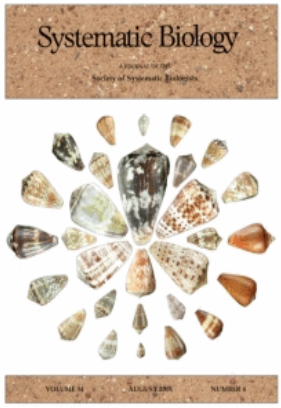

Systematic Biology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713658732

Time Flies, a New Molecular Time-Scale for Brachyceran Fly Evolution Without a Clock

Brian M. Wiegmann [a]; David K. Yeates [b]; Jeffrey L. Thorne [c]; Hirohisa Kishino [d]
[a] Department of Entomology, North Carolina State University, Raleigh, North Carolina 27695, USA; E-mail: bwiegman@unity.ncsu.edu.
[b] CSIRO, Australian National Insect Collection, Natural Resources and Biodiversity Program, GPO Box 1700, Canberra, ACT 2601, Australia.
[c] Departments of Genetics and Statistics, North Carolina State University, Raleigh, North Carolina 27695, USA.
[d] Laboratory of Biometrics, Graduate School of Agriculture and Life Sciences, University of Tokyo, Tokyo, Japan.

First Published on: 01 December 2003

To cite this Article: Wiegmann, Brian M., Yeates, David K., Thorne, Jeffrey L. and Kishino, Hirohisa (2003) 'Time Flies, a New Molecular Time-Scale for Brachyceran Fly Evolution Without a Clock', Systematic Biology, 52:6, 745 — 756

To link to this article: DOI: 10.1080/10635150390250965
URL: http://dx.doi.org/10.1080/10635150390250965

Syst. Biol. 52(6):745–756, 2003

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150390250965

Time Flies, a New Molecular Time-Scale for Brachyceran Fly Evolution Without a Clock

BRIAN M. WIEGMANN,[1] DAVID K. YEATES,[2] JEFFREY L. THORNE,[3] AND HIROHISA KISHINO[4]

[1]*Department of Entomology, North Carolina State University, Raleigh, North Carolina 27695, USA; E-mail: bwiegman@unity.ncsu.edu*
[2]*CSIRO, Australian National Insect Collection, Natural Resources and Biodiversity Program, GPO Box 1700, Canberra, ACT 2601, Australia*
[3]*Departments of Genetics and Statistics, North Carolina State University, Raleigh, North Carolina 27695, USA*
[4]*Laboratory of Biometrics, Graduate School of Agriculture and Life Sciences, University of Tokyo, Tokyo, Japan*

Abstract.— The insect order Diptera, the true flies, contains one of the four largest Mesozoic insect radiations within its suborder Brachycera. Estimates of phylogenetic relationships and divergence dates among the major brachyceran lineages have been problematic or vague because of a lack of consistent evidence and the rarity of well-preserved fossils. Here, we combine new evidence from nucleotide sequence data, morphological reinterpretations, and fossils to improve estimates of brachyceran evolutionary relationships and ages. The 28S ribosomal DNA (rDNA) gene was sequenced for a broad diversity of taxa, and the data were combined with recently published morphological scorings for a parsimony–based phylogenetic analysis. The phylogenetic topology inferred from the combined 28S rDNA and morphology data set supports brachyceran monophyly and the monophyly of the four major brachyceran infraorders and suggests relationships largely consistent with previous classifications. Weak support was found for a basal brachyceran clade comprising the infraorders Stratiomyomorpha (soldier flies and relatives), Xylophagomorpha (xylophagid flies), and Tabanomorpha (horse flies, snipe flies, and relatives). This topology and similar alternative arrangements were used to obtain Bayesian estimates of divergence times, both with and without the assumption of a constant evolutionary rate. The estimated times were relatively robust to the choice of prior distributions. Divergence times based on the 28S rDNA and several fossil constraints indicate that the Brachycera originated in the late Triassic or earliest Mesozoic and that all major lower brachyceran fly lineages had near contemporaneous origins in the mid-Jurassic prior to the origin of flowering plants (angiosperms). This study provides increased resolution of brachyceran phylogeny, and our revised estimates of fly ages should improve the temporal context of evolutionary inferences and genomic comparisons between fly model organisms. [Bayesian analysis; Brachycera; Diptera; divergence times; molecular systematics; 28S ribosomal DNA.]

