

# High Nucleotide Polymorphism and Rapid Decay of Linkage Disequilibrium in Wild Populations of *Caenorhabditis remanei*

Asher D. Cutter,<sup>\*,1</sup> Scott E. Baird<sup>†</sup> and Deborah Charlesworth<sup>\*</sup>

<sup>\*</sup>*Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom and*

<sup>†</sup>*Department of Biological Sciences, Wright State University, Dayton, Ohio 45431*

Manuscript received June 12, 2006  
Accepted for publication August 10, 2006

## ABSTRACT

The common ancestor of the self-fertilizing nematodes *Caenorhabditis elegans* and *C. briggsae* must have reproduced by obligate outcrossing, like most species in this genus. However, we have only a limited understanding about how genetic variation is patterned in such male–female (gonochoristic) *Caenorhabditis* species. Here, we report results from surveying nucleotide variation of six nuclear loci in a broad geographic sample of wild isolates of the gonochoristic *C. remanei*. We find high levels of diversity in this species, with silent-site diversity averaging 4.7%, implying an effective population size close to 1 million. Additionally, the pattern of polymorphisms reveals little evidence for population structure or deviation from neutral expectations, suggesting that the sampled *C. remanei* populations approximate panmixis and demographic equilibrium. Combined with the observation that linkage disequilibrium between pairs of polymorphic sites decays rapidly with distance, this suggests that *C. remanei* will provide an excellent system for identifying the genetic targets of natural selection from deviant patterns of polymorphism and linkage disequilibrium. The patterns revealed in this obligately outcrossing species may provide a useful model of the evolutionary circumstances in *C. elegans*' gonochoristic progenitor. This will be especially important if self-fertilization evolved recently in *C. elegans* history, because most of the evolutionary time separating *C. elegans* from its known relatives would have occurred in a state of obligate outcrossing.

THE combined actions of the population processes of mutation, drift, recombination, demography, and natural selection mold the genomic landscape of genetic variation, leaving signatures that can be detected from patterns of polymorphism in natural populations. To understand the relative importance of these forces in shaping genome evolution, it is crucial to characterize natural genetic variation within and between species. The genus *Caenorhabditis*, a group of bacterivorous nematodes, is rapidly being developed as a model for population genetics, motivated by the genetic model *Caenorhabditis elegans* and by genome sequencing efforts in related members of this group (STEIN *et al.* 2003; FÉLIX 2004). Here, we explore how evolutionary forces shape patterns of nucleotide diversity in natural samples of the obligately outbreeding species *C. remanei*.

Most nematode species, including *C. remanei*, have a gonochoristic (male–female) breeding system. However, establishment of the androdioecious (male–hermaphrodite) *C. elegans* as a biological model has resulted in a research emphasis on the two self-fertile

species in this genus (*C. elegans* and *C. briggsae*), which are thought to have evolved independently from gonochoristic ancestors (KIONTKE *et al.* 2004). Among known species of *Caenorhabditis*, *C. remanei* is related most closely to the androdioecious *C. briggsae* and gonochoristic *C. sp. 5* (JU727) (BRAENDLE and FELIX 2006; KIONTKE and SUDHAUS 2006), sharing a common ancestor many millions of years ago (COGHLAN and WOLFE 2002; STEIN *et al.* 2003; FÉLIX 2004; CUTTER *et al.* 2006). *C. elegans*, as well as the gonochoristic *C. sp. 4* (CB5161) and *C. japonica*, are more distant relatives of *C. remanei*, although all have very similar morphology (FITCH and EMMONS 1995; KIONTKE *et al.* 2004). Despite the morphological parallels among species, they differ greatly at the sequence level (STEIN *et al.* 2003; CHO *et al.* 2004) and no fertile interspecific hybrids between pairwise crosses of *C. elegans*, *C. briggsae*, and *C. remanei* have been observed (BAIRD *et al.* 1992; BAIRD and YEN 2000; HILL and L'HERNAULT 2001; BAIRD 2002).

As *Caenorhabditis* develops further as a model genus for studying comparative evolution, it becomes more urgent to acquire a robust understanding of the gonochoristic species that share common ancestors with the progenitors of these self-fertile species. The consequences of an outcrossing ancestry will be particularly important if hermaphroditism arose recently in the history of *C. elegans* and/or *C. briggsae* (CUTTER *et al.*

<sup>1</sup>Corresponding author: Department of Ecology and Evolutionary Biology, University of Toronto, 25 Harbord St., Toronto, ON M5S 3G5, Canada. E-mail: asher.cutter@utoronto.ca

2006), because most of the evolutionary time separating them from their relatives would have occurred in a gonochoristic state. For example, indications of sexual selection, sex differences in longevity, and selection for codon usage bias in a recent gonochoristic ancestor may persist in selfing descendant species, provided that the origin of selfing is sufficiently recent (McCulloch and GEMS 2003; CUTTER and WARD 2005; CUTTER 2006). *Caenorhabditis* sex determination genes, which might have been involved in breeding system changes in these species, appear to evolve quickly (CLIFFORD *et al.* 2000; HAAG *et al.* 2002; STOTHARD *et al.* 2002; NAYAK *et al.* 2005). Furthermore, female *C. remanei* continue to attract strongly the males of other species (CHASNOV and CHOW 2002) and heterospecific major sperm protein can trigger ovulation (HILL and L'HERNAULT 2001; MILLER *et al.* 2001). However, sufficient time has elapsed in the selfing species to have led to reduced attractiveness of hermaphrodites to males, reduced mating vigor of males, and smaller male sperm relative to gonochoristic species (LAMUNYON and WARD 1999; HILL and L'HERNAULT 2001; CHASNOV and CHOW 2002), either through relaxed selection or through selection disfavoring these traits.

*C. remanei* has been isolated from localities in Europe, North America, and Asia and is found frequently in phoretic associations with invertebrates (principally in the “enduring” dauer larval stage; BAIRD 1999; KIONTKE and SUDHAUS 2006; M. A. FÉLIX, personal communication). Previous work in *C. remanei* has demonstrated that specieswide levels of diversity are substantially greater than those in its self-fertile relatives (GRAUSTEIN *et al.* 2002; JOVELIN *et al.* 2003; HAAG and ACKERMAN 2005). However, it is not yet clear how different regions of the genome vary in their levels of nucleotide diversity, how the genetic variation is distributed across the species' geographic range, or the distance to which linkage disequilibrium extends between pairs of polymorphic sites in the genome. The answers to these questions depend on the extent of demographic processes affecting the whole genome, such as population subdivision and growth, as well as the locus-specific action of natural selection and its interaction with recombination rates and variation in mutation rate among chromosomal regions.

In this study, we evaluate these issues in a sampling of *C. remanei* by surveying nucleotide diversity at six nuclear loci. The pattern of polymorphisms reveals little evidence for population structure or deviation from neutral expectations, indicating that *C. remanei* has a large effective population size and that these samples may approximate panmixis and demographic equilibrium. In combination with the rapid decay of linkage disequilibrium with distance, this suggests that screens for deviations from neutral patterns of polymorphism may prove particularly successful in identifying targets of natural selection in this species.

## MATERIALS AND METHODS

**Nematode populations:** We studied 34 isolates of *C. remanei*, including 6 obtained from the *Caenorhabditis* Genetics Center and 1 (JU825) gifted by M.-A. Félix (Table 1). Several of the strains had been inbred for numerous generations (SB146, PB4641, VT733, CR1014, CR1415, and CR2124), whereas the others are each simple population propagations from the progeny of a single isolated female individual. Most of the strains derive from collections of isopod wood lice under logs in the Wright State University woods from 1998 and 2000. *C. remanei* in Ohio were found in association with the wood lice species *Trachelipus rathkii*, *Porcellio scaber*, *Armadillidium nasatum*, *Oxychilus alliarius*, *Oniscus asellus*, and *Cylisticus convexus* (BAIRD 1999). Three individuals were isolated from the same firefly larva from the Achilles Hill site.

**Molecular methods:** A NaOH digestion protocol (FLOYD *et al.* 2002) was used to isolate DNA from single male individuals for each of the 34 strains of *C. remanei*. We selected for resequencing in these strains the putative orthologs of the six genes from *C. elegans* chromosomes II and X that were assessed for polymorphism in CUTTER (2006) (Table 2). However, the *C. remanei* regions sequenced differ from those analyzed in *C. elegans* (CUTTER 2006). Specifically, because of the high diversity in *C. remanei*, in the present study we primarily sequenced exonic DNA, rather than long introns. These fragments cover roughly a quarter of the coding sequence of the corresponding genes. We use an arbitrary internal naming scheme to refer to these gene fragments (*e.g.*, “p15”). Some individuals were excluded from the analysis of a given locus, due either to failure to amplify or to heterozygosity in sequence traces; only the subset of 27 individuals with complete data at all loci were used for interlocus analyses (Table 1). Both strands were sequenced directly from PCR products, following cleanup with exonuclease I and shrimp alkaline phosphatase, on an ABI 3730 automated sequencer by the University of Edinburgh Sequencing Facility. Sequence data from this article have been deposited in GenBank under the following accession numbers: p15, DQ897401–DQ897433; p17, DQ897434–DQ897465; p18, DQ897466–DQ897499; p21, DQ897500–DQ897532; p22, DQ897533–DQ897565; and p24, DQ897370–DQ897400.

