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 (Coglian 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.

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

E-mail: ashercutter@utoronto.ca.

doi:10.1093/molbev/msn024
Advance Access publication January 24, 2008

© The Author 2008. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org
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 $\mu$ 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.

**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). Divergence time estimates for Caenorhabditis species were obtained 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). 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: FDS152256–FDS153938; C. brenneri: FDS509784–FDS152255; and C. sp. 5: FDS513939–FDS517806). 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, Whotton, 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 data files 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), 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).

**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).**
divergence for *D. ananassae* probably lies between these extremes. Exclusion of *D. ananassae* loci with $d_S < 0.0001$ 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 ($d_S > 0.5$ 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 correct 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; Birne 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. brevieri* = 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 $d_S$/$d_S$ ratio of 0.029 (the average $d_S$/$d_S'$ 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 $L_i = 3 \cdot d_S \cdot p_S + 3 \cdot d_S' \cdot (1 - p_S)$ (Anisimova et al. 2001), where $p_S$ is the fraction of synonymous sites, which was assumed to be 0.25. Observed $d_S$ and $d_S'$ for each species were used to calculate $L_i$ 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
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 \((\mu)\) 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 \(K/\mu\). For \(C.\ elegans\) species, the mutation rate estimate \((\mu)\) 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 \times 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 \times 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 \(\pm 1\) 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.

![Divergence Times in Caenorhabditis and Drosophila](http://mbe.oxfordjournals.org/)

**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 \(9 \times 10^{-9}\) per generation; timescale in years further assumes 60 days per generation. In (\(B\)), the timescale in generations assumes equal substitution and mutation rates of \(5.8 \times 10^{-9}\) per generation; timescale in years further assumes 10 generations per year.

**Table 1**

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Median Loci</th>
<th>Median (d_S^*) (First–Ninth Decile)</th>
<th>Median (d_S) (First–Ninth Decile)</th>
<th>(T) (Generation (\times 10^6))</th>
<th>(T_{90\text{ days}}) (Year (\times 10^6))</th>
<th>(T_{30\text{ days}}) (Year (\times 10^6))</th>
<th>(T_{90\text{ days}}) (Year (\times 10^6))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C.\ sp. 5) + internal branch</td>
<td>354</td>
<td>1.018</td>
<td>1.664 (1.194–2.802)</td>
<td>184.9</td>
<td>7.09</td>
<td>15.20</td>
<td>30.40</td>
</tr>
<tr>
<td>(C.\ elegans)</td>
<td>308</td>
<td>0.660</td>
<td>1.029 (0.396–2.118)</td>
<td>114.3</td>
<td>4.38</td>
<td>9.40</td>
<td>18.79</td>
</tr>
<tr>
<td>(C.\ brenneri)</td>
<td>356</td>
<td>0.523</td>
<td>1.017 (0.674–1.527)</td>
<td>131.0</td>
<td>4.33</td>
<td>9.29</td>
<td>18.58</td>
</tr>
<tr>
<td>(C.\ briggsae)</td>
<td>430</td>
<td>0.518</td>
<td>1.027 (0.666–1.543)</td>
<td>114.1</td>
<td>4.38</td>
<td>9.38</td>
<td>18.76</td>
</tr>
<tr>
<td>(C.\ remanei)</td>
<td>7662</td>
<td>0.637</td>
<td>0.949 (0.592–1.613)</td>
<td>105.4</td>
<td>4.04</td>
<td>8.67</td>
<td>17.33</td>
</tr>
<tr>
<td>(C.\ japonica)</td>
<td>823</td>
<td>0.478</td>
<td>0.809 (0.571–1.207)</td>
<td>89.9</td>
<td>3.45</td>
<td>7.39</td>
<td>14.78</td>
</tr>
<tr>
<td>(C.\ sp. 5)</td>
<td>819</td>
<td>0.415</td>
<td>0.730 (0.481–1.189)</td>
<td>81.1</td>
<td>3.11</td>
<td>6.66</td>
<td>13.32</td>
</tr>
</tbody>
</table>
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 (A) and Drosophila (B). In (A), timescale in generations assumes equal substitution and mutation rates of $9 \times 10^{-9}$ per generation; timescale in years further assumes 60 days per generation. In (B), the timescale in generations assumes equal substitution and mutation rates of $5.8 \times 10^{-9}$ 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.

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 $\mu \pm 1$ 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. breneri, 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 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),
Because 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 $1.17 \times 10^6$ generations ago (table 3; figs. 2 and 3). Other Drosophila lineages have TMRCAs of $3.6–18.9 \times 10^6$ 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 TM RCA 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-Liutard et al. 2007). The lower bound on mutation rate (Haag-Liutard 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

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

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


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


John H. McDonald, Associate Editor

Accepted January 21, 2008