In 1984, Beverley and Wilson (1984) published a seminal work estimating divergence times for *Drosophila* and several other derived flies based on the assumption of a constant rate of protein evolution in larval hemolymph proteins. This early molecule-based time scale has been widely used for dating comparisons between *Drosophila* and other dipterans (e.g., Clark and Henikoff, 1992; Bonneton et al., 1997; Pitnick et al., 1999; Shaw et al., 2001; Bolshakov et al., 2002). Since that time, the existence of a molecular clock has been widely questioned and empirically challenged (e.g., Gillespie, 1986, 1991; Ayala, 2000). A number of new methods have been proposed that incorporate heterogeneity of evolutionary rates over time into divergence time estimation (Thorne et al., 1998; Huelsenbeck et al., 2000; Kishino et al., 2001; Sanderson, 2002). The revolutionary advances of genomic and phylogenetic data and methodology and the need to date precisely the age of divergences between key fly model organisms (e.g., *Drosophila melanogaster*, *Culex pipiens*, *Anopheles gambiae*, and *Musca domestica*) make it essential to revise and extend our current understanding of the time scale for dipteran lineage origins.

The Diptera (true flies) are among the largest radiations of terrestrial eukaryotic organisms. As for other holometabolous insect orders, major diversification of fly lineages occurred in Mesozoic environments (Yeates and Wiegmann, 1999). Flies are arguably the most important insect order in terms of their impact on human and animal health, being the vectors of such devastating afflictions as malaria, yellow fever, and sleeping sickness. In addition, flies such as *Drosophila melanogaster* and *Anopheles gambiae*, are important model organisms for studies in genetics and development. Until recently, our knowledge of the phylogenetic histories of most major fly lineages was meager. With over 75,000 described species in more than 100 families, the brachyceran flies represent an enormous Mesozoic insect radiation. A major lineage within the order, the suborder Brachycera comprises the "higher Diptera," or flies with shortened antennae. This group includes many well-known members, such as fruit flies, horse flies, flower flies, blow flies, and house flies, and numerous less famous relatives. Despite intensive morphological scrutiny over the last 50 years, much of higher level Diptera classification remains contentious, unresolved, or untested by quantitative phylogenetic analyses (Yeates and Wiegmann, 1999).

Recent morphological studies have fueled debate over phylogenetic relationships of the major subgroups of Brachycera (Woodley, 1989; Wiegmann et al., 1993; Griffiths, 1994, 1996; Sinclair et al., 1994; Cumming et al., 1995; Zatwarnicki, 1996; Stuckenberg, 1999, 2001). At the core of this debate are conflicting views on the importance of specific morphological character systems, such as male genitalia and larval mouthparts, and disagreement over the interpretation of key character homologies and transformations (Griffiths, 1972, 1994, 1996; Nagatomi, 1977; Chvála, 1983; Wiegmann et al., 1993; Sinclair et al., 1994; Cumming et al., 1995). Alternative phylogenetic hypotheses based on differing morphological interpretations have also been proposed

(Nagatomi, 1977, 1992, 1996; Griffiths, 1994). Nucleotide data bring new evidence to bear on fly phylogeny, but these data have been applied to relatively few higher level questions (Friedrich and Tautz, 1997; Wiegmann et al., 2000; Krzwinski et al., 2001; Collins and Wiegmann, 2002).

Divergence time estimates derived from comparisons of gene sequence variation provide a valuable temporal context for cladogenesis in the absence of complete fossil histories. Progress in molecular systematics methodology and a wealth of molecular data are leading to more accurate age estimates (e.g., Sanderson, 1997; Wang et al., 1999; Adkins et al., 2001; Wikström et al., 2001). Divergence time estimation methods that do not assume constant rates of evolution are potentially more accurate than clock-based strategies, especially when the relationships under study span a broad range of ages and taxonomic levels or have undergone one or more significant radiations. The basic idea of these new approaches is that rates of evolution tend to be more similar when branches are nearby on the evolutionary tree than when branches are more distant.