**Sequence analysis:** Sequencher v. 4.0 and BioEdit were used for sequence alignment and manual editing to confirm sequence quality and to remove primer sequences. Calculations of diversity from pairwise differences ( $\pi$ ) and from the number of segregating sites ( $\theta$ ) (WATTERSON 1975; NEI and LI 1979), linkage disequilibrium, recombination, and tests of neutrality were performed on each locus using DnaSP v.4.10.04 (ROZAS *et al.* 2003), LDhat (FEARNHEAD and DONNELLY 2001; McVEAN *et al.* 2002), and SplitsTree4 (BRUEN *et al.* 2006; HUDSON and BRYANT 2006). We distinguish between diversity at silent sites (synonymous positions,  $\pi_s$  and  $\theta_s$ ; noncoding intronic positions,  $\pi_{nc}$  and  $\theta_{nc}$ ; synonymous and noncoding positions,  $\pi_{si}$  and  $\theta_{si}$ ) and at nonsynonymous sites ( $\pi_a$  and  $\theta_a$ ). Tests of neutrality, *e.g.*, Tajima's *D* (TAJIMA 1989b), were conducted with silent sites, although values are similar when all sites are considered together. We also performed coalescent simulations with recombination in DnaSP to infer the neutral distributions of test statistics, conditioning on the observed number of segregating sites in the sample. Although we report the standard  $F_{st}$  measure of differentiation between populations (HUDSON *et al.* 1992b), we also tested for population structure with  $K_{st}^*$  in DnaSP because this statistic performs well for small sample sizes coupled with high diversity (HUDSON *et al.* 1992a; HUDSON 2000), as in our data. Note that  $F_{st}$  is based on frequencies of polymorphic sites, assuming independence, whereas  $K_{st}^*$  uses information about pairwise differences

**TABLE 1**  
**Sequenced *C. remanei* samples**

Strain	Locality	State	Country
CR1014	Gloucester	Massachusetts	United States
CR1415	Gloucester	Massachusetts	United States
CR2124	Gloucester	Massachusetts	United States
PB207	Wright St. U. woods	Ohio	United States
PB209	Wright St. U. woods	Ohio	United States
PB210 <sup>a</sup>	Wright St. U. woods	Ohio	United States
PB211	Wright St. U. woods	Ohio	United States
PB212	Wright St. U. woods	Ohio	United States
PB213	Wright St. U. woods	Ohio	United States
PB215	Wright St. U. woods	Ohio	United States
PB219 <sup>b</sup>	Wright St. U. woods	Ohio	United States
PB242	Wright St. U. woods	Ohio	United States
PB243	Wright St. U. woods	Ohio	United States
PB244	Wright St. U. woods	Ohio	United States
PB245 <sup>c</sup>	Wright St. U. woods	Ohio	United States
PB247	Wright St. U. woods	Ohio	United States
PB249 <sup>a</sup>	Wright St. U. woods	Ohio	United States
PB252	Wright St. U. woods	Ohio	United States
PB253	Wright St. U. woods	Ohio	United States
PB275 <sup>d,e</sup>	Wright St. U. woods	Ohio	United States
PB277	Achilles Hill	Ohio	United States
PB278	Achilles Hill	Ohio	United States
PB279 <sup>d</sup>	Achilles Hill	Ohio	United States
PB228	Eastwood Park	Ohio	United States
PB229	Englewood Park	Ohio	United States
PB230	Indian Lake St. Park	Ohio	United States
PB231	Hueston Woods State Park	Ohio	United States
PB232	Punderson State Park	Ohio	United States
PB233	Punderson State Park	Ohio	United States
PB276 <sup>a</sup>	Bloomington	Indiana	United States
VT733	New Haven	Connecticut	United States
PB4641	Brooklyn	New York	United States
JU825	Obernai		France
SB146	Freiburg		Germany

Wright St. U., Wright State University.

<sup>a</sup> Missing for *p24*.

<sup>b</sup> Missing for *p22*.

<sup>c</sup> Missing for *p15*.

<sup>d</sup> Missing for *p17*.

<sup>e</sup> Missing for *p21*.

between sequences. Approximate confidence intervals for  $F_{st}$  and  $K_{st}^*$  were obtained for each locus by resampling random haplotypes with replacement from each putative subpopulation and calculating  $F_{st}$  and  $K_{st}^*$  with a modified version of HUDSON's (2002) samplestats program (kindly provided by E. STAHL, personal communication) or permtest (HUDSON 2000) to derive the distribution of these statistics. Only sites with single-nucleotide polymorphisms were used in this procedure. Structure 2.0 was also used to evaluate population subdivision (PRITCHARD *et al.* 2000). We implemented Structure's linkage model with correlated allele frequencies (given  $\lambda = 0.862$ , which was estimated by Structure for  $K = 1$  and fixed at this value for other  $K$ ) and assigned a genetic distance of  $10^{-7}$  to adjacent sites within a locus and free recombination among loci (burn-in = 50,000; run length = 100,000; three replicates for each  $1 \leq K \leq 5$ ). Analyses of interlocus linkage disequilibrium and population subdivision

with Structure were limited to the 27 individuals with complete data for all loci and excluded indels.

The expected decay of linkage disequilibrium ( $r^2$ ) with distance was modeled for each locus according to Equation 3 of WEIR and HILL (1986),

$$E(r^2) = \frac{10 + \Gamma}{22 + 13\Gamma + \Gamma^2} \times \left[ 1 + \frac{(3 + \Gamma)}{n(22 + 13\Gamma + \Gamma^2)^2} (12 + 12\Gamma + \Gamma^2) \right],$$

where  $n$  is the sample size and  $\Gamma$  is the product of the population recombination parameter ( $\rho = 4N_e r$ ) and distance ( $d$ ). Values of the population recombination parameter inferred from LDhat and from least-squares fitting of the above equation to data resulted in similar estimates (not shown). We also fit this formula to the sample of 230 SNPs

**TABLE 2**  
**Loci sequenced in *C. remanei***

<i>C. remanei</i> locus	<i>C. elegans</i> ortholog	<i>C. briggsae</i> ortholog	Forward primer	Reverse primer	Product size
<i>p15</i>	Y25C1A.5 (II)	CBG19635	CATCGTTCTCGACGTGCTAA	GATCGCATTGCTGAACTCAA	781
<i>p24i</i>	ZK430.1 (II)	CBG03684	AAACCGTGTGCAATTCCTTG	ACGATCGAAAAGAGTGCCAGT	824
<i>p24ii</i>	ZK430.1 (II)	CBG03684	CGTCACTCGTCTCCACAAGA	CGATGCTCTCAACAAGCAAA	881
<i>p17</i>	E01G4.6 (II)	CBG20775	GAGTTCATTCTTCCGTCCA	CCTTTTCAGCTGCATTGACA	868
<i>p18</i>	R160.7 (X)	CBG14460	CGATGAGTGACTTCCGAACA	ACCAGCAATGCACACGAATA	798
<i>p21</i>	T24D11.1 (X)	CBG16368	GTCCGTATGGGGGAGAACTCA	GAGATCCGCATCGTTTTTCAT	822
<i>p22</i>	D1005.1 (X)	CBG24512	AGCGATCTCCTCCATAGCA	AAGCCAATGCCAACTGTACC	877

from 11 strains of *C. elegans* (KOCH *et al.* 2000), except that recombination distances (centimorgans) were used and “ $\rho$ ” was inferred by least-squares regression. For this analysis, pairs of sites from different chromosomes were assigned a distance of 50 cM.

We calculated a measure of codon bias, the frequency of optimal codons,  $F_{op}$ , with CodonW (J. Peden, <http://www.sourceforge.net/codonw>), given the *C. elegans* codon usage table (IKEMURA 1985; STENICO *et al.* 1994). We computed  $F_{op}$  using only the portions of the coding regions that were sequenced for polymorphism analysis, from which we report the maximum value observed among the sampled sequences. Likewise, we used the codeml program in Paml (YANG 1997) to compute Nei–Gojobori estimates of synonymous ( $d_s$ ) and nonsynonymous ( $d_N$ ) substitution rates of the resequenced portions of the coding regions, using SB146 relative to *C. briggsae* orthologs (<http://www.wormbase.org>). We aligned the *C. remanei* and *C. briggsae* sequences using canonical peptide translations with Clustal’s default settings in BioEdit, with manual adjustment as necessary. Neighbor-joining trees and neighbor networks were constructed with SplitsTree4 (HUSON and BRYANT 2006).

