rho/\theta = 4N_e r/4N_e\mu = r/\mu$. In the Ohio sample, ρ_{LDhat}/θ_s has a median of 0.163 ($\rho_{H01}/\theta_s = 0.256$ and $\rho_{H01+f}/\theta_s = 0.020$), and few per-locus values exceed 1. These values are substantially lower than observed for *Drosophila melanogaster* and wild barley (Andolfatto and Przeworski 2000; Morrell *et al.* 2006) and might reflect the high mutation rate in Caenorhabditis (Denver *et al.* 2004).

When measured for the full data set, nucleotide diversity at synonymous sites (π_s and θ_s) correlates positively with the population recombination parameter in this sample ($\pi_s \times \rho_{LDhat}$: $r_{SRC} = 0.52$, $P = 0.0004$; $\theta_s \times \rho_{LDhat}$: $r_{SRC} = 0.52$, $P = 0.0005$; $\pi_s \times \rho_{H01}$: $r_{SRC} = 0.40$, $P = 0.013$; $\theta_s \times \rho_{H01}$: $r_{SRC} = 0.43$, $P = 0.0074$; Figure 3). These data also show no correlation of synonymous-site divergence with ρ or f ($d_S' \times \rho_{LDhat}$: $r_{SRC} = -0.13$, $P = 0.4$; $d_S' \times f$: $r_{SRC} = -0.02$, $P = 0.9$; similarly, nonsignificant results were obtained for d_S and other measures of ρ), which could occur if mutation rates were greater in regions of high recombination and also could cause a θ–ρ correlation. However, for the sample restricted to individuals from Ohio, which likely better represents a single population, a positive correlation between θ and ρ is no longer present ($r_{SRC} < 0.27$, $P > 0.1$ for all measures of θ and ρ). Furthermore, when the relative rate of gene conversion was estimated simultaneously with the population recombination parameter for the full sample, the θ × ρ correlation also disappeared ($\theta_s \times \rho_{H01+f}$: $r_{SRC} = 0.09$, $P = 0.6$; Figure 3). The relative incidence of gene conversion (f), when estimated simultaneously with ρ_{H01+f} for the Ohio sample, is observed to correlate positively with both ρ_{H01} and ρ_{LDhat}, but not with ρ_{H01+f} ($f \times \rho_{H01}$: $r_{SRC} = 0.40$, $P = 0.012$; $f \times \rho_{LDhat}$: $r_{SRC} = 0.45$, $P = 0.0042$; $f \times \rho_{H01+f}$: $r_{SRC} = -0.22$, $P = 0.19$). On the basis of the restricted Ohio sample, the average gene conversion parameter is inferred to be $f = 5.6$ (Table 2), which suggests that unbiased gene conversion is common. It will be important in the future to obtain empirical recombination rate estimates with which to contrast ρ, θ, and gene conversion.

The observation that simultaneous estimation of gene conversion and the population recombination parameter resulted in a reduced θ–ρ correlation for the full data set prompted me to investigate this issue further with coalescent simulations. I simulated neutral genealogies for each of 300 parameter combinations that controlled mutation, recombination, and gene conversion and then calculated the correlation between subsequently estimated values of θ and ρ_{H01} from the simulated data, which differed from the input values