Here, we provide a new estimate of the relationships and temporal diversification of dipteran lineages. Our revised phylogenetic and age estimates are inferred from a data set that combines molecular sequence data from 28S ribosomal DNA (rDNA) with morphological character scorings from Yeates (2002). Our results should provide new impetus and perspectives on the timing of fly diversification, the interpretation of fossil distributions, and the evolution of fly morphological innovations. In addition, they may provide for a more accurate temporal context when calibrating comparisons between fly model systems such as *Drosophila* and *Anopheles*.

Phylogeny and Fossil Record of Brachyceran Diptera

The Diptera are traditionally divided into two major suborders, Nematocera and Brachycera. The Nematocera, now widely considered paraphyletic, comprise six infraorders, the relationships among which remain uncertain (Friedrich and Tautz, 1997; Yeates and Wiegmann, 1999). The Brachycera are a monophyletic lineage whose origins are probably in the Triassic but whose major radiations are thought to have occurred in the Jurassic. As is true for most holometabolous insect groups, the fossil record for Mesozoic Diptera is too sparse to provide precise dates for most of the clades (Hennig, 1981; Evenhuis, 1994; Labandeira, 1994; Grimaldi and Cumming, 1999). The vast majority of fossil Diptera are found in relatively recent Cenozoic amber inclusions. However, the sample of older fossils from key brachyceran clades has recently surged (Grimaldi and Cumming, 1999; Nagatomi and Yang, 1998; Ren 1998a, 1998b; Mostovski and Jarzembowski, 2000). For example, Grimaldi and Cumming (1999) described more than 25 unique fossil Diptera from Cretaceous amber and proposed placements for these taxa within the Eremoneura, the major brachyceran lineage that includes all Empidoidea and Cyclorrhapha. These amber fossils, although extremely useful within radiating clades of late Cretaceous and Tertiary flies, are not old enough to be informative at the deepest branchings of the brachyceran tree, the events that spawned most of the major groups.

Ren (1998b) described newly discovered compression fossils from the late Jurassic deposits of the Yixian Formation of China. These fossils from putative flower-associated lower brachyceran groups (e.g., pangonine Tabanidae, Nemestrinidae, and Apioceridae) significantly increased the hypothesized age of flower or anthophytic fly–flower associations and thereby potentially increase the inferred age of angiosperm origins. Doubts have been raised concerning these fossils because they indirectly infer associations from rather poorly preserved morphological structures (Grimaldi, 1999). Nonetheless, these new fossil discoveries provide important new information about minimum clade age that can be used as calibration points for molecular divergence time estimation.

Phylogenetic estimates for the Brachycera break the group into four monophyletic infraorders: Xylophagomorpha containing the single family Xylophagidae; Tabanomorpha (8 families: horseflies, snipeflies, and relatives); Stratiomyomorpha (3 families: soldier flies and relatives); and Muscomorpha (100+ families: all remaining Brachycera) (Fig. 1). Muscomorpha is further divided into major clades based on traditionally recognized morphological features: Heterodactyla (all brachyceran flies with setiform tarsal empodia); Eremoneura (all Empidoidea and Cyclorrhapha, 13 morphological synapomorphies); Cyclorrhapha (characterized by obtect puparia and many maggotlike larval features); and Schizophora (all Cyclorrhapha that emerge from the puparium via a head structure called the ptilinum) (Fig. 1; Yeates and Wiegmann, 1999).

The oldest clearly interpretable brachyceran fossils are Lower Jurassic (Rhagionidae, *Palaeobolbomyia* Kovalev, 187 million years ago [MYA]; Mostovski, 2000; Mostovski and Jarzembowski, 2000), but a few controversial specimens push the estimate into the Lower/Middle Triassic (240 MYA; Krzeminski and Evenhuis, 2000). Most hypotheses suggest that the four major brachyceran lineages originated contemporaneously in the Jurassic and radiated rapidly into the diverse extant forms (Grimaldi, 1999; Grimaldi and Cumming, 1999). Recent phylogenetic analyses support this view (Fig. 1). Evidence from molecular sequences, fossils, and morphology suggests that the major fly innovations took place in the last 20 million years of the Jurassic, with subsequent radiations—perhaps driven by ecological specialization and associations with plants or herbivores—occurring in the Cretaceous (Grimaldi, 1999). Interpreting phylogenetic results and applying traditional clock-based divergence time methods can be especially difficult for rapidly radiating lineages because these groups exhibit well-documented irregularities in evolutionary rates (Friedrich and Tautz, 1997; Maddison et al., 1999). Therefore, methods that allow rates to change over time

