Divergence Times in *Caenorhabditis* and *Drosophila* Inferred from Direct Estimates of the Neutral Mutation Rate

Asher D. Cutter

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

Accurate inference of the dates of common ancestry among species forms a central problem in understanding the evolutionary history of organisms. Molecular estimates of divergence time rely on the molecular evolutionary prediction that neutral mutations and substitutions occur at the same constant rate in genomes of related species. This underlies the notion of a molecular clock. Most implementations of this idea depend on paleontological calibration to infer dates of common ancestry, but taxa with poor fossil records must rely on external, potentially inappropriate, calibration with distantly related species. The classic biological models *Caenorhabditis* and *Drosophila* are examples of such problem taxa. Here, I illustrate internal calibration in these groups with direct estimates of the mutation rate from contemporary populations that are corrected for interfering effects of selection on the assumption of neutrality of substitutions. Divergence times are inferred among 6 species each of *Caenorhabditis* and *Drosophila*, based on thousands of orthologous groups of genes. I propose that the 2 closest known species of *Caenorhabditis* shared a common ancestor <24 MYA (*Caenorhabditis briggsae* and *Caenorhabditis* sp. 5) and that *Caenorhabditis elegans* diverged from its closest known relatives <30 MYA, assuming that these species pass through at least 6 generations per year; these estimates are much more recent than reported previously with molecular clock calibrations from non-nematode phyla. Dates inferred for the common ancestor of *Drosophila melanogaster* and *Drosophila simulans* are roughly concordant with previous studies. These revised dates have important implications for rates of genome evolution and the origin of self-fertilization in *Caenorhabditis*.

Introduction

A major problem in understanding the history of life is the inference of appropriate dates of common ancestry among species—the difficulty of which is compounded in taxa that lack fossil or biogeographic reference points (Bromham and Penny 2003; Kumar 2005). Paleontological evidence is absent for the nematode model organism *Caenorhabditis elegans* and its relatives, and the Melanogaster subgroup of *Drosophila* relies on biogeographic calibration of molecular divergence with distantly related Hawaiian congeners (Rowan and Hunt 1991; Russo et al. 1995; Tamura et al. 2004). In these and other taxa with poor fossil records, it would be desirable to infer dates of common ancestry without relying on distant external calibration that might inappropriately reflect evolution in the focal taxon. External calibrations are particularly problematic in *Caenorhabditis*, which appears to experience a higher mutation rate than other metazoans (Mushegian et al. 1998; Denver et al. 2004), leading to valid concerns (Félix 2004; Kiontke et al. 2004) about published divergence date estimates (Coghlan and Wolfe 2002; Stein et al. 2003). Explicit incorporation of local or relaxed molecular clock models that accommodate rate variation provide one valuable avenue for inferring divergence times for taxa with evidence of rate differences among them (Thorne et al. 1998; Huelsenbeck et al. 2000; Sanderson 2002; Drummond et al. 2006). In this study, however, I derive dates of divergence by applying internal molecular clock calibration from direct estimates of the rate of neutral mutation in contemporary laboratory populations (Denver et al. 2004; Haag-Liautard et al. 2007) to neutral substitution rates inferred for thousands of orthologs from 6 species in each of these genera.

A key prediction of the neutral theory of molecular evolution holds that the rate of substitution of neutral mutations will be independent of population size and will be equal to the rate of neutral mutation (Kimura 1968). This follows as a consequence of a diploid population producing $2N\mu$ new mutations per generation (neutral mutation rate μ and effective population size N), each with fixation probability $1/2N$ via genetic drift; the substitution rate at equilibrium is therefore $2N\mu \times 1/2N = \mu$ neutral substitutions per generation and is independent of population size. This framework forms the theoretical basis of the notion of the molecular clock (Zuckerkandl and Pauling 1962), in which the time to the most recent common ancestor (TMRCA) (T) of a pair of lineages may be inferred from DNA sequences, assuming that the neutral divergence (K) between them will simply be the product of the substitution rate (μ) and the duration of divergence, summed across both lineages: $K = 2\mu T$. Heterogeneity in divergence among lineages might be caused by differences in the per generation mutation rate or by different generation times. Note that, strictly speaking, the molecular clock prediction applies only to neutral substitutions, although it is often successfully employed to date common ancestors using protein evolution (Kumar and Hedges 1998). Typically, the TMRCA is inferred using one or more paleontological or biogeographic reference points to calibrate the relative divergence between a given set of lineages (Bromham and Penny 2003; Kumar 2005). This standard approach requires that a fossil record exists for the taxa under consideration and that historical dates of divergence can be inferred accurately. A limitation of this traditional method is that taxa with poor fossil preservation must rely on calibrations from distantly related organisms, which may experience drastically different mutation rates or generation times and therefore compromise estimates of divergence time.

Here, I demonstrate the utility of an alternative approach to calibrating a taxonomically local molecular clock that exploits the neutral mutation rate measured directly in

Key words: divergence time, *Caenorhabditis*, *Drosophila*, molecular evolution.

E-mail: asher.cutter@utoronto.ca.

Mol. Biol. Evol. 25(4):778–786. 2008
doi:10.1093/molbev/msn024
Advance Access publication January 24, 2008

species for which this parameter can be estimated in contemporary laboratory populations. A concern of applying contemporary rates of mutation to the deeper timescale of substitution is that comparative estimates of mutation rates generally are lower than those based on pedigrees or mutation accumulation experiments (Ochman et al. 1999; Ochman 2003; Ho et al. 2005; but see Emerson 2007; Bandelt 2008). Consequently, it is necessary to correct measures of divergence to account for potential effects of selection. By focusing on divergence at synonymous sites, I can account for weak selection on codon usage in order to make measures of both divergence and mutations most accurately reflect a neutral process. A second concern is that assumption of a universal molecular clock may be unjustified in analyses of deep timescales (Thorne et al. 1998; Ochman et al. 1999); indeed, this is the root of criticism for current divergence dates of *Caenorhabditis* species (Félix 2004; Kiontke et al. 2004). I minimize the potential impact of mutation rate heterogeneity among taxa by focusing on collections of closely related species and large samples of orthologous loci. By computing lineage-specific neutral substitution rates for thousands of orthologous groups of genes, I apply estimates of μ from mutation accumulation experiments in the nematode *Caenorhabditis elegans* (Denver et al. 2004) and the fruit fly *Drosophila melanogaster* (Haag-Liautard et al. 2007) to infer times to common ancestors for closely related members of these groups for which fossil calibration is lacking or limited.

