Population Genetics of *Caenorhabditis elegans*: The Paradox of Low Polymorphism in a Widespread Species

Arjun Sivasundar and Jody Hey[1]

Department of Genetics, Rutgers University, Piscataway, New Jersey 08854

Manuscript received June 26, 2002
Accepted for publication October 17, 2002

ABSTRACT

Caenorhabditis elegans has become one of the most widely used model research organisms, yet we have little information on evolutionary processes and recent evolutionary history of this widespread species. We examined patterns of variation at 20 microsatellite loci in a sample of 23 natural isolates of *C. elegans* from various parts of the world. One-half of the loci were monomorphic among all strains, and overall genetic variation at microsatellite loci was low, relative to most other species. Some population structure was detected, but there was no association between the genetic and geographic distances among different natural isolates. Thus, despite the nearly worldwide occurrence of *C. elegans*, little evidence was found for local adaptation in strains derived from different parts of the world. The low levels of genetic variation within and among populations suggest that recent colonization and population expansion might have occurred. However, the patterns of variation are not consistent with population expansion. A possible explanation for the observed patterns is the action of background selection to reduce polymorphism, coupled with ongoing gene flow among populations worldwide.

THE nematode *Caenorhabditis elegans* is among the most widely studied model organisms in current biological and biomedical research. In the area of population genetics, however, research on *C. elegans* lags behind that on other research organisms. For example, population genetic research on Drosophila has led to the development of tests of natural selection (Hudson *et al.* 1987; McDonald and Kreitman 1991; Moriyama and Powell 1996; McAllister and McVean 2000), the mapping of quantitative trait loci (Mackay 2001), and the study of diverse biological processes such as speciation (Kliman *et al.* 2000), codon usage bias (Powell and Moriyama 1997), and aging (Schmidt *et al.* 2000). In *Arabidopsis thaliana*, research is increasingly drawing upon population genetic studies to elucidate a range of biochemical and developmental pathways (Purugganan and Suddith 1998, 1999; Maloof *et al.* 2000, 2001), and to develop tools for fine-scale mapping (Nordborg *et al.* 2002). However, in the case of *C. elegans* evolutionary inquiries are often limited to comparisons with the closest known relative, *C. briggsae*, which is thought to have diverged from *C. elegans* 20–120 million years ago (Kent and Zahler 2000), or with more distantly related species such as *Pristionchus pacificus* and *Oscheius* sp. (Delattre and Felix 2001; Grandien and Sommer 2001). Such comparisons cannot effectively address questions at the microevolutionary level. The importance of population genetic studies and baseline data on intraspecific variation in *C. elegans* is only now being recognized (Delattre and Felix 2001).

C. elegans reproduces mainly by self-fertilization, which can reduce the effective population size, and thus polymorphism levels, by up to one-half (Pollak 1987). A further reduction in polymorphism at neutral loci is expected due to the effects of genetic hitchhiking (Maynard Smith and Haigh 1974) and background selection (Charlesworth *et al.* 1993), both of which are enhanced when recombination is reduced, as will occur under high rates of self-fertilization. Notwithstanding these factors, other aspects of *C. elegans* population biology lead to an expectation of high levels of variation, at least on a global scale. *C. elegans* occurs in most parts of the world (Abdul Kader and Cote 1996; Hodgkin and Doniach 1997) and may well be very numerous in soils or other habitats. Even if population genetic factors associated with high rates of self-fertilization cause variation to be reduced within one population, the worldwide distribution leads to the prediction of large amounts of variation among populations. However, very low levels of DNA sequence polymorphism have been observed at four loci (three nuclear and one mitochondrial) on a global scale (Koch *et al.* 2000; Graustein *et al.* 2002).

A species-wide lack of genetic variation in a widespread and numerous organism is paradoxical. In this study, we examine whether these same low levels of variation are seen at numerous loci located throughout the genome. We examine variation at microsatellite loci for evidence of geographic structure, gene flow, and

[1]*Corresponding author:* Department of Genetics, Rutgers University, Nelson Biological Labs, 604 Allison Rd., Piscataway, NJ 08854-8082. E-mail: hey@biology.rutgers.edu

TABLE 1

Wild isolate strains of *C. elegans* used in this study

Reference no.	Strain name	Geographic origin
1	N2	Bristol, England
2	AB1	Australia
3	PB303	Unknown
4	CB4851	Bergerac, France
5	TR388	Madison, WI
6	CB3191	Altadena, CA
7	CB3199	Pasadena, CA
8	CB4932	Taunton, England
9	LSJ1	Berkeley, CA
10	CB4852	Rothamsted, England
11	CB4853	Altadena, CA
12	CB4854	Altadena, CA
13	CB4855	Palo Alto, CA
14	CB4856	Hawaii
15	CB4857	Claremont, CA
16	CB4858	Pasadena, CA
17	TR403	Madison, WI
18	DH424	El Prieto Canyon, CA
19	RC301	Freiburg, Germany
20	DR1349	Pasadena, CA
21	DR1350	Pasadena, CA
22	KR314	Vancouver
23	DR1344	Bergerac, France

All strains were obtained from the Caenorhabditis Genetics Center (University of Minnesota).

population expansion. Our results provide insights into the recent evolutionary history of *C. elegans* and identify processes that might have given rise to the observed patterns of genetic variation on a genome-wide, global scale.

METHODS

Strains: Samples of 23 strains of *C. elegans* isolated from the wild were obtained from the Caenorhabditis Genetics Center (CGC; see Table 1). These represent most of the natural isolates available through the Center. These strains have been isolated at various times from the 1960s to the present and deposited at the CGC. *C. elegans* strains may be maintained frozen and stored indefinitely in liquid nitrogen (Stiernagle 1999). Due to the CGC's practice of freezing strains upon receipt and thawing them when requested, the strains when obtained from the CGC are generally just a few generations away from the original isolate (T. Stiernagle, CGC, personal communication). Specific details for each strain may be obtained from the CGC (http://biosci.umn.edu/CGC/Strains/strains.htm). Worms were grown on 6-cm petri plates at 19° on nematode growth medium, seeded with *Escherichia coli*.

PCR and electrophoresis: From each strain, 20–25 worms were collected and DNA was extracted by standard methods (Johnstone 1999). PCR was performed in 11-μl reaction volumes with 0.38 μl of each primer (1 μm), 1.11 μl of 10× buffer, 0.89 μl $MgCl_2$ (25 mm), 1.11 μl of dNTPs (2.5 mm), 0.44 μl of IRD M13 primer (1 μm; Licor), 0.25 units of *Taq* DNA polymerase, and 1 μl of genomic DNA. PCR products were separated by electrophoresis on a 7% acrylamide gel run on a Licor 4200 automated DNA sequencer. Allele sizes were scored using GeneImagIR software (Licor).