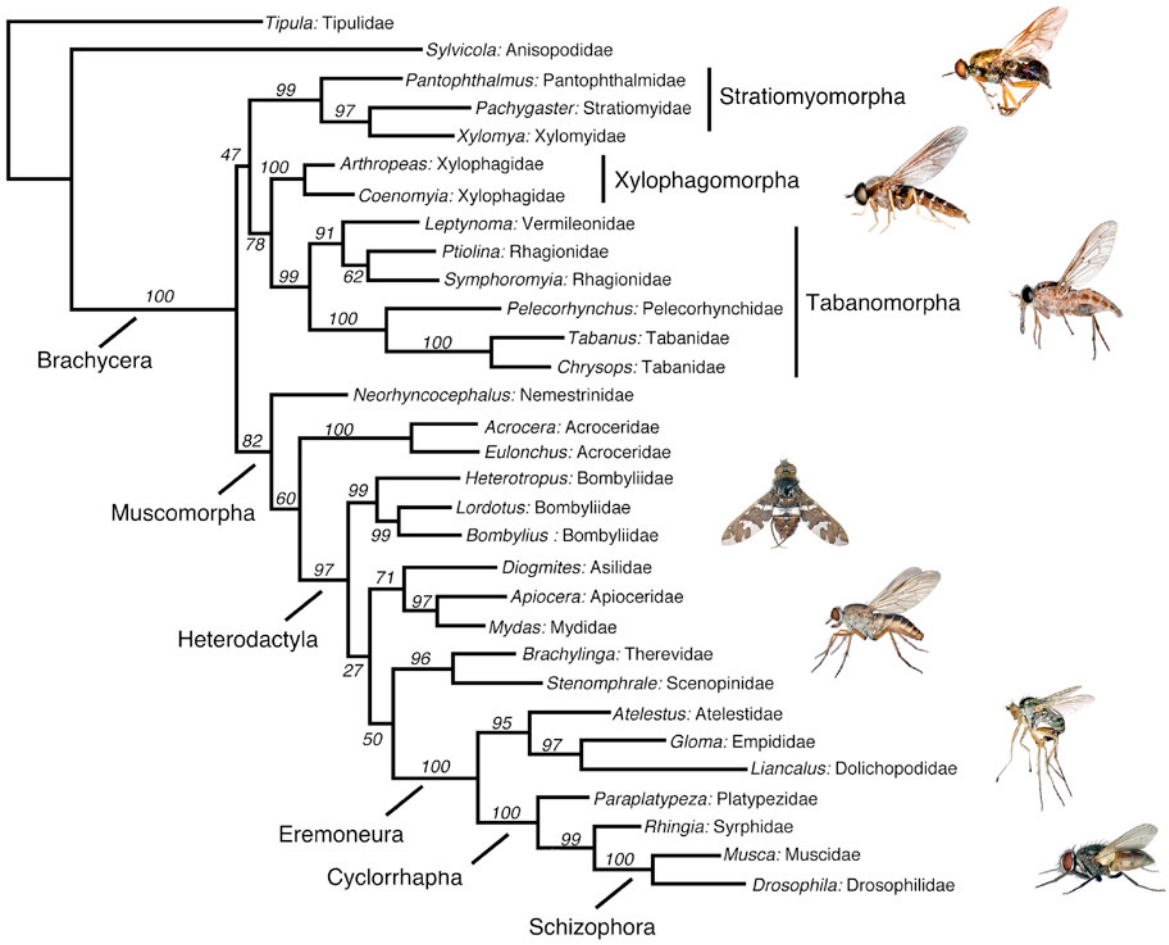

FIGURE 1. Phylogeny of the Brachycera based on combined data from 28S rDNA and morphology. This is the single most-parsimonious tree found by heuristic search with tree bisection–reconnection branch swapping and 20 random additions (1,566 steps, consistency index = 0.57; retention index = 0.56). Node values are nonparametric bootstrap percentages based on 1,000 replicates.

may vastly improve our understanding of brachyceran diversification.