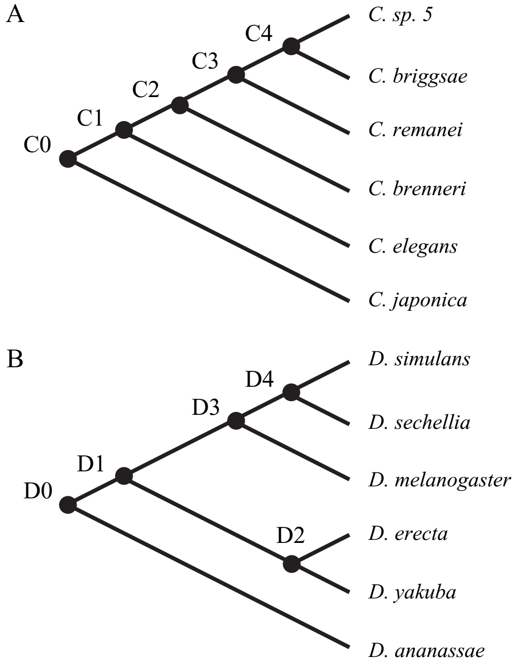

FIG. 1.—Phylogenetic relationships used in calculations of divergence for *Caenorhabditis* (*A*) and *Drosophila* (*B*). Phylogenetic topologies are derived from (Cho et al. 2004; Kiontke et al. 2004, 2007) and from (O'Grady and Kidwell 2002; Pollard et al. 2006), respectively. Divergence estimates for a lineage are calculated from extant species at the tips to the most recent common ancestral node (e.g., for the *C. elegans* lineage, from the tip to node C1; for the *D. melanogaster* lineage, from the tip to node D3).

Materials and Methods

I obtained a distribution of estimates of neutral divergence from multiple sequence alignments for putative coding sequence orthologs of 6 *Caenorhabditis* species (*Caenorhabditis japonica* DF5081, *C. elegans*, *Caenorhabditis brenneri* CB5161, *Caenorhabditis remanei*, *Caenorhabditis briggsae*, and *Caenorhabditis* sp. 5 JU727) and 6 *Drosophila* species (*Drosophila ananassae*, *Drosophila yakuba*, *Drosophila erecta*, *D. melanogaster*, *Drosophila sechellia*, and *Drosophila simulans*). The *Caenorhabditis* divergence values derive from a combination of public genome sequence gene annotations and an expressed sequence tag (EST) sequencing effort, described elsewhere (Cutter et al. 2008). Briefly, coding sequences for *C. elegans*, *C. remanei*, and *C. briggsae* were obtained from Wormbase release WS170, whereas coding sequences for the remaining species of *Caenorhabditis* were inferred from ESTs (GenBank accession numbers for *C. japonica*: FD512256–FD513938; *C. brenneri*: FD509784–FD512255; and *C.* sp. 5: FD513939–FD517806). The raw EST sequences were clustered into unique sequence objects with the PartiGene data pipeline (Parkinson, Anthony, et al. 2004), from which appropriate coding sequence translations were inferred with prot4EST (Wasmuth and Blaxter 2004) and stored in NemBase (Parkinson, Whitton, et al. 2004). The phylogeny relating these *Caenorhabditis* species was established previously (Cho et al. 2004; Kiontke et al. 2004, 2007), and divergence estimates assume this topology (fig. 1). Full-genome coding sequences for *D. melanogaster* were extracted from Flybase release 5.1; coding sequences for the remaining *Drosophila* species were obtained from the supplementary online datafiles reported by Pollard et al. (2006). I follow the previously derived phylogenetic topology for *Drosophila* in calculations of lineage-specific divergence (O'Grady and Kidwell 2002; Pollard et al. 2006; fig. 1).

Based on multiple sequence alignments in each of the focal genera, I used the codeml program in PAML to compute branch-specific synonymous-site divergence (d_S) as a measure of neutral divergence (Goldman and Yang 1994; Yang 1997). For *Drosophila*, I simply used the ortholog designations described by Pollard et al. (2006). The canonical peptide translations of these nuclear-encoded loci were used for multiple sequence alignment with ClustalW using default parameters (Thompson et al. 1994) in an automated fashion with custom Bioperl-based scripts (http://www.bioperl.org), from which branch-specific synonymous-site divergence values were calculated (permitting branch-specific d_N/d_S ratios, i.e., model = 1; codon model F3 × 4). This provided 8088 orthologous groups with representatives in all 6 species of *Drosophila*. Because the lineage-specific divergence for the outgroup species *D. ananassae* cannot be easily defined, 2 versions were computed: 1) the *D. ananassae* divergence values reported by codeml (excluding those with $d_S \leq 0.0001$ or summed $d_S > 5$) and 2) summed values for *D. ananassae* and the internal branch leading from *D. ananassae* to the common ancestor of the remaining taxa (i.e., the distance between nodes D0 and D1 in fig. 1). The true lineage-specific divergence for *D. ananassae* probably lies between these extremes. Exclusion of *D. ananassae* loci with $d_S \leq 0.0001$ or summed $d_S > 5$ led to 7397 and 7889 loci for the 2 approaches to reporting *D. ananassae* divergence.