Microsatellite loci: From an exhaustive list of microsatellites in the *C. elegans* genome (Katti *et al.* 2001), 20 were chosen at random from all six chromosomes (see Table 2). Using the complete genome sequence of *C. elegans* (available at ftp://ncbi.nlm.nih.gov/genbank/genomes/C_elegans/), primers were designed to amplify these repeat regions. One of the primers in each pair was end labeled. All the loci consisted of dinucleotide repeats, with the repeat motif occurring from 10 to 32 times in the reference sequence. In addition, one locus (2003) contained a string of G's following the GA repeat. All loci chosen were in noncoding regions of the genome, either in introns or in intergenic regions.

Statistical analyses: *Expected number of alleles under the infinite-alleles model:* The expected number of alleles in a sample of n chromosomes is given by $k_e = 1 + M/(M + 1) + M/(M + 2) + \ldots + M/(M + n - 1)$ (Ewens 1972), where $M = 4Nu$ (N is the effective population size; u is the mutation rate per locus). At mutation-drift equilibrium, M can be estimated as $M = H/(1 - H)$ (Crow and Kimura 1970), where H is the heterozygosity. However, when using H estimates from a single locus, this formula is biased and a bias-corrected estimate is obtained by solving the equation $M^3 + (7 - M_0)M^2 + (8 - 5M_0)M - 6M_0 = 0$ (Chakraborty and Weiss 1991), where $M_0 = H/(1 - H)$.

Expected number of alleles under the stepwise mutation model: A bias-corrected estimator of M is obtained by solving $1.7M^4 + (25 - 1.7M_0)M^3 + (24.5 - 13M_0)M^2 + (9 - 22.5M_0)M - 6M_0 = 0$ (Estoup *et al.* 1995), where $M_0 = 0.5((1 - H)^{-2} - 1)$ (Ohta and Kimura 1973). The expected number of alleles in a sample of n individuals is estimated using Equation 18 of Kimura and Ohta (1975).

Tests of population expansion: The parameter $\theta = 4Nu$ can be estimated from the genetic variance V and also from the homozygosity P_0 at a locus. Under a population growth model, the variance-based and homozygosity-based estimates of θ deviate from each other. This deviation, measured by the imbalance index β, can be used to detect population expansion (Kimmel *et al.* 1998; King *et al.* 2000). β was estimated in two ways,

$$\ln \hat{\beta}_1 = \ln \hat{\theta}_V - \ln \hat{\theta}_{P_0}$$

and

$$\ln \hat{\beta}_2 = \frac{1}{L}\sum_{i=1}^{L}(\ln \hat{\theta}_V)_i - (\ln \hat{\theta}_{P_0})_i,$$

for a sample of L loci indexed by i (King *et al.* 2000).

TABLE 2

Microsatellite loci examined in this study

Locus	Chromosome	Position	Repeat type	Primers (5′ to 3′)	No. of alleles
1001	1	5370653	$(CA)_{11}$	gagcacatgttttgtggcat gcgtaaccgcttacaagatga	1
1002	1	10894817	$(TA)_{21}$	aagcggcaattctgatgagtg gacgtgcatgcattcgatt	1
2002	2	11286765	$(AG)_{10}$	ttggggtttttttagagaggcg atgaaaaagggcgcgttcta	1
3001	3	4119024	$(TC)_{11}$	tcatcgctgacatatttgtcg ccaccatcatcacacaaggaa	1
3002	3	8166362	$(CA)_{12}$	tttcaagcaacctcatccgt tgtggctctgtcttcaatcaa	1
4002	4	10562854	$(TC)_{10}$	tttattgccccctctatct gcttcgaattgccaatatcg	1
5001	5	7110445	$(GA)_{13}$	agccgtaatggtcgaagagat tttttctcgcgcctctgtctt	1
5002	5	14211871	$(CA)_{10}$	ccccagacaataaggaaccaa tggaaagttgggacgaatga	1
X001	X	5799018	$(CT)_{10}$	aagcgagatcctgaagcaaa aacgcattcttcattgcgtg	1
X002	X	11542117	$(AT)_{32}$	tctggatcacatgtctcgttg ggcgtgcaaaaacagtttc	1
1003	1	6811516	$(AT)_{21}$	gcaaagacttttcgaccaacc agttacggctattgggatcaa	3
1004	1	13131044	$(AC)_{22}$	tttataagctccgaggggttg agccgagcataacgaattga	4
2001	2	6430443	$(TC)_{25}$	catttgggaatgctccaaac tgagaggaggaggaggaaaa	12
2003	2	5254926	$(GA)_{26}(G)_{14}$	gatgcaaacaggcacagaaa gagattgctgctgctggtgt	9
3003	3	3187348	$(TC)_{30}$	tggtcatacttcttctcgcaa ggcaaatagaatgtgaaaggg	6
3004	3	7612601	$(AG)_{20}$	ttacgttcagaaggaggagca aagcgtcattgtttttgcc	5
4001	4	5348085	$(GA)_{10}$	cctattttccgcctccactta ctccaatgaagcttccgaat	3
4004	4	8310545	$(AG)_{33}$	gcacaatgagcaacatgcaa gaagtcaaaaaccaaaaaccc	7
X003	X	2315494	$(CT)_{35}$	ccgacaccgtgatccataaat aggatgtgcagtgatgaaagg	7
X004	X	15253554	$(CA)_{30}$	ccccacattttctgtgtgctt tcacttttatccttccacccc	Polymorphic

Position on the chromosome and repeat type refer to the genome of the N2 strain, which was used to select the loci and design primers. All monomorphic loci are listed first followed by the polymorphic loci.

$\hat{\theta}_V$, the variance-based estimator of θ, and $\hat{\theta}_{P_0}$, the homozygosity estimator of θ, were computed using Equations 2, 3, and 5 from King *et al.* (2000).

For a growing population, the expected distribution of allele lengths is smoothly peaked, as most bifurcations date back to the time of population expansion. However, in the case of a constant-sized population, most pairs of alleles are either closely related or distantly related since genealogy is expected to include a relatively ancient bifurcation associated with the most recent common ancestor of the entire sample. The expectation of a deep genealogy leads in turn to an expectation of an allele length distribution that is ragged and multipeaked. The within-locus *k*-test (Reich and Goldstein 1998) uses the shape of the allele-length distribution to differentiate between smooth and ragged distributions of allele lengths and thereby test for population expansion. The test statistic is given by

$$k = 2.5 \times \mathrm{Sig}^4 + 0.28 \times S^2 - 0.95/n - \mathrm{Gam}_4,$$

where Sig^4 is an estimate of the variance of the allele-length distribution squared, Gam_4 is an estimate of the fourth central moment of the allele-length distribution, and S^2 is the sample variance of the allele-length distribution (Reich and Goldstein 1998).