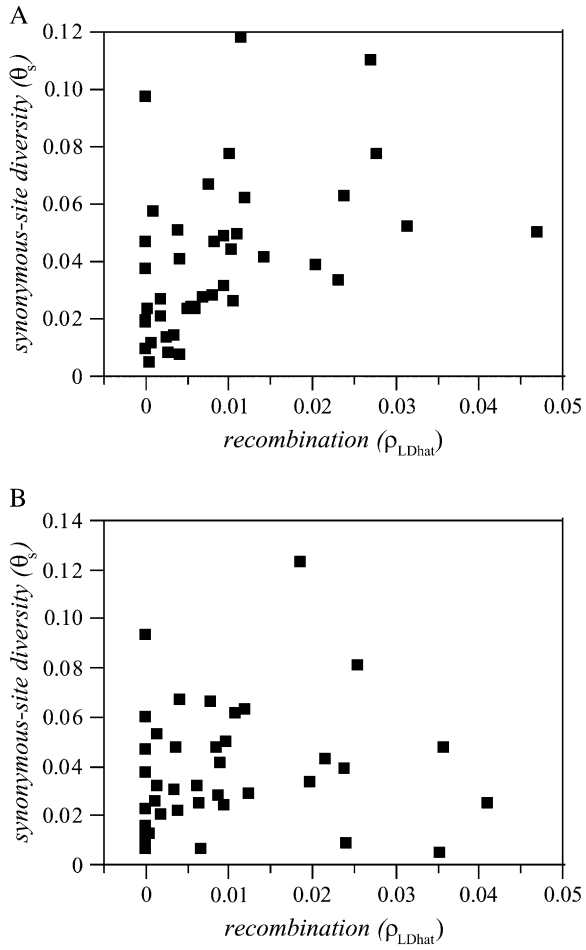

Figure 3.—Correlations between synonymous-site diversity and the population recombination parameter. (A) Diversity and ρ_{LDhat} are positively correlated for the full sample (Spearman-rank correlation = 0.52, $P = 0.0005$). (B) Diversity and ρ_{LDhat} do not correlate for the Ohio samples only (correlation = 0.21, $P = 0.2$).

due to stochasticity in the coalescent process (Figure 4). Each of the 300 correlation coefficients is based on 1000 paired estimates of θ and ρ. An ANOVA model describes 66.5% of the variation in correlation coefficients as a function of each of the input variables, their first-order interactions, and their second-order polynomials (Table 3). The most important factor contributing to variation in the $\theta \times \rho_{H01}$ correlation is the input value of ρ_{sim} (and its quadratic term), followed by the interaction between the input θ_{sim} and ρ_{sim}, the gene conversion parameter (f_{sim}) and its interaction with the input ρ_{sim}, and the interaction between input values of f_{sim} and θ_{sim} (Table 3). A partition analysis also illustrates how the combination of high recombination and high mutation results in a positive correlation between the estimated values of these two parameters, whereas low recombination, infrequent gene conversion, and a high mutation rate together yield negative correlations between estimates of θ and ρ (Figure 4). For the range of parameters considered here, the length of gene conversion tracts did not

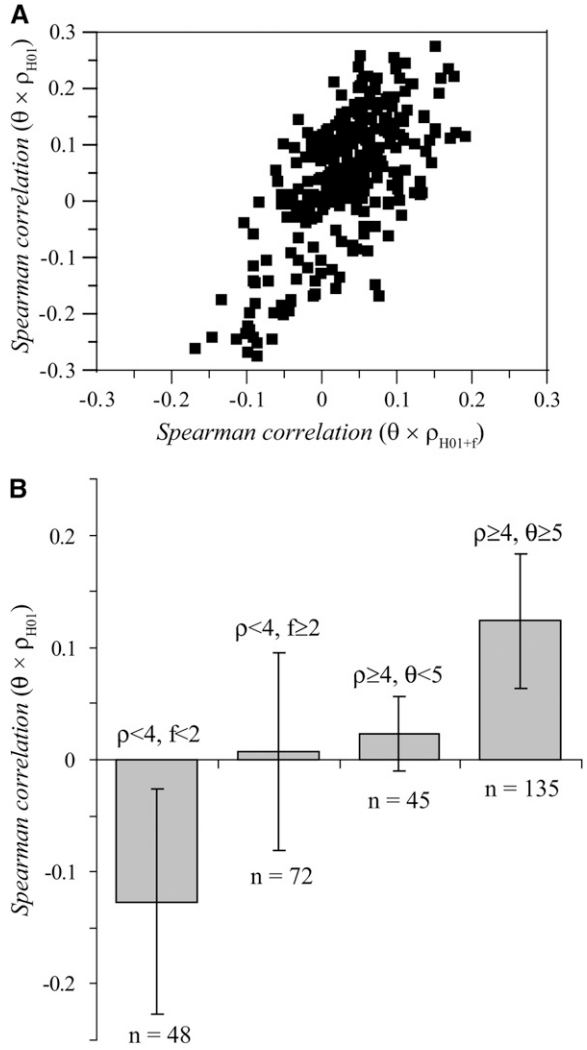

FIGURE 4.—Spearman-rank correlation coefficients between estimators of diversity and recombination. (A) Plot of correlations that are based on estimators of ρ that do or do not simultaneously infer the relative level of gene conversion. Each point represents the correlation between θ and ρ based on 1000 simulated genealogies for a different combination of input parameter values of θ_{sim} (1, 5, 10, or 20), ρ_{sim} (1, 4, 10, or 16), and f_{sim} (0, 1/2, 2, 4, or 8) and gene conversion tract length (100, 400, or 1600 bp). (B) Mean correlation (±1 SD) among simulations for the four categories identified by the first three bifurcations in a partition analysis, explaining 61% of the variation among correlation coefficient values.