As described elsewhere (Cutter et al. 2008), for *Caenorhabditis* species, putative orthologs were inferred with OrthoMCL (Li et al. 2003) through its reciprocal best-hit Blast procedure on canonical peptide translations of the genes (and EST gene fragments). The same procedure described above for *Drosophila* species was then followed for alignment and calculation of branch-specific divergence. The EST collections do not contain representatives for every gene in the genome, so I checked the putative orthologs for evidence of inappropriate orthology classification, based on instances of exceptionally high divergence ($d_N > 0.5$ or $d_S > 5$), and removed such cases from further analysis. This procedure yielded 63 orthologous groups with representatives in all 6 *Caenorhabditis* species, 6,398 orthologous groups specific to *C. elegans–C. remanei–C. briggsae*, and 1,244 orthologous groups with other configurations of 3 or more taxa. For groups of orthologs with incomplete species membership, I used lineage-specific divergence values for only those lineages that led to an ancestral node in the full 6 species phylogeny (e.g., in the *C. elegans–C. remanei–C. briggsae* comparison, only the *C. remanei* divergence values are retained). As for the outgroup species in *Drosophila*, divergence for the outgroup species *C. japonica* was computed with or without summation with the internal branch leading from *C. japonica* to the common ancestor of *C. elegans* and other taxa (i.e., from *C. japonica* to C0 plus the distance between nodes C0 and C1 in fig. 1), excluding loci with values of $d_S = 0$. The true lineage-specific divergence for *C. japonica* probably lies between these extremes. It should be noted that sequence divergence at synonymous sites among *Caenorhabditis* species is saturated in pairwise comparisons (i.e., >1 substitution is expected to have occurred at each synonymous site). To the extent that multiple hits corrections over- or undercompensate for the nonlinear relationship between the number of sequence differences and the number of substitutions, the inferences of divergence time will be directly affected. However, the codon-based maximum likelihood method used here for estimating synonymous-site divergence (Goldman and Yang 1994) performs well (see simulation procedure and results, below).

Because selection for codon bias is evident in both *Caenorhabditis* (Stenico et al. 1994; Duret and Mouchiroud 1999; Cutter and Charlesworth 2006; Cutter et al. 2006) and *Drosophila* (Akashi 1995; Duret and Mouchiroud 1999), the calculations of d_S might not accurately reflect the neutral substitution rate. Therefore, I checked for correlations between lineage-specific d_S and codon bias (effective number of codons, N_c, Wright 1990; J. Peden's program codonW). Most *Drosophila* species demonstrated weak, albeit statistically significant, associations between d_S and N_c (Spearman's nonparametric correlations: *D. sechellia* = −0.08***, *D. simulans* = 0.03*, *D. melanogaster* = −0.11***, *D. erecta* = 0.07***, *D. yakuba* = 0.09***, and *D. ananassae* = −0.16***; ***$P < 0.0001$, *$P < 0.05$). However, because of the small magnitude of the correlations, their inconsistent sign (a positive correlation is expected under selection for codon bias), and previous discussion of how this codon-based method for measuring divergence is generally uncorrelated with codon bias in *D. melanogaster* (Goldman and Yang 1994; Bierne and Eyre-Walker 2003), I do not consider the potential for selection on synonymous sites further in the analysis of *Drosophila* species. Codon bias is relevant to this study only in its potential to lead to artificially low estimates of synonymous-site divergence for loci that experience selection on codon usage. However, this maximum likelihood method generally produces higher divergence estimates than other approaches (Yang 2006), and the lack of strong association with measures of codon bias therefore makes these calculations conservative with respect to underestimating sequence divergence along lineages, as compared with other methods of estimating divergence.

In contrast to *Drosophila* sequences, divergence values for the orthologous groups in *Caenorhabditis* demonstrated strong correlations with codon bias (Spearman's nonparametric correlations: *C. japonica* + internal branch = 0.50, *C. japonica* = 0.24, *C. elegans* = 0.55, *C. brenneri* = 0.60, *C. remanei* = 0.26, *C. briggsae* = 0.58, and *C.* sp. 5 = 0.51; all $P < 0.0001$). This strong effect could be in part due to EST collections disproportionately containing genes with high expression and correspondingly high codon bias due to selection for translational efficiency and/or accuracy (Shields and Sharp 1987). Consequently, synonymous divergence values were adjusted for each species by adding the residuals of a least-squares regression of branch-specific d_S on N_c to the expected d_S for $N_c = 61$ (where codon bias is absent). This adjusted measure of synonymous-site divergence ($d_S{}'$) should reflect neutral patterns of substitution by removing the effects of selection for preferred codons at synonymous sites.

To confirm the accuracy of the method of calculating divergence for *Caenorhabditis*, because the inferred estimates of $d_S{}'$ imply saturation in pairwise comparisons, I simulated 1,000 orthologous groups of sequences for 6 species using the PAML program evolver. Simulation input parameters were selected to mimic the empirical data set. Each sequence contained 300 codons (median length for the empirical data set = 294 codons) and was simulated assuming a transition/transversion ratio of 2.0 and a d_N/d_S ratio of 0.029 (the average $d_N/d_S{}'$ across species), with codon frequencies calculated according to the distribution observed for the coding genome of *C. elegans* version WS170. The tree topology of figure 1 was applied, with branch lengths defined as the number of substitutions per codon for branch i as $L_i = 3 \cdot d_S' \cdot p_S + 3 \cdot d_N \cdot (1 - p_S)$ (Anisimova et al. 2001), where p_S is the fraction of synonymous sites, which was assumed to be 0.25. Observed d_N and $d_S{}'$ for each species were used to calculate L for its respective lineage. Because the true estimate of $d_S{}'$ for *C. japonica* is unknown, I arbitrarily used a value intermediate between the extremes calculated in this study (table 1). The simulated sequence alignments were then passed through codeml, in the same way as for the empirical sequences, to estimate synonymous-site and replacement-site divergence values.

I calculated a range of TMRCA values for species of *Caenorhabditis* and *Drosophila* given the putatively

Table 1
Divergence Estimates for *Caenorhabditis* Lineages from Extant Species to Their Most Recent Common Ancestor

Lineage	Loci	Median d_S	Median d_S' (First–Ninth Decile)	T (Generation × 10^6)	$T_{14\ \mathrm{days}}$ (Year × 10^6)	$T_{30\ \mathrm{days}}$ (Year × 10^6)	$T_{60\ \mathrm{days}}$ (Year × 10^6)
Caenorhabditis japonica + internal branch	354	1.018	1.664 (1.194–2.802)	184.9	7.09	15.20	30.40
Caenorhabditis japonica	308	0.660	1.029 (0.396–2.118)	114.3	4.38	9.40	18.79
Caenorhabditis elegans	356	0.523	1.017 (0.674–1.527)	113.0	4.33	9.29	18.58
Caenorhabditis brenneri	430	0.518	1.027 (0.666–1.543)	114.1	4.38	9.38	18.76
Caenorhabditis remanei	7662	0.637	0.949 (0.592–1.613)	105.4	4.04	8.67	17.33
Caenorhabditis briggsae	823	0.478	0.809 (0.571–1.207)	89.9	3.45	7.39	14.78
Caenorhabditis sp. 5	819	0.415	0.730 (0.481–1.189)	81.1	3.11	6.66	13.32