The interlocus *g*-test (Reich and Goldstein 1998)

employs the principle that the variance of the variance in allele sizes across loci is expected to be larger in a constant-sized population than in a growing one. This results from ancient bifurcations in a growing population having similar dates at all loci, but having considerable variation among loci in a constant-sized population. The test statistic is $g = \mathrm{Var}[V_j]/(4/3\overline{V}^2 + 1/6\overline{V})$, where V_j is the variance at the jth locus, and $\overline{V}$ is the average of all V_j. (Equation 3 of Reich and Goldstein 1998; see also Reich *et al.* 1999).

Maximum-likelihood estimates of population growth rate: Beaumont (1999) developed a Bayesian approach for using microsatellite data to assess recent changes in population size. The method yields posterior distributions of population mutation rates for individual loci, and for all loci collectively it provides distributions of the exponential population growth rate *r*, defined as the ratio of the current population size to that just prior to the period of population size change, and *tf*, which is the time since the population size began to change, expressed in units of the current population size. Distributions were obtained using the computer program MSVAR (Beaumont 1999), which conducts Markov chain Monte Carlo simulations. Noninformative, flat prior distributions were used for all parameters. Several chains were run, yielding ~79,000 points representing $\sim 9 \times 10^7$ parameter updates, and the distributions from multiple independent runs were pooled. Convergence was assessed in two ways: first, by comparing posterior distributions for parameters from independent runs with different starting parameter values, and second, by looking at plots of parameter values against time and checking for "stickiness." For each parameter the maximum-likelihood estimate was identified as the peak of the posterior distribution, and the 95% highest posterior density (HPD) credible set was obtained by finding the shortest interval that contained 95% of the posterior probability.

Analyses of population structure: The software package Arlequin V2.000 (Schneider *et al.* 2000) was used to estimate population parameters. Population substructure was examined using F_{ST}, which does not explicitly take into account a mutational mechanism, as well as R_{ST} (Slatkin 1995), which does. Correlation between genetic and geographic distances was examined using a Mantel test with 10,000 permutations, also using Arlequin V2.000.

RESULTS

Patterns of variation at microsatellite loci: Twenty microsatellite loci were examined in 23 strains of *C. elegans*. No variation was observed within any of the strains at any of the loci, and each strain was represented by a single allele at each locus, as expected if individuals are highly homozygous due to being cultured by self-fertilization for long periods. The fact that a number of worms from each strain were used to extract DNA is not intended to reflect any within-strain variation that may have existed in the original source populations. Any variation within a strain is unlikely to persist in a single plate after several generations of laboratory culture. Thus, essentially, for each locus, 23 chromosomes were sampled worldwide.

Ten loci were monomorphic among all the strains. The remaining 10 loci were polymorphic, with the number of alleles ranging from 3 to 12 (mean = 6.2). The longer loci tended to have more alleles ($r^2 = 0.43$, $P < 0.05$), although one of the longest loci (X002, 32 repeats) was monomorphic. Alleles at one of the polymorphic loci (X004) could not be scored unambiguously; hence this locus was excluded from subsequent analyses.

The strains TR388 from Wisconsin and N2 from England were identical at all 19 loci. Likewise, strains DH424 (El Prieto Canyon, California) and DR1349 (a subclone of strain PA-1; Pasadena, California) were identical at all loci. It is unlikely that cross-contamination of cultures has occurred since the strains have been maintained at the stock center (T. Stiernagle, CGC, personal communication). Although outcrossing rates in the wild are not known, it is possible that *C. elegans* has a low outcrossing rate and that this coupled with occasional migration of strains over long distances might explain the finding of identical genotypes in disparate locations. One way that such a clonal population structure may be revealed is by the presence of significant linkage disequilibrium between loci on different chromosomes. To test this, we classified alleles for each locus as either common or rare and looked for nonrandom associations of alleles, under a two-locus two-allele model, using Fisher's exact test. Of the 36 pairwise comparisons between loci on different chromosomes, only 2 were significant at the 5% level, and the exact probabilities were distributed evenly over the interval from zero to one. Thus, there is no evidence for the presence of any significant level of linkage disequilibrium among the loci examined.

An alternative explanation for the finding of identical strains in distant locations is very recent migration. While such movement of individuals could be effected by association with, for example, birds, it is also possible that human activity has been a factor facilitating *C. elegans* dispersal.

Of the nine variable loci, only two (1004 and 3004) displayed variation entirely in multiples of the motif length (2 bp; see Table 3 and Figure 1). The allele sizes suggest that mutation patterns do not conform strictly to the stepwise model of microsatellite mutation. Table 3 shows the observed number of alleles at each variable locus and the expected number of alleles under infinite-allele and stepwise mutation models. These estimates were corrected for bias (Zouros 1979; Chakraborty and Weiss 1991; Estoup *et al.* 1995), and the concordance of the observed data with the predictions was

TABLE 3

Expected and observed number of alleles under the infinite-alleles model (IAM) and the stepwise mutation model (SMM)

			IAM			SMM		
Locus	n	Observed	Expected	$M = 4N_e u$	N_e	Expected	$M = 4N_e u$	N_e
1004	22	4	4.3275	0.7343	1031.32	3.304876	1.0489	1473.17
2001	23	12	15.3637	6.4672	9083.15	10.14899	31.4544	44177.53
3003	23	6	7.4845	1.7463	2452.67	4.916384	3.2743	4598.74
4001	23	3	1.7995	0.1449	203.51	1.711436	6.3178	8873.31
4004	23	7	8.2121	2.0393	2864.19	5.304924	4.13	5800.56
X003	23	7	11.6322	3.7678	5291.85	7.355142	11.6013	16293.96
1003	22	3	4.2828	0.722	1014.04	3.280925	1.0278	1443.54
2003	23	9	12	4.2399	5954.92	7.602395	14.3884	20208.43
3004	16	5	5.533	1.2573	1765.87	4.06782	2.0719	2909.97

N_e is calculated assuming $u = 1.78 \times 10^{-4}$ (Frisse 1999).

tested by the nonparametric Wilcoxon signed ranks test. The observed data are not significantly different from the predictions of the infinite-alleles model ($P > 0.05$) but are different from the predictions of the stepwise mutation model ($P < 0.01$). These predictions, however, assume an equilibrium population size, while in fact the recent history of the species might have involved large changes in population size. Frisse (1999) conducted