strongly influence the observed correlations. Although correlations between θ and ρ_{H01+f} were observed in these simulations, the magnitudes were generally about half of those seen for θ × ρ_{H01} (Figure 4). These simulations show that this linkage-disequilibrium-based estimator of ρ (ρ_{H01}) can create spurious correlations between estimates of θ × ρ, particularly when the true values of both parameters are large, and even when there is no underlying variation among loci in the population mutation and recombination rates.

I also report that none of Tajima's *D*, F_{op}, γ, or GC content correlates significantly with any measure of ρ

TABLE 3

ANOVA results describing variation explained in estimated θ × ρ_{H01} Spearman-rank correlations from coalescent simulations (r^2_{adj}= 0.665, $F_{14,285}$ = 43.4, P < 0.0001)

Source	d.f.	SS	*F*	*P*	Sign of effect
ρ_{sim}	1	1.493	339.2	<0.0001	+
ρ_{sim}^2	1	0.410	93.0	<0.0001	−
f_{sim}	1	0.386	87.6	<0.0001	+
ρ_{sim} × f_{sim}	1	0.224	50.9	<0.0001	−
θ_{sim} × ρ_{sim}	1	0.215	48.7	<0.0001	+
θ_{sim}	1	0.083	18.8	<0.0001	+
f_{sim}^2	1	0.061	13.9	0.0002	−
θ_{sim} × f_{sim}	1	0.058	13.3	0.0003	+
ρ_{sim} × tract length	1	0.025	5.70	0.018	−
Tract length	1	0.015	3.34	0.069	+
f_{sim} × tract length	1	0.002	0.46	0.50	+
Tract length^2	1	0.002	0.44	0.51	−
θ_{sim} × tract length	1	0.000	0.036	0.85	+
θ_{sim}^2	1	0.000	0.003	0.95	+

(or gene conversion) in this sample. Such associations have previously been used in arguments for genetic hitchhiking (ANDOLFATTO and PRZEWORSKI 2001; STAJICH and HAHN 2005), Hill–Robertson interference (HEY and KLIMAN 2002), and a mutational influence of recombination (MARAIS 2003).

DISCUSSION

Nucleotide diversity is high across the *C. remanei* X chromosome, reflecting a large effective population size for this species, as also inferred from patterns of polymorphism in previously surveyed nuclear loci (GRAUSTEIN *et al.* 2002; JOVELIN *et al.* 2003; HAAG and ACKERMAN 2005; CUTTER *et al.* 2006a). In the population subsample from Ohio, there is no evidence of an overall departure from the neutral expectation in the site-frequency spectrum for putatively neutral sites, implying that panmixis is approximated in the sample. A lack of demographic change is a useful property for population genetic inference of natural selection because it is not necessary to account for complicated influences of population size change or structure in the history of the sample. Therefore, genome scans of selection using silent-site variant frequency spectra may be fruitful in *C. remanei*, provided that weak selection for codon bias is handled appropriately. Here I identify one locus (*Cre-F47A4.5*) that exhibits a significantly skewed frequency spectrum in a manner consistent with balancing selection, as measured by several statistics, motivating further analysis of this region.

I find that selection among alternative degenerate codons generates a robust pattern at polymorphic sites in which preferred codon variants are skewed toward

high frequency. This pattern is particularly evident among loci with strong overall biases in codon usage, indicating the concerted action of contemporary selection pressures shaping allele frequencies and the historical selection that generated fixed codons in genes. This result corroborates previous findings from a smaller sample of loci (Cutter and Charlesworth 2006). One unusual observation is that the selection intensities on preferred *vs.* unpreferred codons seem rather weak for selection on codon usage to have generated such dramatic effects on synonymous-site divergence (McVean and Charlesworth 1999). I speculate that the polymorphism-based method used here to estimate selection on alternative codons might be overly conservative relative to other approaches that permit polarization of changes relative to the ancestral state.