neutral divergence at synonymous sites (d_S' or d_S, respectively), incorporating uncertainty in generation time and standard errors for estimates of the neutral mutation rate. Under a standard neutral, infinite sites model of mutation, the mutation rate (μ) equals the substitution rate (K) (Kimura 1968). Thus, the TMRCA for a pair of sequences is simply $K/(2\mu)$ and for a single branch of a phylogeny the divergence time is K/μ. For *Caenorhabditis* species, the mutation rate estimate (μ) for *C. elegans* of $\mu = 9.0 \times 10^{-9}$ single nucleotide mutations per site per generation on average (standard error of the mean [SEM] = 6×10^{-10}) (Denver et al. 2004; Denver D, personal communication) was used to calibrate the molecular clock for the genus. For *Drosophila*, I applied the single nucleotide mutation rate estimate from *D. melanogaster* of $\mu = 5.8 \times 10^{-9}$ mutations per site per generation on average (SEM = 6×10^{-10}) (Haag-Liautard et al. 2007). Both of these mutation rate estimates derive from direct measurement of new nucleotide sequence differences that accrued over the course of several hundred generations in mutation accumulation lines of *C. elegans* and *D. melanogaster*, respectively. Although there is good reason to expect these laboratory-based estimates of average mutation rate to closely reflect mutation rates in nature, it is important to recognize that the average mutation rate in nature might differ. In particular, longer generation times (more opportunity for oxidative damage to DNA) and environmental insults could contribute to higher average per-generation mutation rates in nature, which would lead to overestimation of divergence time based on laboratory mutation rates. Median values of the distribution of selection-adjusted synonymous-site divergence were used as point estimates of K, and the 10th and 90th percentiles were used as lower and upper bounds. Upper and lower bounds on the mutation rate estimates were taken as ±1 SEM, the results of which are given in supplementary tables 1 and 2 (Supplementary Material online). The resulting TMRCA estimates measure time in units of generations, so a range of plausible generation times was applied to infer TMRCA in units of years. Under benign laboratory conditions, *C. elegans* generation time is 2–6 days, depending on temperature (Wood 1988). However, *Caenorhabditis* are generally found in nature as quiescent dauer larvae (Barrière and Félix 2005; Barrière and Félix 2007), which may persist for months; the true number of generations passed through each year in these species is not known. I conservatively propose an average generation time of 60 days (~6 generations per year) but provide results for calculations of divergence time that assume a 14-day and 30-day turnover. Calculations based on even slower generation times are given in supplementary table 1 (Supplementary Material online). It is commonly assumed that *Drosophila* species experience approximately 10 generations per year, although they may undergo >20 generations per year in laboratory culture. It is likely that *Drosophila* pass through at least 5 generations per year, which I use as a lower limit in calculations of divergence time.

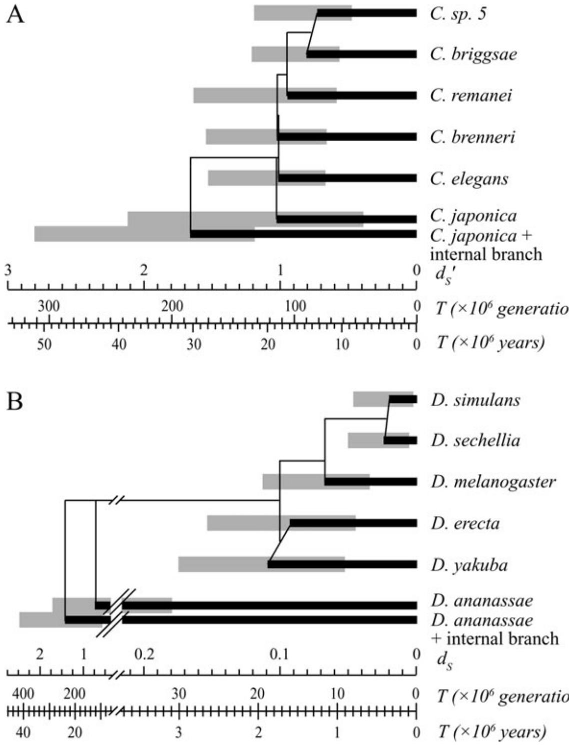

FIG. 2.—Divergence along lineages leading to extant species of *Caenorhabditis* (*A*) and *Drosophila* (*B*). The lengths of the thick, external branches correspond to median d_S' (*A*) or d_S (*B*) values, and the gray boxes indicate the bounds of the 10th and 90th percentiles. Internal branches serve only to graphically connect the external branches in the figure; angled lines that connect lineages simply reflect different lineage-specific median divergence values given that branch tips are fixed at zero (the present day). In (*A*), timescale in generations assumes equal substitution and mutation rates of 9×10^{-9} per generation; timescale in years further assumes 60 days per generation. In (*B*), the timescale in generations assumes equal substitution and mutation rates of 5.8×10^{-9} per generation; timescale in years further assumes 10 generations per year.