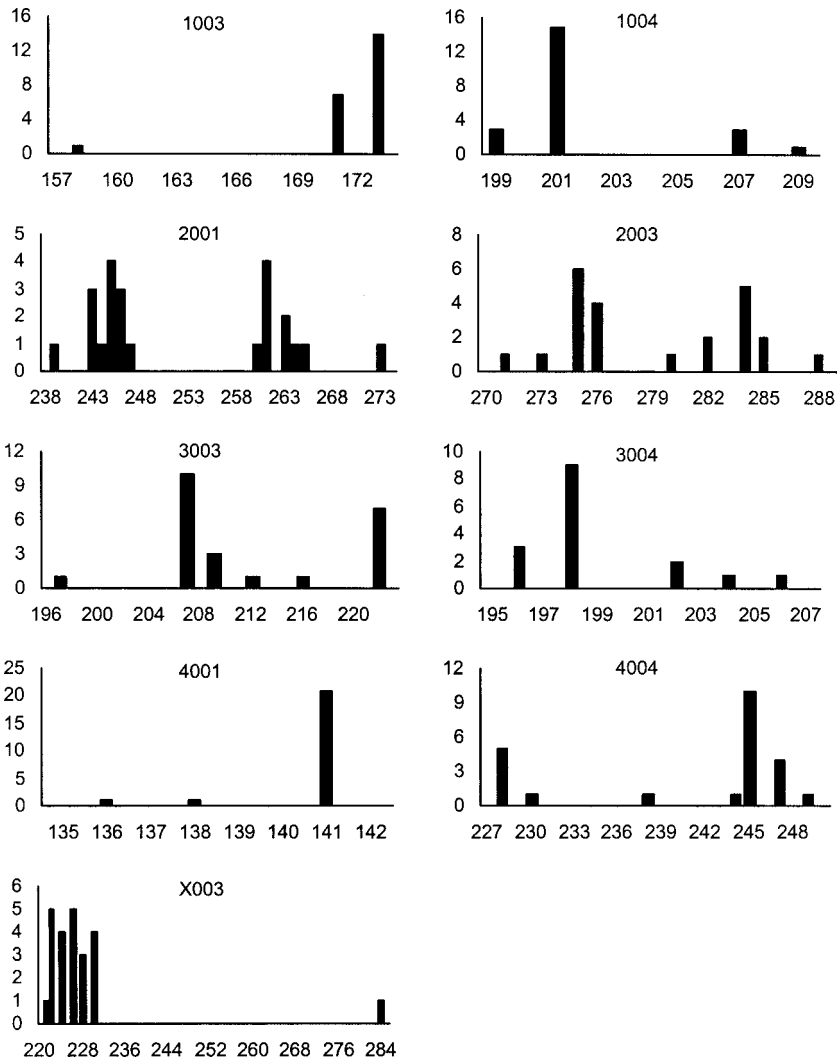

Figure 1.—Distribution of allele sizes for nine microsatellite loci. The horizontal axis depicts the absolute sizes of PCR products for each locus.

mutation-accumulation experiments using 100 lines each started from the descendants of a single, inbred, wild-type worm, which were propagated from a single, randomly picked worm each generation to minimize the effects of selection. It was found that for microsatellites in the range of sizes used in this study, most mutations involved the addition or deletion of a single repeat unit. It is possible that these microsatellites are evolving according to the stepwise model, and the lack of concordance in number of alleles results from the demographic history of the populations.

Correlation between genetic and geographic distances: The geographic origin of one strain (PB303) is not known, so this strain was excluded from the geographic analyses. The strains were assigned to regions (Europe, North America, Australia, and Hawaii) and to populations within regions (in Europe—England, France, and Germany; in North America—northern California, southern California, Vancouver, and Wisconsin; in Australia and Hawaii—one strain each). Partitioning of genetic variation within populations, among populations, and among regions was examined using analysis of molecular variance (AMOVA; Excoffier *et al.* 1992), starting from a matrix of pairwise genetic distances. Using pairwise distances measured as the number of shared alleles, which corresponds to the infinite-alleles model, $F_{ST} = 0.38$ is not significant at the 0.05 level. About 62% of the total variance is contained within populations. However, using the sum of squared allele-size differences, a distance measure corresponding to the stepwise mutation model, 54% of the variance is among regions, and $R_{ST} = 0.73$ is significant ($P < 0.02$). Thus there is some evidence of population structure in *C. elegans* under the assumption of a stepwise model of mutation.

Similarity between pairs of strains was calculated as $S =$ (the number of alleles shared over all loci)/(the number of loci). Pairwise distances were then calculated as $(1 - S)$ (Bowcock *et al.* 1994). Geographic distances between the sites ("as the crow flies") were obtained using a web-based program (available at http://www.indo.com/distance/) and a 22×22 geographic distance matrix was constructed. No correlation was detected between the geographic and genetic distance matrices ($r = 0.02$, $P = 0.348$; Mantel test, 10,000 permutations), using the software package Arlequin V2.000 (Schneider *et al.* 2000).

Tests of population expansion: Population expansion can leave a strong signature on allele-size distributions (Kimmel *et al.* 1998). Populations that have been growing are expected to have distributions that are smooth and have a single peak, whereas populations that have been constant in size or have been shrinking will have more ragged, multipeaked allele-frequency distributions (see methods). Figure 2 provides examples of several simulated distributions under a model of constant population size and a model of recent population growth.

Three statistical tests were employed to detect population expansion events from multilocus microsatellite data. None reveals a signature of population expansion in the data. Of the three tests, the imbalance index β is the most sensitive to population increase. In particular, β is highly responsive to expansions of recent origin (King *et al.* 2000). Populations that have recently expanded exhibit $\beta < 1$ (*i.e.*, $\ln \beta < 0$). The values of $\ln(\beta)$ calculated in two ways ($\ln \beta 1$ and $\ln \beta 2$, see methods) are 3.235 and 2.175, respectively. The within-locus *k*-test (Reich and Goldstein 1998; Reich *et al.* 1999) is based on the shape of the allele-size distribution. The probability of obtaining a positive value of *k* is constrained between 0.515 and 0.55. Population expansion is characterized by fewer loci than expected having a positive *k* value under the null hypothesis of constant population size. A one-tailed binomial test, with the probability of a positive *k* conservatively set to 0.515, can be used to test for departures from the null hypothesis (Reich and Goldstein 1998). By this method, we find no evidence of population expansion (*P* value = 0.463). The interlocus *g*-test (Reich and Goldstein 1998) compares the observed and estimated values of the variance in allele sizes across loci. Again, no sign of population expansion is detected by this test at the 0.05 significance level (table of cutoff values in Reich *et al.* 1999). These tests are based on the assumption of a single-step mutational process. Our data indicate that the loci under study might deviate from this model of mutation. When this assumption is violated due to multistep mutations, the probability of wrongly inferring expansion increases for the within-locus test and decreases for the interlocus test (Reich *et al.* 1999). Since neither of these tests detects an expansion event, we infer that these data reveal no evidence of recent population expansion.