An association between crossover rate and nucleotide diversity is predicted by theory to be caused by background selection (Hudson and Kaplan 1995) or by genetic hitchhiking of neutral variants if selective sweeps are common (Wiehe and Stephan 1993). While supporting evidence for these processes comes from some species (Begun and Aquadro 1992; Andolfatto and Przeworski 2001; Cutter and Payseur 2003b), other species show no such correlation (Nordborg *et al.* 2005; Schmid *et al.* 2005; Wright *et al.* 2006), a weak or inconsistent correlation (Baudry *et al.* 2001; Tenaillon *et al.* 2002; Roselius *et al.* 2005), or a correlation that can be explained by neutral factors (Lercher and Hurst 2002; Hellmann *et al.* 2003). Here, I measured recombination using inverse-linkage disequilibrium estimators of the population recombination parameter ($\rho = 4N_e r$) and found that there is not a consistent significant association between diversity ($\theta = 4N_e\mu$) and ρ. The abundance of polymorphism per recombination unit (*i.e.*, low ρ/θ) suggests that the power to detect selection at linked sites should be high in *C. remanei*. However, empirical crossover rates will provide a more appropriate test for the potential of genetic hitchhiking and background selection to be important forces shaping patterns of genomic diversity (Maynard Smith and Haigh 1974; Charlesworth *et al.* 1993; Wiehe and Stephan 1993; Hudson and Kaplan 1995). It will be invaluable to obtain empirical recombination rate estimates for *C. remanei* to compare with observed levels of diversity so as to contrast with the population recombination parameter estimates (Andolfatto and Przeworski 2000; Przeworski and Wall 2001). In particular, this will be important to determine whether *C. remanei* mirrors the evidence of selection at linked sites found in *C. elegans* using map-based recombination rates (Cutter and Payseur 2003b). It is conceivable that self-fertilization in *C. elegans* reduces the effective recombination rate sufficiently such that the impact of selection at linked sites is easier to detect than for *C. remanei*, in which the window of linked polymorphism will be narrow due to extensive recombination and large effective population size. Therefore, regions of very low recombination in *C. remanei* might be required to detect a general effect of genetic hitchhiking and/or background selection in the form of an association between diversity and recombination.

In addition, I report that coalescent simulations indicate that spurious correlations between θ and linkage-disequilibrium-based estimators of ρ can be generated even under neutrality. Previous simulation work on estimators of the population recombination rate also showed that ρ tends to be overestimated when gene conversion is frequent or θ is high (Smith and Fearnhead 2005). When the relative strength of gene conversion (f) is estimated simultaneously with ρ using a composite maximum-likelihood method (Hudson 2001), the correlations are weaker, suggesting that estimators of ρ that do not account for gene conversion will be more likely to exhibit spurious associations with measures of diversity. Consequently, an artifactual effect of gene conversion on ρ estimators might best explain results such as those of Tenaillon *et al.* (2001, 2002), in which nucleotide diversity correlates with measures of recombination that are based on linkage disequilibrium but not the genetic map, rather than invoking selection or demography as causal factors.

The prevalence of gene conversion in *C. remanei*, as measured by the average ratio of gene conversion to recombination rate (mean $f = 5.6$), is comparable to or slightly higher than values reported in other species (Frisse *et al.* 2001; Padhukasahasram *et al.* 2004; Ptak *et al.* 2004; Morrell *et al.* 2006). In *C. elegans*, gene conversion has been observed to occur at a frequency of $\sim 10^{-5}$ in the 38-kb-long gene *unc-22* (Moerman and Baillie 1979) with tracts extending at least 191 bp (Plasterk and Groenen 1992). Crossover and gene conversion appear to be associated in a number of taxa (Borts and Haber 1987; Jeffreys and May 2004) and crossover rates do correlate with ρ in some species (Andolfatto and Przeworski 2000; Ptak *et al.* 2004). Andolfatto and Nordborg (1998) astutely detailed the important influence of gene conversion in breaking down linkage disequilibrium across short stretches of sequence and Langley *et al.* (2000) suggested that gene conversion might be an important piece of the diversity–recombination puzzle, but the potential role of gene conversion in generating a spurious correlation between diversity and ρ has not been reported previously.