MATERIALS AND METHODS

Phylogenetic Data Sets

We sequenced 2,600 base pairs of 28S rDNA via standard polymerase chain reaction amplification and dye terminator cycle sequencing, as described by Wiegmann et al. (2000). Primers employed to amplify and sequence 28S rDNA are from Yang et al. (2000) and Wiegmann et al. (2000). Sequences were obtained for 31 taxa: 2 basal dipteran taxa used as outgroups and 29 species from 23 brachyceran families. All four brachyceran infraorders are represented in these data. Taxon names, source locality, and GenBank accession numbers are listed in the Appendix. A total of 101 morphological characters were scored from systems traditionally cited as key evidence of higher level brachyceran relationships. These characters represent larval, pupal, and adult life stages, internal and external features, and male and female terminalia drawn from groundplan estimates based on consideration of both extant and fossil taxa (Yeates, 2002). Although fossil evidence is increasingly recognized as an influential source of phylogenetic information (Grimaldi and Cumming, 1999), comparative character data from basal brachyceran fossils is too sparse to score for inclusion in our data sets. The phylogenetic data sets and alignments are available on the Web page of B.M.W. (http://www4.ncsu.edu/unity/users/b/bwiegman/public_html/align.html) and are archived in the EMBL alignment database and TreeBase (SN1487; http://www.treebase.org).

Sequence Alignment, Phylogenetic Analysis, and Branch Length Estimation

Nucleotide sequences were aligned manually with the on-screen multiple alignment editor of Genetic Data Environment 2.2 (Smith et al., 1994). Highly length-variable regions of the 28S rDNA in which ad hoc placement of gaps could affect the phylogenetic outcome were excluded from analyses. Edges of these alignment-ambiguous regions were set by inspection as the last

invariant site that preserves adjacent positional homology without the insertion of a gap. Because our morphological scorings are groundplan estimates for higher level taxa, we combined data sets by appending the morphological scorings to the species sampled for nucleotides, thereby assuming that the individual species sampled unambiguously represent their noncontroversial higher level grouping. For example, the two sequenced tabanid species were given identical morphological scorings from Yeates (2002). The phylogenetic data included 2,220 characters from the 28S rDNA (608 variable and 294 parsimony informative among all taxa; 493 variable and 296 informative within Brachycera) and 101 morphological characters (Yeates, 2002). Phylogenetic analysis of the combined data set was carried out via parsimony with the program PAUP* 4.0 (version b8, Swofford, 2001). Character transformations were treated as unordered for nucleotides and morphology, and alignment gaps were treated as missing data for consistency among parsimony and Bayesian analyses of the data set. For parsimony searches, shortest trees were found by heuristic search with tree bisection–reconnection (TBR) branch swapping and 20 replicate random taxon additions. Bootstrap values were obtained with 1,000 replicate heuristic searches with TBR branch swapping and 20 replicate random taxon additions (Felsenstein, 1985).

Divergence Time Estimation

Software implementing our Bayesian method for divergence time estimation is freely available at http://statgen.ncsu.edu/thorne/multidivtime.html. This method for divergence time estimation relies on a stochastic model for changes of evolutionary rate over time (Kishino et al., 2001; see also Thorne and Kishino, 2002). To disentangle the evolutionary rates and times that are confounded when branch lengths are inferred, the method obtains divergence time estimates by combining sequence data with information such as constraints on node times that are due to fossil data. The method requires an assumed topology. Here, the topology of Figure 1 was assumed because we deemed it the best current working hypothesis based on quantitative analysis of all available data (Wiegmann et al., 2000; Yeates, 2002).