When a model of recent population size change was fit to the data, the maximum-likelihood estimate of the growth rate, *r*, was 0.0027 (95% HPD region: 0.00096–0.01205). For the time parameter, *tf*, the peak occurred at 1.117 (95% HPD region: 0.750–1.717). Since *r* is the ratio of the current to ancestral population sizes, these values correspond to a 500-fold reduction in population size over the past $1.12 \times N_0$ generations, where N_0 is the current population size.

DISCUSSION

Despite our extensive knowledge of cellular, developmental, and genetic mechanisms in *C. elegans*, we do not have detailed knowledge of this species' ecology, population genetics, or evolutionary history (Hodgkin and Doniach 1997; Delattre and Felix 2001). In this study we begin to fill that gap on the basis of inferences from patterns of variation at microsatellite loci.

The patterns of genetic variation observed in a worldwide sample of *C. elegans* are intriguing. In general, widespread species with large population sizes are ex-

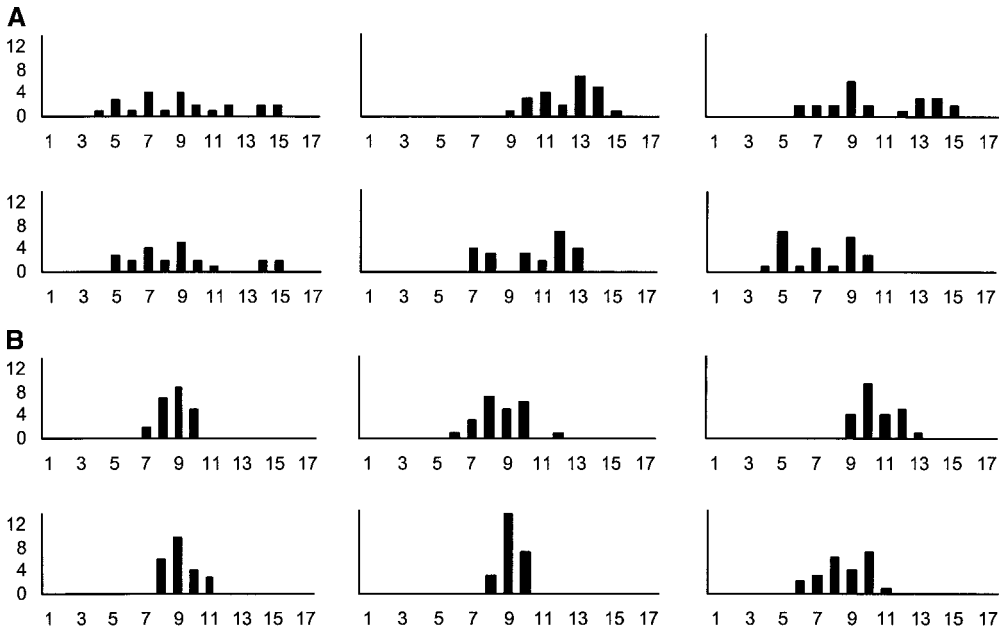

FIGURE 2.—Effect of population growth on the allele-size distribution at microsatellite loci. Coalescent simulations were performed using the program SIMCOAL (http://cmpg.unibe.ch/software/simcoal/). Each panel shows the distribution obtained from six independent simulation runs of sample size 23, under a given set of parameter values. In all cases, a stepwise mutation model with a mutation rate of 0.00178 per locus per generation and a limit of 20 alleles was used, and the current effective population size was set at 14,000, which is approximately the average estimate across loci for N_e under the stepwise mutation model. (A) Constant population size. (B) Exponential growth with $r = 0.005$.

pected to harbor large amounts of genetic variation. However, in organisms such as *C. elegans*, which reproduce primarily by selfing, variation within a population may be reduced. High rates of self-fertilization reduce the effective population size by up to a factor of two (POLLAK 1987). Also, neutral loci may be expected to show reduced variability due to recurrent selection against deleterious mutations at linked loci. This background selection effect is predicted to be especially strong in highly or completely selfing species, or asexually reproducing species, as compared to outcrossing species, since recombination is effectively reduced (NORDBORG *et al.* 1996; CHARLESWORTH *et al.* 1997; NORDBORG 1997). Alternatively, selective sweeps leading to the repeated fixation of favorable alleles might lead to the reduction in diversity at neutral loci (MAYNARD SMITH and HAIGH 1974; KAPLAN *et al.* 1989). This expectation is supported by the finding that the largely self-fertilizing species *C. elegans* and *C. briggsae* harbor very low levels of nucleotide polymorphism compared to the gonochoristic *C. remanei*, another member of the *elegans* group of the genus Caenorhabditis (GRAUSTEIN *et al.* 2002). Thus a high rate of inbreeding, coupled with selection, might serve to reduce effective sizes of *C. elegans* populations. Using observed mutation rates for microsatellite loci (FRISSE 1999), estimates of the effective population size from our data range from ~200 to ~44,000 (Table 3). Although they are found worldwide, actual densities of *C. elegans* in nature have not yet been estimated. However, nematodes are generally very abundant, often attaining densities of millions of individuals per square meter (BLAXTER 1998). If *C. elegans* individuals are indeed numerous, then the polymorphism data suggest the presence of a powerful selective force acting to maintain a low level of genetic variation.

The factors that are expected to reduce variation within populations may be expected to increase variation among populations. In the absence of high levels of gene flow, the effects of local inbreeding and local selective sweeps will remove variation within populations but will not remove variation between populations. However, we find that levels of variation on a global scale are very low. Surveys of microsatellites in various species typically report polymorphism at a large proportion of the loci studied (see Table 4 for some examples). The proportion of polymorphic loci is uncommonly low in *C. elegans*. Our results are consistent with those from other studies of interstrain variation that have utilized various classes of genetic markers. Early studies focusing on a limited number of isolates have found low levels of variation for enzymes (BUTLER *et al.* 1981), minisatellite and simple sequence repeat motifs (UITTERLINDEN *et al.* 1989), and Tc1 transposon patterns (EGILMEZ *et al.* 1995). More recently, DNA sequence polymorphism has been found to be very low for both nuclear and mitochondrial loci (KOCH *et al.* 2000; GRAUSTEIN *et al.* 2002). This genome-wide lack of variation might underlie the absence of strong heterosis for life history traits in crosses between strains of *C. elegans* (JOHNSON and HUTCHINSON 1993).