I thank Scott Baird for sharing nematode strains and Stephen Wright and Bret Payseur for helpful discussions and comments on the manuscript. Three reviewers also provided particularly helpful comments. Some strains were made available by the Caenorhabditis Genetics Center. I also thank the Washington University School of Medicine Genome Sequencing Center for making *C. remanei* genome sequences publicly available. This research was supported by startup funds from the University of Toronto Department of Ecology and Evolutionary Biology and the Connaught Fund.

LITERATURE CITED

Andolfatto, P., and M. Nordborg, 1998 The effect of gene conversion on intralocus associations. Genetics **148:** 1397–1399.

Andolfatto, P., and M. Przeworski, 2000 A genome-wide departure from the standard neutral model in natural populations of Drosophila. Genetics **156:** 257–268.

Andolfatto, P., and M. Przeworski, 2001 Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. Genetics **158:** 657–665.

Barrière, A., and M. A. Félix, 2005 High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. Curr. Biol. **15:** 1176–1184.

Baudry, E., C. Kerdelhue, H. Innan and W. Stephan, 2001 Species and recombination effects on DNA variability in the tomato genus. Genetics **158:** 1725–1735.

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

Borts, R. H., and J. E. Haber, 1987 Meiotic recombination in yeast: alteration by multiple heterozygosities. Science **237:** 1459–1465.

Bulmer, M., 1991 The selection-mutation-drift theory of synonymous codon usage. Genetics **129:** 897–907.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

Cho, S., S. W. Jin, A. Cohen and R. E. Ellis, 2004 A phylogeny of Caenorhabditis reveals frequent loss of introns during nematode evolution. Genome Res. **14:** 1207–1220.

Cutter, A. D., 2006 Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*. Genetics **172:** 171–184.

Cutter, A. D., and B. Charlesworth, 2006 Selection intensity on preferred codons correlates with overall codon usage bias in *Caenorhabditis remanei*. Curr. Biol. **16:** 2053–2057.

Cutter, A. D., and B. A. Payseur, 2003a Rates of deleterious mutation and the evolution of sex in *Caenorhabditis*. J. Evol. Biol. **16:** 812–822.

Cutter, A. D., and B. A. Payseur, 2003b Selection at linked sites in the partial selfer *Caenorhabditis elegans*. Mol. Biol. Evol. **20:** 665–673.

Cutter, A. D., S. E. Baird and D. Charlesworth, 2006a High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. Genetics **174:** 901–913.

Cutter, A. D., M. A. Felix, A. Barriere and D. Charlesworth, 2006b Patterns of nucleotide polymorphism distinguish temperate and tropical wild isolates of *Caenorhabditis briggsae*. Genetics **173:** 2021–2031.

Cutter, A. D., J. Wasmuth and M. L. Blaxter, 2006c The evolution of biased codon and amino acid usage in nematode genomes. Mol. Biol. Evol. **23:** 2303–2315.

Denver, D. R., K. Morris, M. Lynch and W. K. Thomas, 2004 High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. Nature **430:** 679–682.

Duret, L., 2002 Evolution of synonymous codon usage in metazoans. Curr. Opin. Genet. Dev. **12:** 640–649.

Duret, L., and D. Mouchiroud, 1999 Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, Arabidopsis. Proc. Natl. Acad. Sci. USA **96:** 4482–4487.

Frisse, L., R. R. Hudson, A. Bartoszewicz, J. D. Wall, J. Donfack *et al.*, 2001 Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. Am. J. Hum. Genet. **69:** 831–843.

Fu, Y. X., and W. H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Galtier, N., E. Bazin and N. Bierne, 2006 GC-biased segregation of noncoding polymorphisms in Drosophila. Genetics **172:** 221–228.

Goldman, N., and Z. H. Yang, 1994 Codon-based model of nucleotide substitution for protein-coding DNA-sequences. Mol. Biol. Evol. **11:** 725–736.

Graustein, A., J. M. Gaspar, J. R. Walters and M. F. Palopoli, 2002 Levels of DNA polymorphism vary with mating system in the nematode genus Caenorhabditis. Genetics **161:** 99–107.

Haag, E. S., and A. D. Ackerman, 2005 Intraspecific variation in *fem-3* and *tra-2*, two rapidly coevolving nematode sex-determining genes. Gene **349:** 35–42.