To estimate branch lengths on the inferred topology, a discretized gamma distribution with five rate heterogeneity categories (Yang, 1994) was used in conjunction with the Felsenstein 1984 model of nucleotide change (see Felsenstein, 1989). This treatment was selected because it is a reasonable compromise between biological reality and the computational tractability concerns that can arise in Bayesian analyses. Version 3.0c of the PAML software (Yang, 1997) generated maximum likelihood estimates of the amount of rate heterogeneity among sites as well as maximum likelihood estimates of the transition/transversion ratios and nucleotide frequencies. Given these estimated parameters, our own software was employed to approximate the likelihood surface with a multivariate normal distribution centered at the maximum likelihood estimates of branch lengths (see Thorne et al., 1998). This multivariate normal approximation requires much less computation than does calculation of likelihoods via the pruning algorithm of Felsenstein (1981). Computational tractability is an important concern because the likelihood needs to be evaluated or approximated for each of the millions of sets of rates and times that are considered in our Markov chain Monte Carlo (MCMC) procedure (see below).

Our Bayesian method requires the specification of prior distributions for parameters. Specifically, prior distributions are required for the rate at the ingroup root, the node times, and a parameter that determines on a log-scale the expected amount of rate change per unit time. For the node time prior, there can be as few as one, but the more nodes specified the better. In each case, we selected prior distributions that were biologically plausible but otherwise vague.

Unless otherwise specified, the prior distribution for the rate at the ingroup root node for all analyses presented here was a gamma distribution with a mean of 0.02 and an SD of 0.015 changes per time unit, where 1 time unit in this analysis represents 100 million years. The mean of this prior distribution was selected by examining the sum of the estimated branch lengths separating an ingroup tip and the ingroup root. Although this sum varies widely among ingroup tips in this data set, the median value is very roughly 0.04 expected substitutions per site. Because branch lengths are rates multiplied by time durations and because the time since the ingroup root was believed a priori to be about 200 million years (i.e., 2.0 time units), the prior mean of the ingroup root rate was set at 0.02. This inspection of branch lengths to set priors is technically a violation of the definition of a prior as representing beliefs before data analysis. For this reason, we intentionally set the SD of the prior for the ingroup root rate (0.015) as large relative to the mean.

The assumption our Bayesian method makes about rate change is that the logarithm of the rate at the end of a branch has a normal distribution such that the expected rate at the end of the branch is equal to the rate at the beginning of the branch. The variance of this normal distribution is the product of the time duration of the branch and the rate change parameter. The average rate on a branch is assumed to be the mean of the rates at the nodes that begin and end the branch. The Bayesian method can analyze data with the constant rate assumption of a molecular clock by setting the rate change parameter equal to zero. Unless otherwise specified, for all analyses presented here where rates were allowed to change over time the prior distribution that we selected for this rate change parameter was a gamma distribution with mean and SD both equal to 0.5. Choice of this prior for the rate variation parameter was influenced by previous analysis of simulated and real data sets. In practice, we have found that a value of 1 or 2 for the product of the prior mean for the rate variation parameter and the prior mean for the number of time units since the ingroup root yields satisfactory divergence time estimates in a wide

range of cases (J.L.T., pers. obs.). Our decision to set the prior SD equal to the prior rate was also motivated by previous experience.

Divergence time estimation from molecular sequence data is improved by calibration from external dating information provided by fossils or other sources. In our Bayesian implementation, information external to the molecular sequence data is represented by constraining node times on the phylogenetic tree. Individual node times can be constrained so as to be either earlier or later than some specific date. Five constraints on node times were employed in this analysis. We used four fossil-based dates for minimum clade age estimates and added one maximum age constraint (Fig. 3). The maximum and minimum bounds on the date of origin of the Brachycera were set at 250–187 MYA, extending well into the Triassic. This maximum is 10 million years earlier than any postulated fossil evidence or hypothesis for Brachyceran origins but is within the hypothesized origin of the Diptera (Triassic; Hennig, 1981). The minimum age, 187 MYA, is based on a number of unequivocally brachyceran fossils from the Lower Jurassic of China, Siberia, and England (reviewed by Evenhuis, 1994; Grimaldi and Cumming, 1999). The remaining minimum clade ages used were 170 MYA for Bombyliidae (*Palaeoplatypygus*, Siberia; Kalugina and Kovalev, 1985), 120 MYA for Cyclorrhapha, (*Chimeromyia*, Lebanon; Grimaldi and Cumming, 1999), and 70 MYA for Schizophora (*Cretaformia*, Canada; McAlpine, 1970).