## RESULTS

**Patterns of nucleotide polymorphism:** Across six loci comprising 3950 bp of sequence in 34 *C. remanei* individuals, we identified very high levels of nucleotide polymorphism (Table 3). Silent-site diversity ( $\pi_{si}$ ) for the six loci ranged from 1.5 to 11.8%, with an average of 5.7% (1252 bp in total, 761 bp of which are coding), and diversity at nonsynonymous sites ( $\pi_a$ ) averaged 0.33% (2611 bp in total), ranging from 0 to 0.64% (Table 3). Two loci (*p22*, which is probably X-linked, as explained below, and the putatively autosomal *p24*) were particularly polymorphic, with silent-site diversities of  $\sim 10\%$  (Table 3). For these loci, individual pairs of haplotypes can be up to  $K_s = 25\%$  divergent at silent sites. This high diversity is not an artifact due to alternative gene duplicates being sequenced, as evidenced by the extensive recombination detected within these polymorphic loci, the clear decline of linkage disequilibrium with distance, the lack of clearly divergent sets of haplotypes, and their unimodal distributions of pairwise allelic divergence. Exclusion of these two loci lowers the average diversity of the remaining four loci to  $\pi_{si} = 3.1\%$ .

Diversity at synonymous sites does not differ significantly from that at noncoding sites ( $P > 0.5$ ; Table 3). Sixteen polymorphic sites across four loci exhibited more than two segregating bases, and insertion/deletion (indel) polymorphisms were present in five of the six loci (including an in-frame 3-bp coding region indel polymorphism in *p17*). Among the 252 polymorphic sites with two allelic states, there are 158 transitions (ts) and 94 transversions (tv), yielding a ts/tv ratio of 1.68. We also note that strain PB4641, the strain used for full-genome sequencing, contains a C to T transition in the *p17* locus that generates a predicted premature stop codon (TAA).

Residual heterozygosity in the isofemale lines was detected in some male individuals for the three loci expected to be autosomal on the basis of the locations of their orthologs in *C. elegans* (*p15*, *p24*, and *p17*). The three orthologs of *C. elegans* X-linked genes (*p18*, *p22*, and *p21*), however, never exhibited residual heterozygosity in male *C. remanei*, consistent with X-linkage in this species as well. This is also supporting evidence that these are single-copy genes. We observed no difference between the levels of genetic variation of the three putatively autosomal loci (*p15*, *p24*, and *p17*) relative to the three putatively X-linked loci, *p18*, *p22*, and *p21* (the mean for both sets of loci is  $\pi_{si} = 0.057$ , so a  $4/3 N_e X$  chromosome correction results in a trend of greater diversity on putative X-linked loci), and no correlation with polymorphism levels in orthologous genes in *C. elegans* or *C. briggsae* (CUTTER 2006; CUTTER *et al.* 2006).

We tested for associations between the diversity values and several genic features. Synonymous-site diversity correlates positively with G + C content at fourfold degenerate sites (GC3s; Spearman’s  $\rho = 0.84$ ,  $P = 0.036$ ) and with the population recombination parameter estimates,  $\rho = 4N_e r$ , see below (Spearman’s  $\rho = 0.83$ ,  $P = 0.042$ ), but not with other features of the genes such as codon usage bias ( $F_{op}$ ) or divergence between *C. briggsae* and *C. remanei* ( $d_N$ ,  $d_S$  or  $d_N/d_S$ ). Identical associations were observed for silent-site diversity ( $\pi_{si} \times GC3s$  Spearman’s  $\rho = 0.93$ ,  $P = 0.008$ ;  $\pi_{si} \times \rho_{4N_e r}$  Spearman’s  $\rho = 0.89$ ,  $P = 0.019$ ), and polymorphism at nonsynonymous sites was not significantly correlated

**TABLE 3**  
**Summary of per-locus diversity in *C. remanei* in the entire sample**

Locus	<i>n</i>	<i>S</i> <sub>2</sub> <sup>a</sup>	<i>S</i> <sub>3</sub> <sup>b</sup>	Sequence length			Diversity				Diversity				Reference	
				Total	Syn.	Noncod.	Nonsyn.	$\pi_{si}$	$\pi_s$	$\pi_{nc}$	$\pi_a$	$\theta_{si}$	$\theta_s$	$\theta_{nc}$		$\theta_a$
<i>p15</i>	33	54	2	693	119.3	137	390.7	0.055	0.050	0.058	0.0013	0.055	0.045	0.063	0.0006	This study
<i>p24</i>	31	43	6	551	124.1	0	424.9	0.096	0.096	n/a	0.0050	0.097	0.097	n/a	0.0053	This study
<i>p17</i>	32	27	1	686	152.6	0	522.4	0.019	0.019	n/a	0.0064	0.023	0.023	n/a	0.0071	This study
<i>p22</i>	33	75	7	661	140.3	68	435.7	0.118	0.129	0.094	0.0028	0.096	0.110	0.069	0.0045	This study
<i>p18</i>	34	26	0	672	83.9	236	345.1	0.015	0.033	0.008	0.0043	0.014	0.032	0.007	0.0057	This study
<i>p21</i>	33	27	0	687	140.8	50	492.2	0.037	0.035	0.044	0.0000	0.035	0.035	0.030	0.0000	This study
<i>Cr-gfp-1</i>	12			426	81.3	56	281.7	0.046	0.058	0.027	0.0036	0.041	0.049	0.029	0.0035	GRAUSTEIN <i>et al.</i> (2002)
<i>Cr-tra-2</i>	12			418	63.6	120	227.4	0.021	0.010	0.026	0.0025	0.018	0.010	0.022	0.0029	GRAUSTEIN <i>et al.</i> (2002)
<i>Cr-odr-3</i>	11			2060	210.3	1049	779.8	0.021	0.020	0.021	0.0002	0.019	0.017	0.020	0.0004	JOVELIN <i>et al.</i> (2003)
<i>Cr-fem-3</i>	11			2637	295.6	547	1015.4	0.047	0.046	0.047	0.0107	0.060	0.051	0.064	0.0138	HAAG and ACKERMAN (2005)
Average (this study)								0.0567	0.0605	0.0512	0.0033	0.0531	0.0569	0.0422	0.0039	
Average (all)								0.0474	0.0497	0.0407	0.0037	0.0457	0.0469	0.0380	0.0044	

Syn., synonymous; noncod., noncoding; nonsyn., nonsynonymous.

<sup>a</sup>Diallelic segregating sites.

<sup>b</sup>Triallelic segregating sites.

with nonsynonymous divergence ( $\pi_a \times d_N/d_S$  Spearman's  $\rho = 0.77$ ,  $P = 0.072$ ). Diversity at nonsynonymous sites did not correlate with synonymous-site diversity ( $P = 0.7$ ).

**Population structure:** We evaluated the *C. remanei* samples for evidence of population structure using three approaches, recognizing that the available samples are imperfect for testing the full scale and extent of structure in the species as a whole. First, we used the data from the three sampling locations with three or more individuals [Wright State University (WSU), Achilles Hill, and Gloucester; Table 1] to compute standard measures of interpopulation genetic variation ( $F_{st}$ ,  $K_{st}^*$ ) for each locus, tested for significance with  $K_s^*$  (HUDSON *et al.* 1992a,b). Using this approach, we identified evidence of structure only for locus *p24*, an effect that appears to be driven by the lack of diversity among the samples from Gloucester (Table 4). For all other loci,  $F_{st}$ -values were very close to zero (Table 4). A resampling approach to infer confidence intervals on  $F_{st}$  (Table 4) allows us to exclude the presence of strong population structure like that observed in other *Caenorhabditis* species (BARRIÈRE and FÉLIX 2005; HABER *et al.* 2005; SIVASUNDAR and HEY 2005; CUTTER 2006; CUTTER *et al.* 2006). We obtained similar results with the  $S_{nn}$  nearest-neighbor population structure statistic (HUDSON 2000), again with only locus *p24* showing significant evidence of structure after Bonferroni multiple-tests correction ( $S_{nn} = 0.88$ ,  $P < 0.001$ ). Furthermore, overall levels of diversity were similar whether all samples are analyzed together or separately for the three sampling localities or for the collection of singleton samples each from different localities (Table 4).