Haag, E. S., H. Chamberlin, A. Coghlan, D. H. Fitch, A. D. Peters *et al.*, 2007 Caenorhabditis evolution: if they all look alike, you aren't looking hard enough. Trends Genet. **23:** 101–104.

Hellmann, I., I. Ebersberger, S. E. Ptak, S. Paabo and M. Przeworski, 2003 A neutral explanation for the correlation of diversity with recombination rates in humans. Am. J. Hum. Genet. **72:** 1527–1535.

Hey, J., and R. M. Kliman, 2002 Interactions between natural selection, recombination and gene density in the genes of Drosophila. Genetics **160:** 595–608.

Hey, J., and J. Wakeley, 1997 A coalescent estimator of the population recombination rate. Genetics **145:** 833–846.

Hill, W. G., and A. Robertson, 1966 Effect of linkage on limits to artificial selection. Genet. Res. **8:** 269–294.

Hillier, L. W., R. D. Miller, S. E. Baird, A. Chinwalla, L. A. Fulton *et al.*, 2007 Comparison of *C. elegans* and *C. briggsae* genome sequences reveals extensive conservation of chromosome organization and synteny. PLoS Biol. **5:** e167.

Hilliker, A. J., G. Harauz, A. G. Reaume, M. Gray, S. H. Clark *et al.*, 1994 Meiotic gene conversion tract length distribution within the *rosy* locus of *Drosophila melanogaster*. Genetics **137:** 1019–1026.

Hudson, R. R., 2001 Two-locus sampling distributions and their application. Genetics **159:** 1805–1817.

Hudson, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics **18:** 337–338.

Hudson, R. R., and N. L. Kaplan, 1995 Deleterious background selection with recombination. Genetics **141:** 1605–1617.

Hudson, R. R., M. Kreitman and M. Aguade, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

Ikemura, T., 1985 Codon usage and transfer-RNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2:** 13–34.

Jeffreys, A. J., and C. A. May, 2004 Intense and highly localized gene conversion activity in human meiotic crossover hot spots. Nat. Genet. **36:** 151–156.

Jovelin, R., B. C. Ajie and P. C. Phillips, 2003 Molecular evolution and quantitative variation for chemosensory behaviour in the nematode genus Caenorhabditis. Mol. Ecol. **12:** 1325–1337.

Kiontke, K., N. P. Gavin, Y. Raynes, C. Roehrig, F. Piano *et al.*, 2004 *Caenorhabditis* phylogeny predicts convergence of hermaphroditism and extensive intron loss. Proc. Natl. Acad. Sci. USA **101:** 9003–9008.

Langley, C. H., B. P. Lazzaro, W. Phillips, E. Heikkinen and J. M. Braverman, 2000 Linkage disequilibria and the site frequency spectra in the *su(s)* and *su(w(a))* regions of the *Drosophila melanogaster* X chromosome. Genetics **156:** 1837–1852.

Lercher, M. J., and L. D. Hurst, 2002 Human SNP variability and mutation rate are higher in regions of high recombination. Trends Genet. **18:** 337–340.

Li, H., A. Coghlan, J. Ruan, L. J. Coin, J.-K. Heriche *et al.*, 2006 TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Res. **34:** D572–D580.

Li, W. H., 1987 Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol. **24:** 337–345.

Marais, G., 2003 Biased gene conversion: implications for genome and sex evolution. Trends Genet. **19:** 330–338.

Maynard Smith, J., and J. Haigh, 1974 Hitch-hiking effect of a favorable gene. Genet. Res. **23:** 23–35.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in Drosophila. Nature **351:** 652–654.

McVean, G. A. T., and B. Charlesworth, 1999 A population genetic model for the evolution of synonymous codon usage: patterns and predictions. Genet. Res. **74:** 145–158.

McVean, G. A. T., and B. Charlesworth, 2000 The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics **155:** 929–944.

McVean, G., P. Awadalla and P. Fearnhead, 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics **160:** 1231–1241.

Merkl, R., 2003 A survey of codon and amino acid frequency bias in microbial genomes focusing on translational efficiency. J. Mol. Evol. **57:** 453–466.