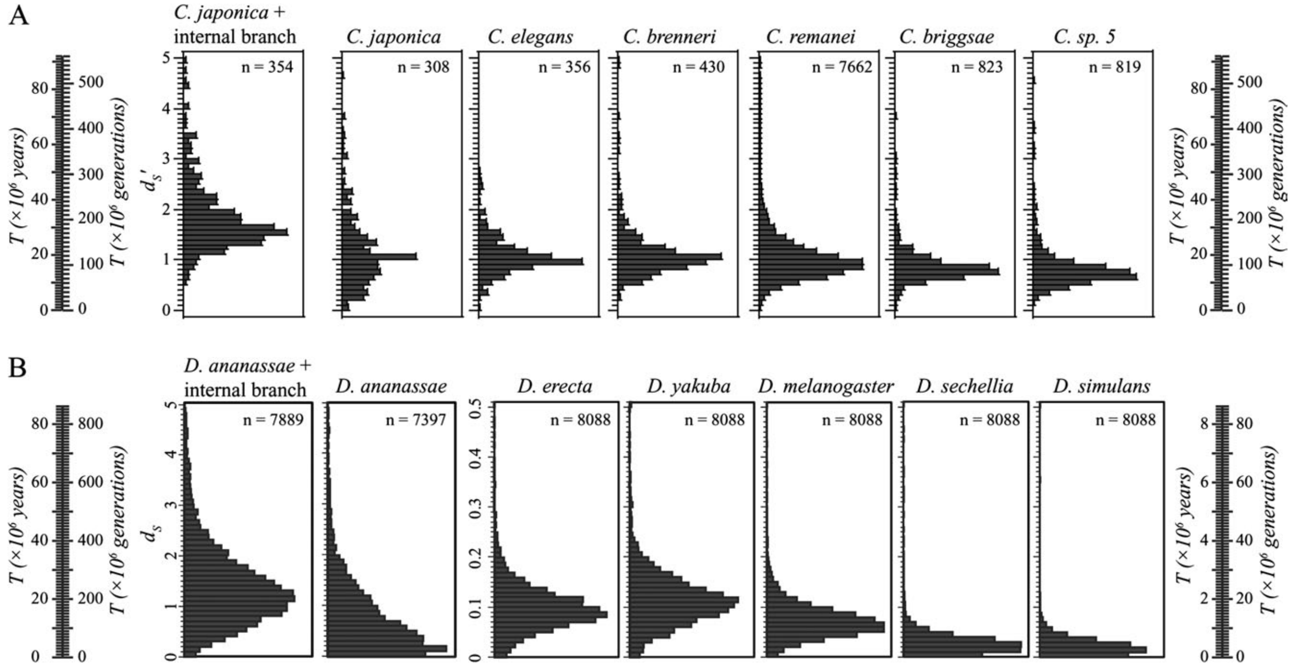

FIG. 3.—Distribution of lineage-specific synonymous-site substitution rates and their corresponding divergence times for species of *Caenorhabditis* (*A*) and *Drosophila* (*B*). In (*A*), timescale in generations assumes equal substitution and mutation rates of 9 × 10^{-9} per generation; timescale in years further assumes 60 days per generation. In (*B*), the timescale in generations assumes equal substitution and mutation rates of 5.8 × 10^{-9} per generation; timescale in years further assumes 10 generations per year. The few points corresponding to $0.5 < d_S < 5$ are excluded from (*B*) for clarity. Note the different scales for *Drosophila ananassae* relative to the other *Drosophila*. Sample size, *n*, is indicated in each panel.

Results and Discussion

I estimated neutral divergence from lineage-specific rates of synonymous-site substitution (d_S' or d_S) for orthologous groups of genes in *Caenorhabditis* and *Drosophila* (table 1; figs. 2 and 3). Lineage ages (*T*) were inferred by applying median synonymous-site divergence values and direct measures of the average per-site mutation rate in *C. elegans* ($\mu = 9.0 \times 10^{-9}$ mutations per generation) (Denver et al. 2004) and *D. melanogaster* ($\mu = 5.8 \times 10^{-9}$ mutations per generation) (Haag-Liautard et al. 2007) to the prediction from the neutral theory of molecular evolution that $d_S = \mu T$ (table 1; figs. 2 and 3). Contemporary measures of the mutation rate in these 2 species provide the most proximate calibration of molecular divergence in these groups, given the phylogenetic distance to other taxa for which fossil or biogeographic calibrations could be made. For *Caenorhabditis*, I adjusted the divergence values upward, accounting for selection on codon usage in order to accurately reflect patterns of neutral substitution; this correction was unnecessary in *Drosophila*.

For *C. elegans*, the median selection-adjusted, lineage-specific divergence ($d_S' = 1.017$; table 1) translates to an estimated common ancestry date of 113.0 × 10^6 generations ago (table 1). Although generations time in nature is unknown for *Caenorhabditis* species, the long-lived dauer stage likely predominates, leading to few generations per year relative to the laboratory (Barrière and Félix 2005; Cutter 2006; Barrière and Félix 2007). I conservatively propose a 60-day average generation time (~6 generations per year), which yields a point estimate of the TMRCA for *C. elegans* and its sister clade (including *C. briggsae*) of ~18 MYA and for the 2 closest relatives (*C. briggsae* and *C.* sp. 5) of ~14 MYA (supplementary table 1). Calculations using values of μ ± 1 SEM and extreme deciles of d_S' span a range of TMRCAs 11.6–29.9 MYA for the *C. elegans* lineage (table 1); however, ruling out any time 5–30 MYA is difficult given uncertainty in generation time, with faster generation times consistent with shorter TMRCAs (fig. 4). Median divergence values for other lineages suggest dates of common ancestry of 81.1–114.3 × 10^6 generations ago, depending on the lineage (table 1; figs. 2 and 3). The times to common ancestry among taxa suggest that speciation occurred in their history over a relatively short interval, as proposed previously (Fitch et al. 1995). Some TMRCAs could be overestimated if certain lineages experienced higher mutation rates (Baer et al. 2005) or for species inhabiting warmer climates (e.g., *C. briggsae* and *C. brenneri*, Kiontke and Sudhaus 2006; Sudhaus and Kiontke 2007) that might turn over generations faster than cool region species.

These *Caenorhabditis* divergence dates are substantially more recent than reported for analyses dependent on nonnematode calibration (30–180 MYA, Prasad and Baillie 1989; Kennedy et al. 1993; Coghlan and Wolfe 2002; Stein et al. 2003) (but see Heschl and Baillie 1990, 23–32 MYA) but accord with arguments that a ~100 MYA TMRCA is too long ago (Félix 2004; Kiontke et al. 2004). Because nematodes appear to experience a higher mutation rate than other eukaryotes (Mushegian et al. 1998; Denver et al. 2004), also reflected in long-branch lengths typical for nematode lineages in metazoan phylogenies (Aguinaldo et al. 1997; Peterson and Eernisse 2001; Wolf et al. 2004; Philippe et al. 2005),