In a second approach to identifying population subdivision, we used the program Structure to attempt to infer the maximum-likelihood number of subpopulations ( $K$ ) from the pattern of polymorphism in the sample (PRITCHARD *et al.* 2000). This approach yielded higher log-likelihood values for  $K$  increasing from 1 to 5. However, the program assigns roughly equal proportions of each subpopulation to each individual for all values of  $K$  that we explored (*i.e.*,  $\sim 1/K$  of each individual is assigned as being derived from each putative subpopulation; Figure 1). This suggests little, or no, population subdivision. Although it should be noted that this program was not intended to be used with sequence data (*i.e.*, from closely linked polymorphic sites), and this use can lead to overestimation of  $K$ , we used a model allowing recombination to make these data fit the assumptions required by Structure as closely as possible. Finally, we represented the relationships between the sampled individuals graphically using neighbor networks (HUDSON and BRYANT 2006), which qualitatively illustrate the lack of any clear association between sequences and sampling locality (Figure 2; supplemental Figure 1 at <http://www.genetics.org/supplemental/>). Overall, these data provide very little

**TABLE 4**  
**Silent-site diversity for sample subsets and tests for population structure**

Locus	Diversity ( $\pi_{si}$ )					Population structure <sup>g</sup>					
	All <sup>a</sup>	WSU <sup>b</sup>	Achilles <sup>c</sup>	Gloucester <sup>c</sup>	Singleton samples <sup>d</sup>	$F_{st}$	$K_{st}^*$	$P$	$F_{st}$ (95% C.I.) <sup>h</sup>		$K_{st}^*$ (95% C.I.) <sup>h</sup>
<i>p15</i>	0.055	0.061	0.054	0.046	0.045	-0.028	-0.004	0.54	-0.013 (-0.149-0.193)		0.047 (-0.006-0.139)
<i>p24</i>	0.096	0.096	0.064	0.000	0.108 <sup>e</sup>	0.500	0.099	<0.001	0.143 (-0.004-0.317)		0.135 (0.095-0.218)
<i>p17</i>	0.019	0.015	0.026 <sup>f</sup>	0.013	0.022	0.085	0.037	0.16	0.059 (-0.258-0.534)		0.062 (-0.016-0.171)
<i>p22</i>	0.118	0.128	0.106	0.081	0.109	0.050	-0.001	0.41	-0.077 (-0.223-0.091)		0.046 (0.003-0.120)
<i>p18</i>	0.015	0.009	0.029	0.027	0.014	-0.041	0.108	0.055	-0.358 (-0.644-0.476)		0.166 (0.001-0.422)
<i>p21</i>	0.037	0.040	0.038	0.035	0.028	0.032	0.034	0.16	0.001 (-0.254-0.295)		0.089 (0.007-0.230)
Average	0.0567	0.0582	0.0530	0.0336	0.0544	0.100	0.045				

<sup>a</sup> Sample size  $31 \leq n \leq 34$ .

<sup>b</sup>  $15 \leq n \leq 17$ .

<sup>c</sup>  $n = 3$ .

<sup>d</sup>  $n = 9$  from localities with only a single sample.

<sup>e</sup>  $n = 8$ .

<sup>f</sup>  $n = 2$ .

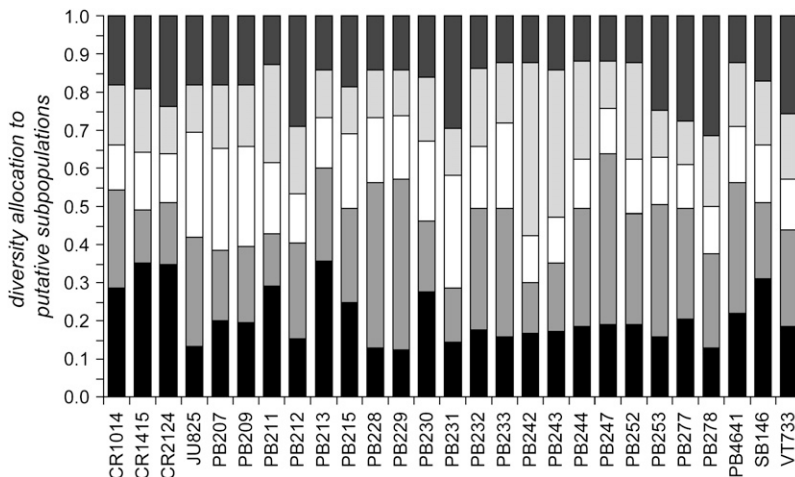
<sup>g</sup> Based on three sampling localities (Wright State University woods, Achilles Hill, Gloucester).

<sup>h</sup> Median, 2.5%, and 97.5% percentiles from resampled distribution.

evidence of consistent population structure, indicating that these samples approximate panmixia.

**Recombination and linkage disequilibrium:** Evidence for recombination is prevalent both within and among loci, with significant test results for intralocus recombination for five of the six loci (Table 5). Using a four-gamete test to infer the minimum number of recombination events in the sample for each locus (HUDSON and KAPLAN 1985), at least two recombination events are detected in the two loci with the least evidence of recombination (Table 5). We also calculated the population recombination parameter ( $\rho = 4N_e r$ ; effective population size  $N_e$  and recombination rate  $r$ ) for each locus using the maximum-likelihood method implemented in LDhat (FEARNHEAD and DONNELLY 2001), yielding an average value of  $\rho = 0.0346$  (Table 5). Consistent with crossing over in these regions, pairwise linkage disequilibrium (LD) decays significantly over

the ~600 bp of sequence in four of the six loci (Figure 3), with significant LD present between only 0.4% of sites (after multiple-tests correction). Within loci,  $r^2$ -values average 0.208 (Table 5). No significant LD is detected between sites in different loci, and the average  $r^2$  between loci equals 0.050. To illustrate the expected decay of LD with distance given the inferred recombination parameter ( $\rho$ ), we overlay WEIR and HILL's (1986) Equation 3 on the plots of pairwise  $r^2$  and distance values for each locus (Figure 3). Extrapolating this decline in linkage disequilibrium with distance to the point at which the interlocus value is reached ( $r^2 = 0.05$ ) suggests that linkage equilibrium should occur after ~1-2 kb on average (Table 5). In a reanalysis of published sequence variation in 3 kb of the *Cr-fem-3* gene (HAAG and ACKERMAN 2005), linkage disequilibrium also decays significantly with distance (Spearman's  $\rho = 0.30$ ,  $P < 0.0001$ ), reaching background levels after



**FIGURE 1.**—Partitioning of genetic variation among putative subpopulations identified by Structure in an example run with  $K = 5$ . Different shadings representing different putative subpopulations are present at similar frequency within each sample, indicating absence of population structure.

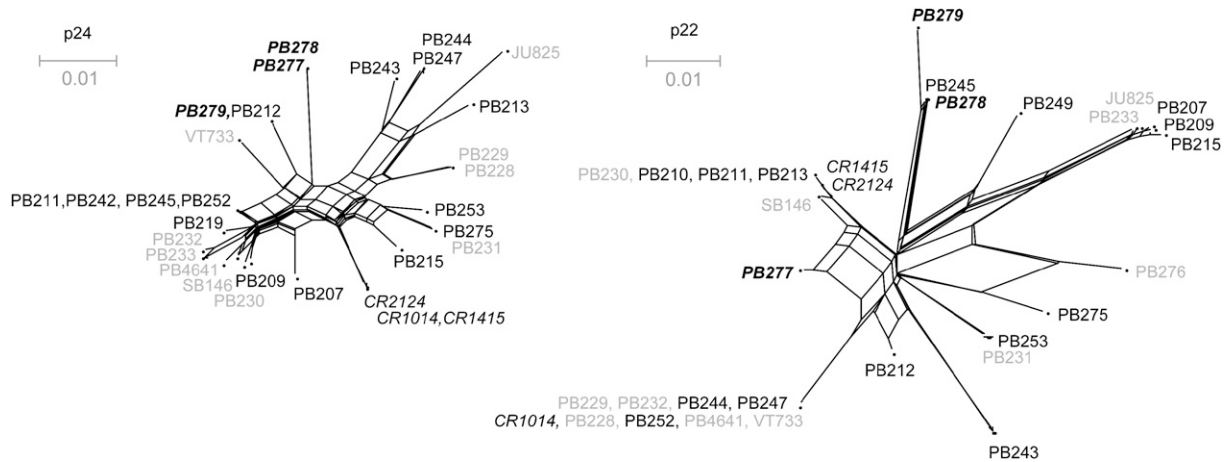


FIGURE 2.—Neighbor networks for two loci. Population structure was identified for locus *p24*, but not for locus *p22*. Reticulation of the branches indicates recombination. Samples from WSU are in regular type (see Table 1), those from Achilles Hill are in boldface italics (PB277–PB279), those from Gloucester are in italics (CR1014, CR1415, and CR2124), and samples from a singleton locality are in shaded type. See supplemental Figure 1 at <http://www.genetics.org/supplemental/> for all loci.

534 bp (given the estimate of the recombination parameter  $\rho$  from LDhat of 0.0325,  $n = 11$ ). For comparison, linkage disequilibrium in *C. elegans* for a genomewide sample of 230 SNPs from 11 strains (KOCH *et al.* 2000) decays on a scale of tens of centimorgans (Figure 4), which implies LD over many megabases given an average recombination rate of  $\sim 3$  cM/Mb (BARNES *et al.* 1995).