The population genetics portrait that begins to emerge from this and previous studies is of a widespread species that exhibits little variation, either within geographic regions or on a global scale. It is unlikely that this picture has been caused by a biased or small sample of loci, as our sample consists of randomly selected microsatellites located on all the chromosomes of *C. elegans*. Alternatively, a low mutation rate for *C. elegans* microsatellites might account for the low level of polymorphism. However, the mutation rate at microsatellite

TABLE 4

Examples of proportions of polymorphic microsatellite loci in various organisms

Organism	No. of loci	No. polymorphic	%	Source
Caenorhabditis elegans	20	10	50	This study
Linepithema humile	19	8	42 (NS)	Krieger and Keller (1999)
Drosophila melanogaster	10	8	80	England *et al.* (1996)
D. melanogaster	18	17	94	Goldstein and Clark (1995)
D. melanogaster	133	131[a]	98	Kauer *et al.* (2002)[b]
D. persimilis	10	9	90	Noor *et al.* (2000)
D. virilis	11	11	100	Schlotterer and Harr (2000)
D. lummei	11	11	100	Schlotterer and Harr (2000)
D. pseudoobscura	10	10	100	Noor *et al.* (2000)
D. subobscura	18	18	100	Pascual *et al.* (2000)
Anopheles gambiae	11	11	100	Lanzaro *et al.* (1995)
Sheep	105	76	72	Reed and Beattie (2001)
Mouse	7300[a]	6580[c]	90	Dietrich *et al.* (1996)
Human	2506	2327	93	Weissenbach *et al.* (1992)
Saccharomyces cerevisiae	7	7	100	Hennequin *et al.* (2001)
S. cerevisiae	20	20	100	Field and Wills (1998)
Arabidopsis thaliana	55	30[d]	55 (NS)	Bell and Ecker (1994)
Magnolia sieboldii japonica	4	4	100	Kikuchi and Isagi (2002)
Maize	34	34	100	Taramino and Tingey (1996)

The proportion of polymorphic loci is significantly different from that in *C. elegans* ($P < 0.05$) in all but two cases (which are marked NS).

[a] Inferred.

[b] Supplemental data online.

[c] 48% among lab strains; 94% lab strains *vs.* CAST or SPR.

[d] Bell and Ecker (1994): "Of 18 $(CA)_n$ repeats ($n > 13$), only one was polymorphic. In contrast, 25 of 30 $(GA)_n$ repeats, 2 of 3 (AT) in repeats, and 2 of 4 $(A)_n$ repeats were polymorphic."

loci observed from mutation-accumulation experiments ranges from 8.93×10^{-5} to 1.85×10^{-2} (Frisse 1999) for a variety of repeat sizes and motifs. This mutation rate is not unusually low compared to other organisms (*e.g.*, Crawford and Cuthbertson 1996; Goldstein and Schlotterer 1999, p.6 and references therein; Vazquez *et al.* 2000; Thuillet *et al.* 2002; Yue *et al.* 2002). In this light it is noteworthy that the microsatellite mutation rates in Drosophila are very low (Vazquez *et al.* 2000), yet far more polymorphism is observed in this species (Table 4).

A. thaliana may be expected to resemble *C. elegans* in some respects, and in recent years it has been the subject of many population genetic studies (Todokoro *et al.* 1995; Innan *et al.* 1997; Purugganan and Suddith 1998, 1999; Maloof *et al.* 2000, 2001). Like *C. elegans*, *A. thaliana* reproduces mainly by self-fertilization, with an estimated selfing rate of >99% (Bergelson *et al.* 1998), and it has a widespread distribution. Findings in *A. thaliana*, however, are conflicting. On the one hand, studies using microsatellite loci have found no variation within an ecotype, with all individuals homozygous at all loci, but large amounts of variation among ecotypes (Todokoro *et al.* 1995; Innan *et al.* 1997). The picture is one of population structure and possibly recent expansion of the species throughout the world. On the other hand, Bergelson *et al.* (1998) report low levels of both intra- and interpopulation divergence using RFLP markers and no association between geographic and genetic distance. These observations might reflect processes specific to the different classes of genetic markers used in these studies. In *C. elegans*, however, our study and others have observed low levels of variation for a range of different markers.

A possible explanation to account for the low levels of variation in our data and those from other studies is that *C. elegans* has only recently spread throughout the world. If this has been the case then we should see evidence of recent population growth in the allelic distributions of microsatellite loci. We conducted several tests that were designed to detect recent population expansion on the basis of microsatellite data and that are capable of detecting recent expansion events with a high probability (King *et al.* 2000). None of the tests revealed a signature of such an event in the recent history of the organism. We also fit the data to a model of recent exponential size change (Beaumont 1999). The resulting maximum-likelihood estimates suggest that the global population of *C. elegans* has actually been shrinking and are consistent with a recent 500-fold reduction in size. The method assumes that mutation occurs under the stepwise mutation model (SMM; Beaumont 1999), and so the fact that the data do not correspond to this model may partly explain why the

allele-frequency distributions depart so strongly from that expected under population growth. However, the tests are not all similarly sensitive to this assumption (King *et al.* 2000). On balance, it seems unlikely that present *C. elegans* genetic diversity is the result of recent colonization and expansion events.

The patterns of variation can also be considered in light of models of natural selection. If recombination is low, due to low rate of outcrossing, then both selective sweeps and background selection can greatly reduce levels of genetic variation at neutral loci. However, two points argue against selective sweeps as the main cause of low polymorphism. First, selective sweeps that are associated with local adaptation will remove variation only over the local geographic area and will effectively promote population differentiation at the expense of more widespread polymorphisms. However, *C. elegans* reveals low levels of polymorphism globally and only moderate amounts of population structure. Second, a recent selective sweep exerts a very similar effect on patterns of variation at linked neutral loci as does a recent population expansion. In effect, a selective sweep causes a severe bottleneck for linked loci. However, none of the analyses of recent population history found evidence for recent population growth, as would be expected if much of the genome was linked to regions that had recently experienced selective sweeps on a global scale.

C. elegans presents us with an interesting paradox. The species occurs almost worldwide in areas with widely different climates and environments. Yet only a low level of genetic variation exists on a global scale. There appears to be some population structure, but this has not led to appreciable divergence between geographic regions. We also find that the allele-frequency distributions are not as expected if *C. elegans* had recently spread throughout the world or if there had been recent selective sweeps that reduced variation. It is possible that background selection (Charlesworth *et al.* 1993), greatly enhanced in effect due to very low levels of recombination, is the major cause of low levels of variation.