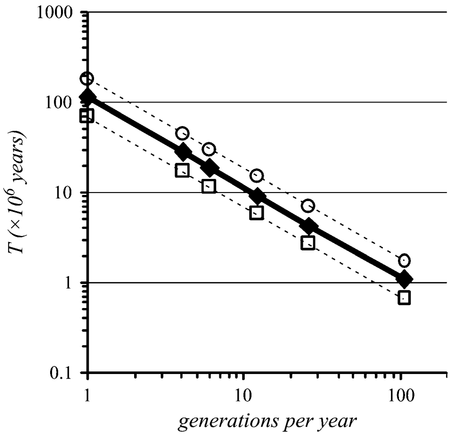

FIG. 4.—Calculated divergence times as a function of the assumed generation time for the *Caenorhabditis elegans* lineage leading to the most recent common ancestor with other species. Solid diamonds correspond to median $d_S{}'$ values; open squares and circles correspond to, respectively, 10th and 90th percentiles with mutation rate +1 SEM and −1 SEM.

one must assume an implausibly slow 1 generation per year to reconcile a 100 MYA TMRCA (fig. 4). Furthermore, given the conservation of chromosomal synteny between different *Caenorhabditis* species (Hillier et al. 2007), partial development of interspecific hybrids (Baird et al. 1992; Geldziler et al. 2006), and conservation of female-secreted mate attraction signals (Chasnov et al. 2007), the dates suggested here are more biologically plausible than very ancient common ancestry. However, these revised TMRCA estimates imply that rates of genome rearrangement (Coghlan and Wolfe 2002; Stein et al. 2003), intron gain and loss (Gotoh 1998; Robertson 1998; Cho et al. 2004; Coghlan and Wolfe 2004; Kiontke et al. 2004), and gene family dynamics (Robertson 1998; Lynch and Conery 2000, 2003; Katju and Lynch 2003; Cho et al. 2004; Thomas et al. 2005; Thomas 2006) are even greater than previously realized. These revised date estimates also narrow the maximum potential duration of self-fertilization in the lineages leading to *C. elegans* and to *C. briggsae*.

Because synonymous-site divergence for the *Caenorhabditis* species shows evidence of saturation (i.e., pairwise $d_S{}' > 1$), I conducted simulations to assess the robustness of the method of estimating sequence divergence. Using input parameter values matched to those observed for *Caenorhabditis*, the simulations of DNA sequence evolution indicate that the method of calculating divergence generally performs well (table 2). Simulated synonymous-site divergence values tend to be slightly underestimated, but all median estimates for within-group taxa fall within ~10% of the true simulated value. Thus, the ~5-fold more recent divergence times reported here relative to other studies are not an artifact of extreme underestimation of neutral substitution rates.

Based on 8088 orthologs shared between 6 *Drosophila* species, I calculate from a median $d_S = 0.068$ that *D. melanogaster* diverged from its common ancestor with *D. simulans* and *D. sechellia* approximately 11.7 × 10^6 generations ago (table 3; figs. 2 and 3). Other *Drosophila* lineages have TMRCAs of 3.6–18.9 × 10^6 generations ago (except for the much more divergent *D. ananassae*; table 3). The nominally longer median lineage-specific divergence for *D. yakuba* relative to its sister species *D. erecta* (fig. 2) implies that *D. yakuba* might pass through more generations per year or potentially experience an elevated mutation rate, although it is conceivable that widespread incomplete lineage sorting could contribute to a similar pattern (Pollard et al. 2006). Assuming 10 generations per year leads to a comparable or somewhat more recent TMRCA between *D. melanogaster* and *D. simulans* (~1.17 MYA) than is generally believed (0.8–5.4 MYA; Caccone et al. 1988; Russo et al. 1995; Li et al. 1999; Tamura et al. 2004). However, the disparity with other studies in date estimates is greater for the common ancestor of *D. melanogaster* and *D. yakuba*. Discrepancy could result from several factors: 1) slower generation turnover than assumed, 2) unaccounted selection on synonymous sites, 3) upwardly biased mutation rate estimate, and/or 4) mutation rate difference between melanogaster group species and taxa used for external molecular clock calibration. Issue 1) cannot fully compensate for the difference (table 3) and 2) is unlikely here because the method of calculating d_S should preclude such an effect by accounting for selection on codon bias and by yielding higher divergence estimates than other methods (Goldman and Yang 1994; Bierne and Eyre-Walker 2003); also, we observe no strong, consistent association between d_S and codon bias in *Drosophila* (see Materials and Methods). In addition, synonymous sites in *Drosophila* experience less selective constraint than noncoding sites (Andolfatto 2005). Issue 3) could arise if mutation rates differ at coding and noncoding sites, perhaps as a by-product of nucleotide-content differences, although there is no evidence for such an effect (Haag-Liautard et al. 2007). The lower bound on mutation rate (Haag-Liautard et al. 2007) is consistent with a *D. melanogaster*–*D. simulans* divergence of up to 4.3 MYA (table 3), which reconciles easily with traditional dates. Note also that the

Table 2
Results of Simulation Estimates of Divergence

Lineage for Comparison	Loci	Assumed Branch Length (L)	Assumed Lineage-Specific d_S	Median Estimated d_S (25th–75th Percentile)
Caenorhabditis japonica + internal branch	1000	1.103	1.384	1.184 (0.990–1.374)
Caenorhabditis elegans	1000	0.827	1.017	0.923 (0.773–1.099)
Caenorhabditis brenneri	1000	0.831	1.027	0.926 (0.781–1.078)
Caenorhabditis remanei	1000	0.777	0.949	0.870 (0.749–1.016)
Caenorhabditis briggsae	1000	0.678	0.809	0.778 (0.662–0.918)
Caenorhabditis sp. 5	1000	0.613	0.730	0.708 (0.599–0.835)

Table 3
Divergence Estimates for *Drosophila* Lineages from Extant Species to Their Most Recent Common Ancestor

Lineage	Loci	Median d_S (First–Ninth Decile)	T (Generation × 10^6)	$T_{5\ \text{generations per year}}$ (Year × 10^6)	$T_{10\ \text{generations per year}}$ (Year × 10^6)	$T_{20\ \text{generations per year}}$ (Year × 10^6)
Drosophila ananassae + internal branch	7889	1.365 (0.637–2.501)	235	47.07	23.53	11.77
D. ananassae	7397	0.662 (0.180–1.760)	114	22.83	11.42	5.71
Drosophila erecta	8088	0.094 (0.046–0.154)	16.1	3.22	1.61	0.81
Drosophila yakuba	8088	0.110 (0.054–0.175)	18.9	3.79	1.89	0.95
Drosophila melanogaster	8088	0.068 (0.035–0.113)	11.7	2.35	1.17	0.59
Drosophila sechellia	8088	0.025 (0.006–0.050)	4.2	0.85	0.42	0.21
Drosophila simulans	8088	0.021 (0.003–0.046)	3.6	0.71	0.36	0.18

estimates of the mutation rate and neutral divergence reported here are somewhat higher than assumed in previous studies (Li et al. 1999; Tamura et al. 2004); these 2 features partially cancel out, leading to the roughly similar divergence date estimates for *D. melanogaster–D. simulans*. Determining whether contemporary mutation rates or biogeographic calibrations from Hawaiian taxa are most appropriate for the Melanogaster subgroup remains a challenge.