**Tests of neutrality:** Given the findings of limited, if any, population structure and extensive recombination, it is also informative to consider the potential for demography and selection to perturb equilibrium neutral patterns of polymorphism at the surveyed loci. Genetic diversity measures based on pairwise differences ( $\pi$ ) or the number of segregating sites ( $\theta$ ) are expected to yield approximately equal values at equilibrium under a standard neutral model (WATTERSON 1975; NEI and LI 1979; TAJIMA 1983). Departures from this neutral expectation are quantified by statistics such as Tajima's *D*, for which values different from zero may reflect the action of selection or nonneutral demographic processes (TAJIMA 1989a). For all six loci sampled here, no significant departures from the neutral expectation were observed by either Tajima's *D* or Fu and Li's *D*<sup>\*</sup>, assuming no recombination (Table 6). However, the distribution of *D* and *D*<sup>\*</sup> calculated from coalescent simulations with free recombination in DnaSP indicated a departure from neutrality for *p22* alone (*D*,  $P = 0.001$ ; *D*<sup>\*</sup>,  $P = 0.001$ ), as well as when calculated using observed levels of recombination for *D*<sup>\*</sup> ( $P = 0.027$ ). The average value of *D* of 0.196 across loci is slightly positive, and *D*<sup>\*</sup> is similar, as also reported for two other nuclear loci in this species (GRAUSTEIN *et al.* 2002). This could be a signature of a recent bottleneck or population contraction; however, data from many more loci are necessary to determine whether this is a general pattern in the genome. Tests of neutrality that are based on

divergence, *e.g.*, HKA tests (HUDSON *et al.* 1987), unfortunately cannot be applied here because of saturated synonymous-site divergence relative to the nearest known congener (Table 6). However, the relatively high amino acid divergence and low codon usage bias of *p18* (Table 6) suggest that this locus likely experiences much weaker purifying selection than the other loci.

**Effective population size:** Given values of the population parameters  $\rho$  and  $\theta$  and the assumption of demographic equilibrium, it is possible to calculate corresponding estimates of the effective population size ( $N_e$ ) because  $\rho = 4N_e r$  and  $\theta = 4N_e \mu$  (where  $r$  is the recombination rate and  $\mu$  is the mutation rate; LI 1997). From the average  $\pi_{si}$  used as an estimate of  $\theta$ , and applying the *C. elegans* neutral mutation rate of  $9.0 \times 10^{-9}$ /bp rate to *C. remanei* (DENVER *et al.* 2004; KEIGHTLEY and CHARLESWORTH 2005), we obtain a diversity-based estimate of  $N_e$  in *C. remanei* of  $\sim 1.6 \times 10^6$  (per locus range:  $4.2 \times 10^5$ – $3.3 \times 10^6$ ). It is not yet known whether the mutation rate estimated in *C. elegans* applies in other species, but patterns of divergence suggest roughly equivalent mutation rates among Caenorhabditis species (CUTTER and PAYSEUR 2003), although *C. briggsae* may have a slightly higher rate than *C. elegans* (BAER *et al.* 2005). To make a similar calculation on the basis of  $\rho$ , a measure of recombination rate ( $r$ ) is required. *C. remanei* has six chromosomes, each of which likely experience only a single crossover per meiosis (HILLERS and VILLENEUVE 2003), and its total genome size is slightly larger than those of *C. elegans* and *C. briggsae* ( $\sim 130$  Mb; J. S. JOHNSTON and R. WATERSTON, personal communication), suggesting an average recombination rate of  $\sim 2.3$  cM/Mb ( $6 \times 50$  cM/130 Mb). Using this genomic average recombination rate as an estimate of  $r$ , the observed mean  $\rho$  of 0.0346/bp implies an effective size of  $N_e \sim 3.7 \times 10^5$  (per locus range:  $4.0 \times 10^4$ – $1.5 \times 10^6$ ). This calculation based on  $\rho$  depends on the

**TABLE 5**  
**Intralocus recombination and linkage disequilibrium**

Locus	$R_m^a$	$\rho^b$	Corr( $r^2$ , $d$ ) <sup>c</sup>	Significance <sup>d</sup>	Mean $r^2$	$d_{r^2=0.05}^e$
<i>p15</i>	12	0.0162	-0.096	+ + -	0.190	1068.8
<i>p24</i>	11	0.1350	-0.117	+ + +	0.101	128.3
<i>p17</i>	2	0.0052	-0.040	- - -	0.158	3329.8
<i>p22</i>	19	0.0269	-0.080	+ + +	0.162	643.7
<i>p18</i>	2	0.0037	0.070	- + -	0.404	4679.4
<i>p21</i>	4	0.0207	-0.276	+ + +	0.235	863.4
Average	8.3	0.0346	-0.090	NA	0.208	1781.1

<sup>a</sup> Minimum number of recombination events (HUDSON and KAPLAN 1985).

<sup>b</sup>  $\rho = 4N_e r$  (per base pair; FEARNHEAD and DONNELLY 2001).

<sup>c</sup> Correlation between pairwise linkage disequilibrium ( $r^2$ ) and distance between sites ( $d$ ).

<sup>d</sup>  $P < 0.05$  (+) or  $P \geq 0.05$  (-) left to right for corr( $r^2$ ,  $d$ ),  $L_{k_{max}}$ , and  $\phi_w$  tests for the presence of recombination (McVEAN *et al.* 2002; BRUEN *et al.* 2006).

<sup>e</sup> Distance in base pairs at which background level of linkage disequilibrium is reached.

appropriateness of our estimate of the average genomic recombination rate, which may deviate from the local rate of recombination for any given locus. Consequently, if the six loci tend to reside in regions of lower than average recombination, the  $\rho$ -based  $N_e$  calculation will be an underestimate. If real, the difference in  $N_e$  inferred from  $\theta$  and  $\rho$  could also indicate excess linkage disequilibrium, which could result from departures from demographic equilibrium or from natural selection acting at or near these loci (ANDOLFATTO and PRZEWSKI 2000; ANDOLFATTO and WALL 2003). Nevertheless, it is plausible from both of these calculations that *C. remanei* has an effective population size on the order of 1 million.

## DISCUSSION

### Patterns of nucleotide polymorphism in *C. remanei*:

Natural population samples of *C. remanei* are characterized by very high silent-site nucleotide diversity, some loci having  $\pi_{si}$  as high as 10%. This does not appear to be an artifact of population structure, as intralocus measures of diversity paint the same picture and there is little evidence for population structure. The range of diversity among the six loci surveyed here encompasses the values previously reported for four other nuclear loci studied in different samples from this species (Table 3; GRAUSTEIN *et al.* 2002; JOVELIN *et al.* 2003; HAAG and ACKERMAN 2005). Considering all 10 loci together yields an average  $\pi_{si}$  of 4.7%, implying that nearly 1 in 20 silent sites differs between any two randomly selected *C. remanei* sequences.

In stark contrast to *C. remanei*, the self-fertilizing congeners *C. elegans* and *C. briggsae* exhibit ~20-fold less diversity in orthologous loci (GRAUSTEIN *et al.* 2002; JOVELIN *et al.* 2003; CUTTER 2006; CUTTER *et al.* 2006). This difference in diversity between outcrossing and selfing *Caenorhabditis* is consistent with the prediction

of lower genetic variation within inbreeding populations, although it is not entirely clear which aspects of inbreeding might have led to such a dramatic reduction in diversity in selfing *Caenorhabditis* (GRAUSTEIN *et al.* 2002; SIVASUNDAR and HEY 2005; CUTTER 2006; CUTTER *et al.* 2006). Loci from some other species of nematodes also demonstrate diversity at nuclear loci as high as the ~10% seen in *C. remanei*, such as  $\beta$ -tubulin in *Haemonchus contortus* (BEECH *et al.* 1994) and ITS1-rDNA of *Longidorus bififormis* (YE *et al.* 2004), although most reports indicate levels of silent-site diversity of  $\leq 2\%$  (ANDERSON *et al.* 1998; CUTTER *et al.* 2006). Similarly high levels of nucleotide polymorphism have also been reported in several species of mycophagous *Drosophila* (DYER and JAENIKE 2004; SHOEMAKER *et al.* 2004), suggesting that intraspecific per-nucleotide diversities of 5–10% may not be unusual in a variety of taxa, in spite of the predominant view of low-to-moderate levels of genetic variation within genetic model organisms: *Drosophila melanogaster*  $\pi_{si} \sim 1.6\%$  (ANDOLFATTO 2001), *Arabidopsis thaliana*  $\theta_{si} \sim 1.6\%$ , maize  $\theta_{si} \sim 1.5\%$  (WRIGHT and GAUT 2005), and human  $\pi_{si} \sim 0.08\%$  (ZHANG 2000; YU *et al.* 2001).