Moerman, D. G., and D. L. Baillie, 1979 Genetic organization in *Caenorhabditis elegans*: fine-structure analysis of the *unc-22* gene. Genetics **91:** 95–103.

Morrell, P. L., D. M. Toleno, K. E. Lundy and M. T. Clegg, 2006 Estimating the contribution of mutation, recombination and gene conversion in the generation of haplotypic diversity. Genetics **173:** 1705–1723.

Nei, M., and W. H. Li, 1979 Mathematical-model for studying genetic-variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76:** 5269–5273.

Nordborg, M., T. T. Hu, Y. Ishino, J. Jhaveri, C. Toomajian *et al.*, 2005 The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol. **3:** e196.

Norman, K. R., and D. G. Moerman, 2002 Alpha spectrin is essential for morphogenesis and body wall muscle formation in *Caenorhabditis elegans*. J. Cell Biol. **157:** 665–677.

Padhukasahasram, B., P. Marjoram and M. Nordborg, 2004 Estimating the rate of gene conversion on human chromosome 21. Am. J. Hum. Genet. **75:** 386–397.

Plasterk, R. H., and J. T. Groenen, 1992 Targeted alterations of the *Caenorhabditis elegans* genome by transgene instructed DNA double strand break repair following Tc1 excision. EMBO J. **11:** 287–290.

Przeworski, M., and J. D. Wall, 2001 Why is there so little intragenic linkage disequilibrium in humans? Genet. Res. **77:** 143–151.

Ptak, S. E., K. Voelpel and M. Przeworski, 2004 Insights into recombination from patterns of linkage disequilibrium in humans. Genetics **167:** 387–397.

Roselius, K., W. Stephan and T. Stadler, 2005 The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. Genetics **171:** 753–763.

Schmid, K. J., S. Ramos-Onsins, H. Ringys-Beckstein, B. Weisshaar and T. Mitchell-Olds, 2005 A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. Genetics **169:** 1601–1615.

Sharp, P. M., and W. H. Li, 1987 The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. **4:** 222–230.

Sivasundar, A., and J. Hey, 2003 Population genetics of *Caenorhabditis elegans*: the paradox of low polymorphism in a widespread species. Genetics **163:** 147–157.

Smith, N. G., and P. Fearnhead, 2005 A comparison of three estimators of the population-scaled recombination rate: accuracy and robustness. Genetics **171:** 2051–2062.

Stajich, J. E., and M. W. Hahn, 2005 Disentangling the effects of demography and selection in human history. Mol. Biol. Evol. **22:** 63–73.

Stein, L. D., Z. Bao, D. Blasiar, T. Blumenthal, M. R. Brent *et al.*, 2003 The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. **1:** 166–192.

Stenico, M., A. T. Lloyd and P. M. Sharp, 1994 Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. Nucleic Acids Res. **22:** 2437–2446.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Tenaillon, M. I., M. C. Sawkins, A. D. Long, R. L. Gaut, J. F. Doebley *et al.*, 2001 Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). Proc. Natl. Acad. Sci. USA **98:** 9161–9166.

Tenaillon, M. I., M. C. Sawkins, L. K. Anderson, S. M. Stack, J. Doebley *et al.*, 2002 Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Genetics **162:** 1401–1413.

Thein, M. C., G. McCormack, A. D. Winter, I. L. Johnstone, C. B. Shoemaker *et al.*, 2003 *Caenorhabditis elegans* exoskeleton collagen COL-19: an adult-specific marker for collagen modification and assembly, and the analysis of organismal morphology. Dev. Dyn. **226:** 523–539.

Wall, J. D., 1999 Recombination and the power of statistical tests of neutrality. Genet. Res. **74:** 65–79.

Watterson, G. A., 1975 Number of segregating sites in genetic models without recombination. Theor. Popul. Biol. **7:** 256–276.

Wiehe, T. H. E., and W. Stephan, 1993 Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 842–854.

Wiuf, C., and J. Hein, 2000 The coalescent with gene conversion. Genetics **155:** 451–462.

Wright, S. I., J. P. Foxe, L. DeRose-Wilson, A. Kawabe, M. Looseley *et al.*, 2006 Testing for effects of recombination rate on nucleotide diversity in natural populations of *Arabidopsis lyrata*. Genetics **174:** 1421–1430.

Communicating editor: D. Begun