How confident can we be in the accuracy of divergence dates inferred from calibration of a molecular clock based on contemporary rates of mutation? It has been argued that de novo mutation rates generally exceed substitution rates inferred from phylogenies (Ochman et al. 1999; Ochman 2003; Ho et al. 2005), so population-based measures of mutation should be applied cautiously to deeper timescales (Ho and Larson 2006). I have minimized the potential problems in this context by focusing on synonymous sites for which the molecular clock hypothesis is most appropriate (rather than amino acid substitutions), by correcting synonymous-site divergence for the effects of selection, by using the median values for thousands of orthologous genes, and by limiting the scope of taxa to close relatives.

Conclusions

For taxa that lack divergence time landmarks from fossils or biogeography, such as nematodes, direct estimates of the mutation rate from contemporary laboratory populations provide the best available tool for calibrating molecular clocks. With the technical ability to determine contemporary mutation rates at hand, this approach is now feasible in a variety of organisms. *Caenorhabditis* exemplifies this method's merits, in which dates of common ancestry derived from genome-scale data sets suggest that *Caenorhabditis* species diverged from one another much more recently than inferred from external calibration points and the dubious assumption of a universal molecular clock (e.g., TMRCA of mammals, insects, and nematodes; Coghlan and Wolfe 2002). This result is conservative with respect to assumptions about generation time and methods of calculating divergence and corresponds more plausibly with a variety of biological phenomena.

Supplementary Material

Supplementary tables 1 and 2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

I thank the Washington University School of Medicine Genome Sequencing Center for making *C. remanei* genome sequences publicly available. I am grateful to A. Eyre-Walker, M.A. Félix, and several anonymous reviewers for critical comments on the manuscript. This work was supported by funds from the National Science Foundation (Doctoral Dissertation Improvement Grant and International Research Fellowship Program grant 0401897) and by startup funds from the Department of Ecology and Evolutionary Biology at the University of Toronto.

Literature Cited

Aguinaldo AMA, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature. 387:489–493.

Akashi H. 1995. Inferring weak selection from patterns of polymorphism and divergence at silent sites in *Drosophila* DNA. Genetics. 139:1067–1076.

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. Nature. 437:1149–1152.

Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol. 18:1585–1592.

Baer CF, Shaw F, Steding C, et al. (11 co-authors). 2005. Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. Proc Natl Acad Sci USA. 102:5785–5790.

Baird SE, Sutherlin ME, Emmons SW. 1992. Reproductive isolation in Rhabditidae (Nematoda, Secernentea): mechanisms that isolate 6 species of 3 genera. Evolution. 46:585–594.

Bandelt HJ. 2008. Clock debate: when times are a-changin': time dependency of molecular rate estimates: tempest in a teacup. Heredity. 100:1–2.

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

Barrière A, Félix MA. 2007. Temporal dynamics and linkage disequilibrium in natural *Caenorhabditis elegans* populations. Genetics. 176:999–1011.

Bierne N, Eyre-Walker A. 2003. The problem of counting sites in the estimation of the synonymous and nonsynonymous substitution rates: implications for the correlation between the synonymous substitution rate and codon usage bias. Genetics. 165:1587–1597.

Bromham L, Penny D. 2003. The modern molecular clock. Nat Rev Genet. 4:216–224.

Caccone A, Amato GD, Powell JR. 1988. Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics. 118:671–683.

Chasnov JR, So WK, Chan CM, Chow KL. 2007. The species, sex, and stage specificity of a *Caenorhabditis* sex pheromone. Proc Natl Acad Sci USA. 104:6730–6735.

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

Coghlan A, Wolfe KH. 2002. Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. Genome Res. 12:857–867.

Coghlan A, Wolfe KH. 2004. Origins of recently gained introns in *Caenorhabditis*. Proc Natl Acad Sci USA. 101:11362–11367.

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

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

Cutter AD, Wasmuth J, Blaxter ML. 2006. The evolution of biased codon and amino acid usage in nematode genomes. Mol Biol Evol. 23:2303–2315.

Cutter AD, Wasmuth JD, Washington NL. Forthcoming 2008. Patterns of molecular evolution in *Caenorhabditis* preclude ancient origins of selfing. Genetics.

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

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88.

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

Emerson BC. 2007. Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. Syst Biol. 56:337–345.

Félix MA. 2004. Genomes: a helpful cousin for our favourite worm. Curr Biol. 14:R75–R77.

Fitch DHA, Bugajgaweda B, Emmons SW. 1995. 18S ribosomal-RNA gene phylogeny for some Rhabditidae related to *Caenorhabditis*. Mol Biol Evol. 12:346–358.

Geldziler B, Chatterjee I, Kadandale P, Putiri E, Patel R, Singson A. 2006. A comparative study of sperm morphology, cytology and activation in *Caenorhabditis elegans*, *Caenorhabditis remanei* and *Caenorhabditis briggsae*. Dev Genes Evol. 216:198–208.

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

Gotoh O. 1998. Divergent structures of *Caenorhabditis elegans* cytochrome P450 genes suggest the frequent loss and gain of introns during the evolution of nematodes. Mol Biol Evol. 15:1447–1459.

Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Charlesworth B, Keightley PD. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. Nature. 445:82–85.

Heschl MFP, Baillie DL. 1990. Functional elements and domains inferred from sequence comparisons of a heat-shock gene in 2 nematodes. J Mol Evol. 31:3–9.

Hillier LW, Miller RD, Baird SE, Chinwalla A, Fulton LA, Koboldt DC, Waterston RH. 2007. Comparison of *C. elegans* and *C. briggsae* genome sequences reveals extensive conservation of chromosome organization and synteny. PLoS Biol. 5:e167.