As in other obligately outbreeding species, linkage disequilibrium between pairs of polymorphic sites in *C. remanei* decays rapidly with distance (LONG *et al.* 1998; REMINGTON *et al.* 2001; TENAILLON *et al.* 2001). For most of the loci examined here, linkage disequilibrium declines significantly over just a few hundred base pairs at a rate that suggests linkage equilibrium will generally be reached at distances  $>1$ –2 kb. *D. melanogaster* and maize show similar rapidly decaying linkage disequilibrium (LONG *et al.* 1998; REMINGTON *et al.* 2001; TENAILLON *et al.* 2001), whereas 50–60 kb is required for its decay in humans and in *A. thaliana* (REICH *et al.* 2001; NORDBORG *et al.* 2005). The self-fertile wild barley, *Hordeum vulgare*, exhibits LD to a comparable extent as maize (MORRELL *et al.* 2005), whereas in the self-fertile *C. elegans*, linkage



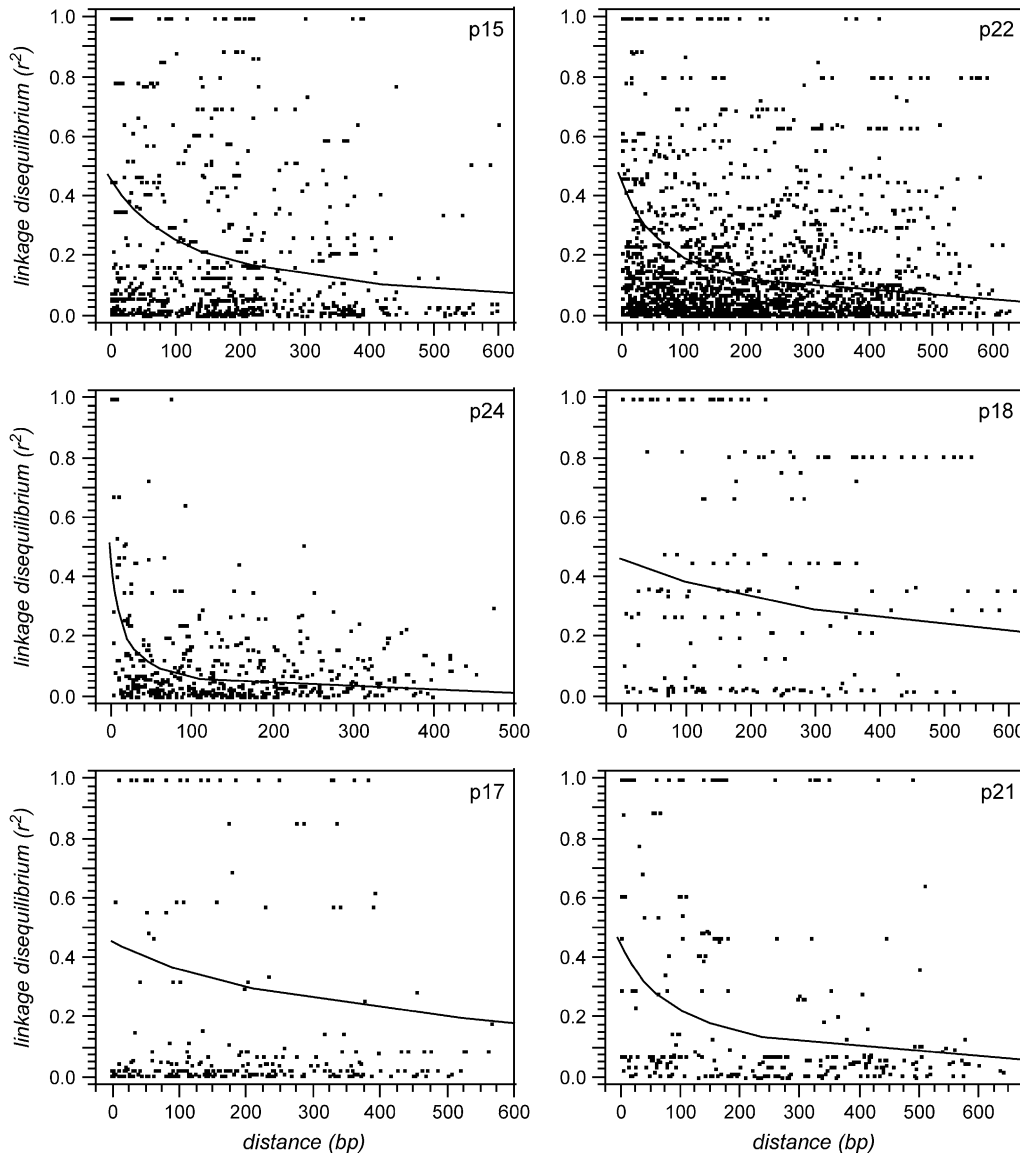


FIGURE 3.—Decay of linkage disequilibrium with distance for each locus. Curves indicate the expected relationship given the inferred estimate of the recombination parameter  $\rho$  and sample size of each locus.

disequilibrium decays at a scale of megabases and even extends between pairs of sites on different chromosomes (Figure 4), which is consistent with very low outcrossing rates (BARRIÈRE and FÉLIX 2005; HABER *et al.* 2005; CUTTER 2006). The precise relationship between LD and distance for individual loci, however, is defined by the combined effects of local rates of recombination and mutation, selection regimes, and stochasticity inherent to the coalescent process. For example, the correlation that we observe between diversity levels ( $\pi$  or  $\theta$ ) and estimates of the recombination parameter ( $\rho$ ) in these data could indicate a reduced efficacy of natural selection in lower recombination regions due to background selection or genetic hitchhiking (MAYNARD SMITH and HAIGH 1974; CHARLESWORTH *et al.* 1993; WIEHE and STEPHAN 1993; HUDSON and KAPLAN 1995). Alternatively, mutation or biased gene conversion rates correlated with recombination could result in greater diversity in high recombination regions (MARAIS and

DURET 2001; MARAIS 2003). The additional observation of a GC-content  $\times$   $\rho$  correlation, without an association between  $\rho$  and codon usage bias, suggests the operation of biased gene conversion, although further data are required to adequately test this.

We also note that values of  $\rho/\theta$  in *C. remanei* are somewhat lower than expected from inferences of genome size, map length, and mutation rate (*C. remanei*  $r \sim 2.3 \times 10^{-8}$  M/bp/generation, *C. elegans*  $\mu \sim 9 \times 10^{-9}$  mutations/bp/generation; DENVER *et al.* 2004; KEIGHTLEY and CHARLESWORTH 2005; J. S. JOHNSTON and R. WATERSTON, personal communication). At equilibrium,  $\rho = 4N_c r$  and  $\theta = 4N_c \mu$ , so  $\rho/\theta$  provides an index of the average number of recombination events per mutation event ( $r/\mu$ ). Here, we observe mean  $\rho/\theta = 0.5$  and expect  $r/\mu \sim 2.5$ . A more formal analysis of these issues with additional data is warranted; low values of  $\rho/\theta$ , reflecting excess linkage disequilibrium due to historical demographic perturbations, have also been described

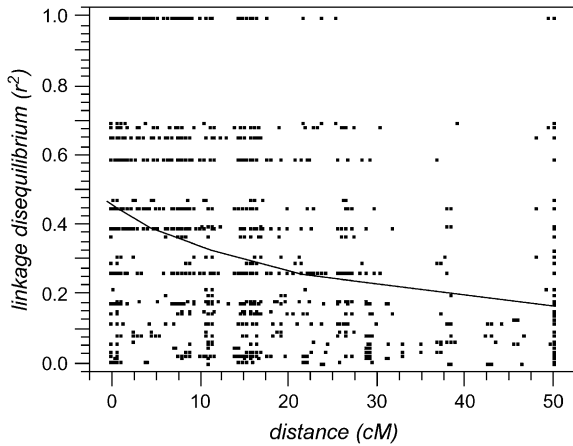


FIGURE 4.—Decay of linkage disequilibrium with genetic distance in *C. elegans*. The solid curve indicates the least-squares regression fit of linkage disequilibrium values to their expected decline with distance. Points at 50 cM indicate pairs of sites from different chromosomes; data are from KOCH *et al.* (2000).

for *D. melanogaster* (ANDOLFATTO and PRZEWSKI 2000; ANDOLFATTO and WALL 2003; HADDRILL *et al.* 2005).

We found no difference in levels of diversity between putatively autosomal and X-linked loci. Findings of comparable levels of nucleotide diversity at autosomal and X-linked loci also have been reported for several species of *Drosophila* (ANDOLFATTO 2001; DYER and JAENIKE 2004), despite the potentially reduced effective population size of the X chromosome (by virtue its hemizyosity in males). Some populations of *D. melanogaster* and *D. simulans*, however, appear to have elevated (African) or reduced (non-African) levels of diversity on the X chromosome (BEGUN and WHITLEY 2000; ANDOLFATTO 2001; KAUER *et al.* 2002; GLINKA *et al.* 2003; THORNTON and ANDOLFATTO 2006); thus, future evaluation of diversity for *C. remanei* population samples from different continents might provide additional insights into the processes patterning genetic variation.