One possible factor that could explain the low level of variation worldwide *and* the absence of evidence of population expansion is ongoing gene flow over long distances among populations of *C. elegans* worldwide. During periods of environmental stress, L2 larvae of *C. elegans* enter a state of diapause, the dauer larva. During this facultative life stage, larvae display behaviors not seen in other stages. The dauer larva can survive long periods without feeding and resist desiccation. The dauer stage may be an adaptation not only for survival during periods of stress, but also for dispersal by birds, insects, or other animals (Riddle 1988; Hodgkin and Doniach 1997). If we assume an equilibrium migration model then the F_{ST} estimate and the relationship, $F_{ST} = 1/(1 + 4N_e m)$ (Wright 1951), where m is the migration rate among populations, lead to an estimate of $N_e m$ of 0.41 individuals migrating among populations per generation. Taken together, the data and analyses provide indirect evidence for movement over long distances of *C. elegans* individuals among populations. This movement may be association with human activity; however, it does not appear to have been solely a recent phenomenon, as we find no evidence for population expansion. Another interpretation of the pattern observed here is that *C. elegans* occurs in very low numbers in nature. Given the rapid growth rate observed in laboratory cultures, the ability to go into the dauer state in unfavorable environments, and the general observation that nematodes usually attain very high densities, this seems an unlikely explanation. However, the gaps in our knowledge of the ecology of this species preclude us from entirely ruling out this possibility.

We thank Christian Schlotterer, Garth Patterson, Chi-hua Chiu, Carlos Machado, Andrew Singson, and two anonymous reviewers for useful discussions and comments on earlier versions of the manuscript. All the *C. elegans* strains used in this work were provided by the Caenorhabditis Genetics Center, which is funded by the National Institutes of Health National Center for Research Resources.

LITERATURE CITED

Abdul Kader, N., and M. G. Cote, 1996 Isolation, identification and characterization of some strains of *Caenorhabditis elegans* (Maupas, 1900) from Quebec. Fundam. Appl. Nematol. **19:** 381–389.

Beaumont, M. A., 1999 Detecting population expansion and decline using microsatellites. Genetics **153:** 2013–2029.

Bell, C. J., and J. R. Ecker, 1994 Assignment of 30 microsatellite loci to the linkage map of Arabidopsis. Genomics **19:** 137–144.

Bergelson, J., E. Stahl, S. Dudek and M. Kreitman, 1998 Genetic variation within and among populations of *Arabidopsis thaliana*. Genetics **148:** 1311–1323.

Blaxter, M., 1998 *Caenorhabditis elegans* is a nematode. Science **282:** 2041–2046.

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd *et al.*, 1994 High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368:** 455–457.

Butler, M. H., S. M. Wall, K. R. Luehrsen, G. E. Fox and R. M. Hecht, 1981 Molecular relationships between closely related strains and species of nematodes. J. Mol. Evol. **18:** 18–23.

Chakraborty, R., and K. M. Weiss, 1991 Genetic variation of the mitochondrial DNA genome in American Indians is at mutation-drift equilibrium. Am. J. Phys. Anthropol. **86:** 497–506.

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

Charlesworth, B., M. Nordborg and D. Charlesworth, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. **70:** 155–174.

Crawford, A. M., and R. P. Cuthbertson, 1996 Mutations in sheep microsatellites. Genome Res. **6:** 876–879.

Crow, J. F., and M. Kimura, 1970 *An Introduction To Population Genetics Theory.* Harper & Row, New York.

Delattre, M., and M. A. Felix, 2001 Microevolutionary studies in nematodes: a beginning. Bioessays **23:** 807–819.

Dietrich, W. F., J. Miller, R. Steen, M. A. Merchant, D. Damron-Boles *et al.*, 1996 A comprehensive genetic map of the mouse genome. Nature **380:** 149–152.

Egilmez, N. K., R. H. Ebert and R. J. Shmookler Reis, 1995 Strain evolution in *Caenorhabditis elegans*: transposable elements as markers of interstrain evolutionary history. J. Mol. Evol. **40:** 372–381.

England, P. R., D. A. Briscoe and R. Frankham, 1996 Microsatellite polymorphisms in a wild population of Drosophila melanogaster. Genet. Res. **67:** 285–290.

Estoup, A., L. Garnery, M. Solignac and J. M. Cornuet, 1995 Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. Genetics **140:** 679–695.

Ewens, W. J., 1972 The sampling theory of selectively neutral alleles. Theor. Popul. Biol. **3:** 87–112.

Excoffier, L., P. E. Smouse and J. M. Quattro, 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131:** 479–491.

Field, D., and C. Wills, 1998 Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. Proc. Natl. Acad. Sci. USA **95:** 1647–1652.

Frisse, L. M., 1999 Understanding the mechanisms of microsatellite formation and mutation using the model organism *Caenorhabditis elegans*. Ph.D. Dissertation, University of Missouri, Kansas City, MO.

Goldstein, D. B., and A. G. Clark, 1995 Microsatellite variation in North American populations of Drosophila melanogaster. Nucleic Acids Res. **23:** 3882–3886.

Goldstein, D. B., and C. Schlotterer, (Editors), 1999 *Microsatellites: Evolution and Applications*. Oxford University Press, Oxford.

Grandien, K., and R. J. Sommer, 2001 Functional comparison of the nematode Hox gene lin-39 in C. elegans and P. pacificus reveals evolutionary conservation of protein function despite divergence of primary sequences. Genes Dev. **15:** 2161–2172.

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.

Hennequin, C., A. Thierry, G. F. Richard, G. Lecointre, H. V. Nguyen *et al.*, 2001 Microsatellite typing as a new tool for identification of Saccharomyces cerevisiae strains. J. Clin. Microbiol. **39:** 551–559.

Hodgkin, J., and T. Doniach, 1997 Natural variation and copulatory plug formation in *Caenorhabditis elegans*. Genetics **146:** 149–164.

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

Innan, H., R. Terauchi and N. T. Miyashita, 1997 Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. Genetics **146:** 1441–1452.

Johnson, T. E., and E. W. Hutchinson, 1993 Absence of strong heterosis for life span and other life history traits in *Caenorhabditis elegans*. Genetics **134:** 465–474.

Johnstone, I. L., 1999 Molecular biology, pp. 201–225 in *C. elegans—A Practical Approach*, edited by I. A. Hope. Oxford University Press, London.

Kaplan, N., R. R. Hudson and C. H. Langley, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

Katti, M. V., P. K. Ranjekar and V. S. Gupta, 2001 Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol. Biol. Evol. **18:** 1161–1167.

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.

Kent, W. J., and A. M. Zahler, 2000 Conservation, regulation, synteny, and introns in a large-scale C. briggsae-C. elegans genomic alignment. Genome Res. **10:** 1115–1125.

Kikuchi, S., and Y. Isagi, 2002 Microsatellite genetic variation in small and isolated populations of *Magnolia sieboldii* ssp. *japonica*. Heredity **88:** 313–321.

Kimmel, M., R. Chakraborty, J. P. King, M. Bamshad, W. S. Watkins *et al.*, 1998 Signatures of population expansion in microsatellite repeat data. Genetics **148:** 1921–1930.