Ho SY, Larson G. 2006. Molecular clocks: when times are a-changin'. Trends Genet. 22:79–83.

Ho SY, Phillips MJ, Cooper A, Drummond AJ. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol. 22:1561–1568.

Huelsenbeck JP, Larget B, Swofford D. 2000. A compound Poisson process for relaxing the molecular clock. Genetics. 154:1879–1892.

Katju V, Lynch M. 2003. The structure and early evolution of recently arisen gene duplicates in the *Caenorhabditis elegans* genome. Genetics. 165:1793–1803.

Kennedy BP, Aamodt EJ, Allen FL, Chung MA, Heschl MFP, McGhee JD. 1993. The gut esterase gene (*ges-1*) from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. J Mol Biol. 229:890–908.

Kimura M. 1968. Evolutionary rate at molecular level. Nature. 217:624–626.

Kiontke K, Barriere A, Kolotuev I, Podbilewicz B, Sommer R, Fitch DH, Felix MA. 2007. Trends, stasis, and drift in the evolution of nematode vulva development. Curr Biol. 17:1925–1937.

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

Kiontke K, Sudhaus W. 2006. Ecology of *Caenorhabditis* species (January 09, 2006). In: Fitch DHA, editor. Wormbook: The *C. elegans* Research Community. [Internet]. Available from: http://www.wormbook.org. doi/10.1895/wormbook.1.37.1

Kumar S. 2005. Molecular clocks: four decades of evolution. Nat Rev Genet. 6:654–662.

Kumar S, Hedges SB. 1998. A molecular timescale for vertebrate evolution. Nature. 392:917–920.

Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178–2189.

Li YJ, Satta Y, Takahata N. 1999. Paleo-demography of the *Drosophila melanogaster* subgroup: application of the maximum likelihood method. Genes Genet Syst. 74:117–127.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science. 290:1151–1155.

Lynch M, Conery JS. 2003. The evolutionary demography of duplicate genes. J Struct Funct Genomics. 3:35–44.

Mushegian AR, Garey JR, Martin J, Liu LX. 1998. Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes. Genome Res. 8:590–598.

O'Grady PM, Kidwell MG. 2002. Phylogeny of the subgenus sophophora (Diptera: Drosophilidae) based on combined analysis of nuclear and mitochondrial sequences. Mol Phylogenet Evol. 22:442–453.

Ochman H. 2003. Neutral mutations and neutral substitutions in bacterial genomes. Mol Biol Evol. 20:2091–2096.

Ochman H, Elwyn S, Moran NA. 1999. Calibrating bacterial evolution. Proc Natl Acad Sci USA. 96:12638–12643.

Parkinson J, Anthony A, Wasmuth J, Schmid R, Hedley A, Blaxter M. 2004. PartiGene—constructing partial genomes. Bioinformatics. 20:1398–1404.

Parkinson J, Whitton C, Schmid R, Thomson M, Blaxter M. 2004. NEMBASE: a resource for parasitic nematode ESTs. Nucleic Acids Res. 32p. D427–D430.

Peterson KJ, Eernisse DJ. 2001. Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. Evol Dev. 3:170–205.

Philippe H, Lartillot N, Brinkmann H. 2005. Multigene analyses of bilaterian animals corroborate the monophyly of

Ecdysozoa, Lophotrochozoa and Protostomia. Mol Biol Evol. 22:1246–1253.

Pollard DA, Iyer VN, Moses AM, Eisen MB. 2006. Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. PLoS Genet. 2:e173.

Prasad SS, Baillie DL. 1989. Evolutionarily conserved coding sequences in the *dpy-20-unc-22* region of *Caenorhabditis elegans*. Genomics. 5:185–198.

Robertson HM. 1998. Two large families of chemoreceptor genes in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* reveal extensive gene duplication, diversification, movement, and intron loss. Genome Res. 8:449–463.

Rowan RG, Hunt JA. 1991. Rates of DNA change and phylogeny from the DNA sequences of the alcohol dehydrogenase gene for five closely related species of Hawaiian *Drosophila*. Mol Biol Evol. 8:49–70.

Russo CA, Takezaki N, Nei M. 1995. Molecular phylogeny and divergence times of drosophilid species. Mol Biol Evol. 12:391–404.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol. 19:101–109.

Shields DC, Sharp PM. 1987. Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. Nucleic Acids Res. 15:8023–8040.

Stein LD, Bao Z, Blasiar D, et al. (36 co-authors). 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. 1:166–192.

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

Sudhaus W, Kiontke K. 2007. Comparison of the cryptic nematode species *Caenorhabditis brenneri* sp. n. and *C. remanei* (Nematoda: Rhabditidae) with the stem species pattern of the *Caenorhabditis elegans* group. Zootaxa. 1456:45–62.

Tamura K, Subramanian S, Kumar S. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. Mol Biol Evol. 21:36–44.

Thomas JH. 2006. Concerted evolution of two novel protein families in *Caenorhabditis* species. Genetics. 172:2269–2281.

Thomas JH, Kelley JL, Robertson HM, Ly K, Swanson WJ. 2005. Adaptive evolution in the SRZ chemoreceptor families of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Proc Natl Acad Sci USA. 102:4476–4481.

Thompson JD, Higgins DG, Gibson TJ. 1994. Clustal-W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol. 15:1647–1657.

Wasmuth J, Blaxter M. 2004. prot4EST: translating expressed sequence tags from neglected genomes. BMC Bioinformatics. 5:187.

Wolf YI, Rogozin IB, Koonin EV. 2004. Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis. Genome Res. 14:29–36.

Wood WB. 1988. The nematode *Caenorhabditis elegans*. New York: Cold Spring Harbor Laboratory Press.

Wright F. 1990. The effective number of codons used in a gene. Gene. 87:23–29.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13:555–556.

Yang Z. 2006. Computational molecular evolution. New York: Oxford University Press.

Zuckerkandl E, Pauling L. 1962. Molecular disease, evolution, and genetic heterogeneity. In: Marsha M, Pullman B, editors. Horizons in biochemistry. New York: Academic Press. p. 189–225.

John H. McDonald, Associate Editor

Accepted January 21, 2008