Locus *p24*, the putative ortholog of the *C. elegans* “undescribed” gene ZK430.1, has the most extreme values for several population statistics in this sample of loci,

including the highest  $\rho$ ,  $\theta_{si}$ , and  $F_{st}$  and lowest  $r^2$ . Is this locus or a linked locus subject to selection, or is the genomic region subject to unusual characteristics? Tests of the frequency spectrum (*e.g.*, Tajima’s  $D$ ) do not suggest nonneutrality, and the coding sequence exhibits only an average rate of protein evolution ( $d_N$ ). However,  $d_S$ , a measure of the mutation rate from interspecific comparisons, is also the most extreme value among the loci in this sample. Thus, *p24* may simply reside in a region of both high mutation and high recombination.

**Consistency with panmixia and demographic equilibrium:** The obligately outcrossing *C. remanei* stands in striking contrast to *C. elegans* and *C. briggsae* in terms of showing little evidence of population structure (BARRIÈRE and FÉLIX 2005; HABER *et al.* 2005; SIVASUNDAR and HEY 2005; CUTTER 2006; CUTTER *et al.* 2006). The results of analyses of population subdivision and tests of neutrality point to the conclusion that these samples roughly conform to the predictions of a panmictic population at demographic equilibrium. However, it is important to recognize that (1) more extensive population subsampling will improve our understanding of how genetic variation is distributed across space and (2) the analysis of more loci is required to infer the degree to which population growth or contraction may have affected diversity patterns. For example, should larger samples of loci confirm a genomewide trend of a deficit of rare alleles (*e.g.*, positive Tajima’s  $D$ ) and a consistently larger estimate of  $N_e$  inferred from diversity than from linkage disequilibrium, then this could suggest recent population contraction in the history of these, primarily American, *C. remanei* samples (TAJIMA 1989a; ANDOLFATTO and WALL 2003). Population sampling for disparate regions (Europe, Asia) will also aid in our understanding of the historical processes that shape genome evolution in this species. The extensive recombination and large effective population size inferred for *C. remanei*, on the order of  $N_e \sim 10^6$ , imply that selection will operate efficiently on traits subject to even very weak selection. Consequently, scans of the genome for deviant patterns of polymorphism and linkage disequilibrium hold great promise for identifying regions that may be subject to natural selection in the wild, without

TABLE 6  
Divergence from *C. briggsae*, G + C content, and tests of neutrality

Locus	$d_N$	$d_S$	max $F_{op}$	GC3s	GCnc	Tajima’s $D$	Fu and Li’s $D^*$
<i>p15</i>	0.013	1.21	0.460	0.46	0.35	−0.019	0.192
<i>p24</i>	0.102	2.08	0.394	0.49	NA	0.333	0.209
<i>p17</i>	0.095	1.63	0.472	0.42	NA	−0.478	0.675
<i>p22</i>	0.056	1.55	0.650	0.47	0.35	0.857	1.083
<i>p18</i>	0.247	1.38	0.311	0.42	0.36	0.252	0.545
<i>p21</i>	0.026	1.88	0.411	0.42	0.24	0.228	0.969
Average	0.090	1.62	0.450	0.45	0.32	0.196	0.612

Calculated for sequenced portions of genes only.

requiring extensive corrections for the confounding effects of complex demographic scenarios. Furthermore, should the high molecular diversity correspond to substantial phenotypic variation, *C. remanei* may provide an exceptionally good system for dissecting the genetic basis of quantitative traits based on natural variation (JOVELIN *et al.* 2003).

**General implications for *Caenorhabditis*:** Obligate outcrossing defines the ancestral mode of reproduction among *Caenorhabditis* nematodes, from which self-fertilization has evolved independently in at least the two lineages leading to *C. elegans* and *C. briggsae* (KIONTKE *et al.* 2004). The molecular mechanisms controlling germline differentiation and sex determination appear to have diverged dramatically among these species (CLIFFORD *et al.* 2000; HAAG *et al.* 2002; STOTHARD *et al.* 2002; NAYAK *et al.* 2005). However, the duration of selfing in the *C. elegans* and *C. briggsae* lineages remains an open question: extant intraspecific diversity would permit an origin as recently as a few hundred thousand years ago despite a much more ancient common ancestor of known species of *Caenorhabditis* (COGHLAN and WOLFE 2002; STEIN *et al.* 2003; CUTTER 2006; CUTTER *et al.* 2006). Regardless of the dating of the origin of self-fertilization, much of the evolutionary histories of these species would have occurred in an obligately outcrossing ancestor. If *C. remanei* adequately reflects many of the characteristics common to obligately outcrossing *Caenorhabditis* species, then it will provide a useful model for the evolutionary circumstances surrounding the gonochoristic progenitors of *C. elegans* and *C. briggsae*. This should prove particularly insightful for understanding the role of natural selection in shaping such traits as longevity, sex-limited behaviors and chemosensation, codon usage bias, and transitions in the mode of reproduction.

We thank M.-A. Félix for providing strain JU825, E. Haag and M. Palopoli for sharing sequence data files with us, and E. Stahl for help with samplestats. We also gratefully acknowledge the comments of D. Begun, K. Dyer, P. Haddrill, and two anonymous reviewers on previous drafts of the manuscript and the suggestions of B. Charlesworth and W. G. Hill relating to several of the analyses in this work. The *Caenorhabditis* Genetics Center kindly provided some strains that were used in this study. This work was funded by the National Science Foundation with International Research Fellowship Program grant no. 0401897 to A.D.C.

#### LITERATURE CITED

- ANDERSON, T. J. C., M. S. BLOUIN and R. N. BEECH, 1998 Population biology of parasitic nematodes: applications of genetic markers. *Adv. Parasitol.* **41**: 219–283.
- ANDOLFATTO, P., 2001 Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **18**: 279–290.
- ANDOLFATTO, P., and M. PRZEWSKI, 2000 A genome-wide departure from the standard neutral model in natural populations of *Drosophila*. *Genetics* **156**: 257–268.
- ANDOLFATTO, P., and J. D. WALL, 2003 Linkage disequilibrium patterns across a recombination gradient in African *Drosophila melanogaster*. *Genetics* **165**: 1289–1305.
- BAER, C. F., F. SHAW, C. STEDING, M. BAUMGARTNER, A. HAWKINS *et al.*, 2005 Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. *Proc. Natl. Acad. Sci. USA* **102**: 5785–5790.
- BAIRD, S. E., 1999 Natural and experimental associations of *Caenorhabditis remanei* with *Trachelipus rathkii* and other terrestrial isopods. *Nematology* **1**: 471–475.
- BAIRD, S. E., 2002 Haldane's rule by sexual transformation in *Caenorhabditis*. *Genetics* **161**: 1349–1353.
- BAIRD, S. E., and W.-C. YEN, 2000 Reproductive isolation in *Caenorhabditis*: terminal phenotypes of hybrid embryos. *Evol. Dev.* **2**: 9–15.
- BAIRD, S. E., M. E. SUTHERLIN and S. W. EMMONS, 1992 Reproductive isolation in Rhabditidae (Nematoda, Secernentea): mechanisms that isolate 6 species of 3 genera. *Evolution* **46**: 585–594.
- BARNES, T. M., Y. KOHARA, A. COULSON and S. HEKIMI, 1995 Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. *Genetics* **141**: 159–179.
- 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.
- BEECH, R. N., R. K. PRICHARD and M. E. SCOTT, 1994 Genetic variability of the beta-tubulin genes in benzimidazole-susceptible benzimidazole-resistant strains of *Haemonchus contortus*. *Genetics* **138**: 103–110.
- BEGUN, D. J., and P. WHITLEY, 2000 Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA* **97**: 5960–5965.
- BRAENDLE, C., and M. A. FELIX, 2006 Sex determination: ways to evolve a hermaphrodite. *Curr. Biol.* **16**: R468–R471.
- BRUEN, T. C., H. PHILIPPE and D. BRYANT, 2006 A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**: 2665–2681.
- CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- CHASNOV, J. R., and K. L. CHOW, 2002 Why are there males in the hermaphroditic species *Caenorhabditis elegans*? *Genetics* **160**: 983–994.
- 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.
- CLIFFORD, R., M. H. LEE, S. NAYAK, M. OHMACHI, F. GIORGINI *et al.*, 2000 FOG-2, a novel F-box containing protein, associates with the GLD-1 RNA binding protein and directs male sex determination in the *C. elegans* hermaphrodite germline. *Development* **127**: 5265–5276.
- COGHLAN, A., and K. H. WOLFE, 2002 Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. *Genome Res.* **12**: 857–867.
- 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. A. PAYSEUR, 2003 Rates of deleterious mutation and the evolution of sex in *Caenorhabditis*. *J. Evol. Biol.* **16**: 812–822.
- CUTTER, A. D., and S. WARD, 2005 Sexual and temporal dynamics of molecular evolution in *C. elegans* development. *Mol. Biol. Evol.* **22**: 178–188.
- CUTTER, A. D., M.-A. FÉLIX, A. BARRIÈRE and D. CHARLESWORTH, 2006 Patterns of nucleotide polymorphism distinguish temperate and tropical wild isolates of *Caenorhabditis briggsae*. *Genetics* **173**: 2021–2031.
- 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.
- DYER, K. A., and J. JAENIKE, 2004 Evolutionarily stable infection by a male-killing endosymbiont in *Drosophila innubila*: molecular evidence from the host and parasite genomes. *Genetics* **168**: 1443–1455.
- FEARNHEAD, P., and P. DONNELLY, 2001 Estimating recombination rates from population genetic data. *Genetics* **159**: 1299–1318.
- FÉLIX, M. A., 2004 Genomes: a helpful cousin for our favourite worm. *Curr. Biol.* **14**: R75–R77.