Kimura, M., and T. Ohta, 1975 Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. Proc. Natl. Acad. Sci. USA **72:** 2761–2764.

King, J. P., M. Kimmel and R. Chakraborty, 2000 A power analysis of microsatellite-based statistics for inferring past population growth. Mol. Biol. Evol. **17:** 1859–1868.

Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman *et al.*, 2000 The population genetics of the origin and divergence of the *Drosophila simulans* complex species. Genetics **156:** 1913–1931.

Koch, R., H. G. van Luenen, M. van der Horst, K. L. Thijssen and R. H. Plasterk, 2000 Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*. Genome Res. **10:** 1690–1696.

Krieger, M. J. B., and L. Keller, 1999 Low polymorphism at 19 microsatellite loci in a French population of Argentine ants (*Linepithema humile*). Mol. Ecol. **8:** 1078–1080.

Lanzaro, G. C., L. Zheng, Y. T. Toure, S. F. Traore, F. C. Kafatos *et al.*, 1995 Microsatellite DNA and isozyme variability in a west African population of *Anopheles gambiae*. Insect Mol. Biol. **4:** 105–112.

Mackay, T. F., 2001 Quantitative trait loci in Drosophila. Nat. Rev. Genet. **2:** 11–20.

Maloof, J. N., J. O. Borevitz, D. Weigel and J. Chory, 2000 Natural variation in phytochrome signaling. Semin. Cell Dev. Biol. **11:** 523–530.

Maloof, J. N., J. O. Borevitz, T. Dabi, J. Lutes, R. B. Nehring *et al.*, 2001 Natural variation in light sensitivity of Arabidopsis. Nat. Genet. **29:** 441–446.

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

McAllister, B. F., and G. A. McVean, 2000 Neutral evolution of the sex-determining gene transformer in Drosophila. Genetics **154:** 1711–1720.

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

Moriyama, E. N., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in Drosophila. Mol. Biol. Evol. **13:** 261–277.

Noor, M. A., M. D. Schug and C. F. Aquadro, 2000 Microsatellite variation in populations of *Drosophila pseudoobscura* and *Drosophila persimilis*. Genet. Res. **75:** 25–35.

Nordborg, M., 1997 Structured coalescent processes on different time scales. Genetics **146:** 1501–1514.

Nordborg, M., B. Charlesworth and D. Charlesworth, 1996 Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. Proc. Biol. Sci. **263:** 1033–1039.

Nordborg, M., J. O. Borevitz, J. Bergelson, C. C. Berry, J. Chory *et al.*, 2002 The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat. Genet. **30:** 190–193.

Ohta, T., and M. Kimura, 1973 A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. **22:** 201–204.

Pascual, M., M. D. Schug and C. F. Aquadro, 2000 High density of long dinucleotide microsatellites in Drosophila subobscura. Mol. Biol. Evol. **17:** 1259–1267.

Pollak, E., 1987 On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics **117:** 353–360.

Powell, J. R., and E. N. Moriyama, 1997 Evolution of codon usage bias in Drosophila. Proc. Natl. Acad. Sci. USA **94:** 7784–7790.

Purugganan, M. D., and J. I. Suddith, 1998 Molecular population genetics of the Arabidopsis CAULIFLOWER regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. Proc. Natl. Acad. Sci. USA **95:** 8130–8134.

Purugganan, M. D., and J. I. Suddith, 1999 Molecular population genetics of floral homeotic loci: departures from the equilibrium-neutral model at the *APETALA3* and *PISTILLATA* genes of *Arabidopsis thaliana*. Genetics **151:** 839–848.

Reed, K. M., and C. W. Beattie, 2001 Isolation of 105 microsatellite loci from an ovine genomic library enriched for microsatellites. Anim. Biotechnol. **12:** 77–86.

Reich, D. E., and D. B. Goldstein, 1998 Genetic evidence for a Paleolithic human population expansion in Africa. Proc. Natl. Acad. Sci. USA **95:** 8119–8123.

Reich, D., M. Feldman and D. Goldstein, 1999 Statistical properties of two tests that use multilocus data sets to detect population expansions. Mol. Biol. Evol. **16:** 453–466.

Riddle, D. L., 1988 The dauer larva, pp. 393–412 in *The Nematode Caenorhabditis elegans*, edited by W. Wood. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schlotterer, C., and B. Harr, 2000 *Drosophila virilis* has long and highly polymorphic microsatellites. Mol. Biol. Evol. **17:** 1641–1646.

Schmidt, P. S., D. D. Duvernell and W. F. Eanes, 2000 Adaptive

evolution of a candidate gene for aging in Drosophila. Proc. Natl. Acad. Sci. USA **97:** 10861–10865.

Schneider, S., D. Roessli and L. Excoffier, 2000 *Arlequin: A Software for Population Genetics Data Analysis.* Department of Anthropology, University of Geneva, Geneva.

Slatkin, M., 1995 A measure of population subdivision based on microsatellite allele frequencies. Genetics **139:** 457–462.

Stiernagle, T., 1999 Maintenance of *C. elegans*, pp. 51–68 in *C. elegans—A Practical Approach*, edited by I. A. Hope. Oxford University Press, London.

Taramino, G., and S. Tingey, 1996 Simple sequence repeats for germplasm analysis and mapping in maize. Genome **39:** 277–287.

Thuillet, A.-C., D. Bru, J. David, P. Roumet, S. Santoni *et al.*, 2002 Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp durum desf. Mol. Biol. Evol. **19:** 122–125.

Todokoro, S., R. Terauchi and S. Kawano, 1995 Microsatellite polymorphisms in natural-populations of *Arabidopsis thaliana* in Japan. Jpn. J. Genet. **70:** 543–554.

Uitterlinden, A. G., P. E. Slagboom, T. E. Johnson and J. Vijg, 1989 The *Caenorhabditis elegans* genome contains monomorphic minisatellites and simple sequences. Nucleic Acids Res. **17:** 9527–9530.

Vazquez, J. F., T. Perez, J. Albornoz and A. Dominguez, 2000 Estimation of microsatellite mutation rates in *Drosophila melanogaster*. Genet. Res. **76:** 323–326.

Weissenbach, J., G. Gyapay, C. Dib, A. Vignal, J. Morissette *et al.*, 1992 A second-generation linkage map of the human genome. Nature **359:** 794–801.

Wright, S., 1951 The genetical structure of populations. Ann. Eugen. **15:** 323–354.

Yue, G. H., P. Beeckmann and H. Geldermann, 2002 Mutation rate at swine microsatellite loci. Genetica **114:** 113–119.

Zouros, E., 1979 Mutation rates, population sizes and amounts of electrophoretic variation of enzyme loci in natural populations. Genetics **92:** 623–646.

Communicating editor: P. D. Keightley