- FITCH, D. H. A., and S. W. EMMONS, 1995 Variable cell positions and cell contacts underlie morphological evolution of the rays in the male tails of nematodes related to *Caenorhabditis elegans*. *Dev. Biol.* **170**: 564–582.
- FLOYD, R., E. ABEBE, A. PAPERT and M. BLAXTER, 2002 Molecular barcodes for soil nematode identification. *Mol. Ecol.* **11**: 839–850.
- GLINKA, S., L. OMETTO, S. MOUSSET, W. STEPHAN and D. DE LORENZO, 2003 Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. *Genetics* **165**: 1269–1278.
- 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., S. P. WANG and J. KIMBLE, 2002 Rapid coevolution of the nematode sex-determining genes *fem-3* and *tra-2*. *Curr. Biol.* **12**: 2035–2041.
- HABER, M., M. SCHUNGEL, A. PUTZ, S. MULLER, B. HASERT *et al.*, 2005 Evolutionary history of *Caenorhabditis elegans* inferred from microsatellites: evidence for spatial and temporal genetic differentiation and the occurrence of outbreeding. *Mol. Biol. Evol.* **22**: 160–173.
- HADDRILL, P. R., K. R. THORNTON, B. CHARLESWORTH and P. ANDOLFATTO, 2005 Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res.* **15**: 790–799.
- HILL, K. L., and S. W. L'HERNAULT, 2001 Analyses of reproductive interactions that occur after heterospecific matings within the genus *Caenorhabditis*. *Dev. Biol.* **232**: 105–114.
- HILLERS, K. J., and A. M. VILLENEUVE, 2003 Chromosome-wide control of meiotic crossing over in *C. elegans*. *Curr. Biol.* **13**: 1641–1647.
- HUDSON, R. R., 2000 A new statistic for detecting genetic differentiation. *Genetics* **155**: 2011–2014.
- 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, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA-sequences. *Genetics* **111**: 147–164.
- 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.
- HUDSON, R. R., D. D. BOOS and N. L. KAPLAN, 1992a A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* **9**: 138–151.
- HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992b Estimation of levels of gene flow from DNA-sequence data. *Genetics* **132**: 583–589.
- HUSON, D. H., and D. BRYANT, 2006 Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**: 254–267.
- IKEMURA, T., 1985 Codon usage and transfer-RNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* **2**: 13–34.
- JOVELIN, R., B. C. AJJE and P. C. PHILLIPS, 2003 Molecular evolution and quantitative variation for chemosensory behaviour in the nematode genus *Caenorhabditis*. *Mol. Ecol.* **12**: 1325–1337.
- KAUER, M., B. ZANGERL, D. DIERINGER and C. SCHLOTTERER, 2002 Chromosomal patterns of microsatellite variability contrast sharply in African and non-African populations of *Drosophila melanogaster*. *Genetics* **160**: 247–256.
- KEIGHTLEY, P. D., and B. CHARLESWORTH, 2005 Genetic instability of *C. elegans* comes naturally. *Trends Genet.* **21**: 67–70.
- KIONTKE, K., and W. SUDHAUS, 2006 Ecology of *Caenorhabditis* species, in *Wormbook*, edited by THE C. ELEGANS RESEARCH COMMUNITY (<http://www.wormbook.org>).
- 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.
- KOCH, R., H. G. A. M. VAN LUENEN, M. VAN DER HORST, K. L. THIJSSSEN and R. H. A. PLASTERK, 2000 Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*. *Genome Res.* **10**: 1690–1696.
- LAMUNYON, C. W., and S. WARD, 1999 Evolution of sperm size in nematodes: sperm competition favours larger sperm. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **266**: 263–267.
- LI, W. H., 1997 *Molecular Evolution*. Sinauer, Sunderland, MA.
- LONG, A. D., R. F. LYMAN, C. H. LANGLEY and T. F. C. MACKAY, 1998 Two sites in the Delta gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics* **149**: 999–1017.
- MARAIS, G., 2003 Biased gene conversion: implications for genome and sex evolution. *Trends Genet.* **19**: 330–338.
- MARAIS, G., and L. DURET, 2001 Synonymous codon usage, accuracy of translation, and gene length in *Caenorhabditis elegans*. *J. Mol. Evol.* **52**: 275–280.
- MAYNARD SMITH, J., and J. HAIGH, 1974 Hitch-hiking effect of a favorable gene. *Genet. Res.* **23**: 23–35.
- MCCULLOCH, D., and D. GEMS, 2003 Evolution of male longevity bias in nematodes. *Aging Cell* **2**: 165–173.
- MCVEAN, G., P. AWADALLA and P. FEARNHEAD, 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231–1241.
- MILLER, M. A., V. Q. NGUYEN, M.-H. LEE, M. KOSINSKI, T. SCHEDL *et al.*, 2001 A sperm cytoskeletal protein that signals oocyte meiotic maturation and ovulation. *Science* **291**: 2144–2147.
- MORRELL, P. L., D. M. TOLENO, K. E. LUNDY and M. T. CLEGG, 2005 Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. *Proc. Natl. Acad. Sci. USA* **102**: 2442–2447.
- NAYAK, S., J. GOREE and T. SCHEDL, 2005 *fog-2* and the evolution of self-fertile hermaphroditism in *Caenorhabditis*. *PLoS Biol.* **3**: e6.
- 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.
- PRITCHARD, J. K., M. STEPHENS and P. DONNELLY, 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- REICH, D. E., M. CARGILL, S. BOLK, J. IRELAND, P. C. SABETI *et al.*, 2001 Linkage disequilibrium in the human genome. *Nature* **411**: 199–204.
- REMINGTON, D. L., J. M. THORNSBERRY, Y. MATSUOKA, L. M. WILSON, S. R. WHITT *et al.*, 2001 Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA* **98**: 11479–11484.
- ROZAS, J., J. C. SANCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.
- SHOEMAKER, D. D., K. A. DYER, M. AHRENS, K. MCAABEE and J. JAENIKE, 2004 Decreased diversity but increased substitution rate in host mtDNA as a consequence of Wolbachia endosymbiont infection. *Genetics* **168**: 2049–2058.
- SIVASUNDAR, A., and J. HEY, 2005 Sampling from natural populations using RNAi reveals high outcrossing and population structure in *Caenorhabditis elegans*. *Curr. Biol.* **15**: 1598–1602.
- STEIN, L. D., Z. BAO, D. BLASIAZ, 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.
- STOTHARD, P., D. HANSEN and D. PILGRIM, 2002 Evolution of the PP2C family in *Caenorhabditis*: rapid divergence of the sex-determining protein FEM-2. *J. Mol. Evol.* **54**: 267–282.
- TAJIMA, F., 1983 Evolutionary relationship of DNA-sequences in finite populations. *Genetics* **105**: 437–460.
- TAJIMA, F., 1989a The effect of change in population-size on DNA polymorphism. *Genetics* **123**: 597–601.
- TAJIMA, F., 1989b 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.

- THORNTON, K., and P. ANDOLFATTO, 2006 Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics* **172**: 1607–1619.
- WATTERSON, G. A., 1975 Number of segregating sites in genetic models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- WEIR, B. S., and W. G. HILL, 1986 Nonuniform recombination within the human beta-globin gene cluster. *Am. J. Hum. Genet.* **38**: 776–778.
- 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.
- WRIGHT, S. I., and B. S. GAUT, 2005 Molecular population genetics and the search for adaptive evolution in plants. *Mol. Biol. Evol.* **22**: 506–519.
- YANG, Z. H., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**: 555–556.
- YE, W. M., A. L. SZALANSKI and R. T. ROBBINS, 2004 Phylogenetic relationships and genetic variation in Longidorus and Xiphinema species (Nematoda: Longidoridae) using ITS1 sequences of nuclear ribosomal DNA. *J. Nematol.* **36**: 14–19.
- YU, N., Z. ZHAO, Y. X. FU, N. SAMBUUGHIN, M. RAMSAY *et al.*, 2001 Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. *Mol. Biol. Evol.* **18**: 214–222.
- ZHANG, D. Y., 2000 Resource allocation and the evolution of self-fertilization in plants. *Am. Nat.* **155**: 187–199.

Communicating editor: D. BEGUN