For all sets of node times that satisfy the constraints, the software requires specification of how likely the node times are a priori. In the absence of constraints, this prior distribution would have two parts (see Kishino et al., 2001). The first part would be a gamma distribution for the time since the ingroup root. We adopted a gamma distribution with a mean of 200 million years and SD of 15 million years. The second part would be a Dirichlet distribution modified to tree structures that defines the proportion of times between tips and internal nodes relative to the time between tips and the ingroup root. Because the prior distribution for node times is conditional upon the constraints and therefore must satisfy them, the prior distribution becomes more complicated and needs to be approximated (see Kishino et al., 2001).

The MCMC method was employed as described by Kishino et al. (2001) to approximate both prior and posterior distributions. Here, initial parameter values were randomly selected to initialize the Markov chain, and then a burn-in period of 100,000 cycles of proposed changes to the current state of the Markov chain was completed before parameters were sampled from the chain. Thereafter, the state of the Markov chain was sampled every 100 cycles until a total of 10,000 samples had been collected. Prior and posterior distributions were approximated based upon the 10,000 samples.

MCMC approaches would be guaranteed to perfectly approximate the distributions of interest if the Markov chains could be run for an infinite amount of time. There are various diagnostics available to determine whether the chains have been run long enough to give a good approximation. In our opinion, the most simple approach and one of the best is simply to repeat the MCMC procedure multiple times from different starting points and then determine whether the approximations obtained by different repetitions are sufficiently similar to one another. The variability of approximations among repetitions is known as Monte Carlo error, and techniques exist for quantifying this variability. However, visual inspection of our results indicates that this Monte Carlo error is small. Different MCMC runs tend to give us about the same answer for the first three significant figures of most parameters.

Therefore, convergence of the Markov chain was assessed for each analysis by subsequently running another Markov chain from a different randomly selected initial state and then verifying that the posterior distribution approximations based upon the first and second MCMC analyses were highly similar. Although the analyses assuming a molecular clock converge more slowly than do the analyses used to approximate the posterior in the presence of rate variation over time, all of the Markov chain analyses appear to have yielded satisfactory approximations of the distributions of interest.

One key issue with this and any divergence time analysis is robustness. To explore robustness, we investigated all 12 combinations of three diverse prior distributions for the rate at the ingroup root and four diverse prior distributions for the rate variation parameter. The three ingroup root rate priors explored were a gamma distribution with a prior mean of 0.002 and prior SD of 0.0015, a gamma distribution with a prior mean of 0.02 and SD of 0.015, and a gamma distribution with a prior mean of 0.2 and SD of 0.15. For the rate variation priors, we explored the case of a perfect clock (i.e., no rate variation allowed), a gamma distribution with a mean of 0.05 and SD of 0.05, a gamma distribution with a mean of 0.5 and SD of 0.5, and a gamma distribution with a mean of 2.0 and SD of 2.0.

Because a limitation of our divergence time estimation software is that it assumes a known topology, we explored the effect of other plausible topologies on our estimates of key divergence dates. In addition to the topology from the combined data, we also obtained topologies from parsimony and from Bayesian analyses of the 28S rDNA data set alone (Fig. 2). The Bayesian analyses of the nucleotide data were carried out with the program MrBayes 3.0B4 (Huelsenbeck and Ronquist, 2001) under the HKY + gamma model of nucleotide substitution with the shape prior for rate variation being a uniform distribution between 0.1 and 50, nucleotide frequencies being empirically estimated, and the transition/transversion rate ratio prior set to beta (1.0, 1.0). For each of several runs from different randomly selected starting points, the Bayesian analysis was run for 1 million generations and sampled every 1,000 generations.

For each MCMC analysis, four chains were simultaneously run, one cold and three incrementally heated. Plots of log-likelihood scores versus generation time indicated that a burn-in of 200,000 generations was satisfactory. Sample points prior to this burn-in level were discarded.

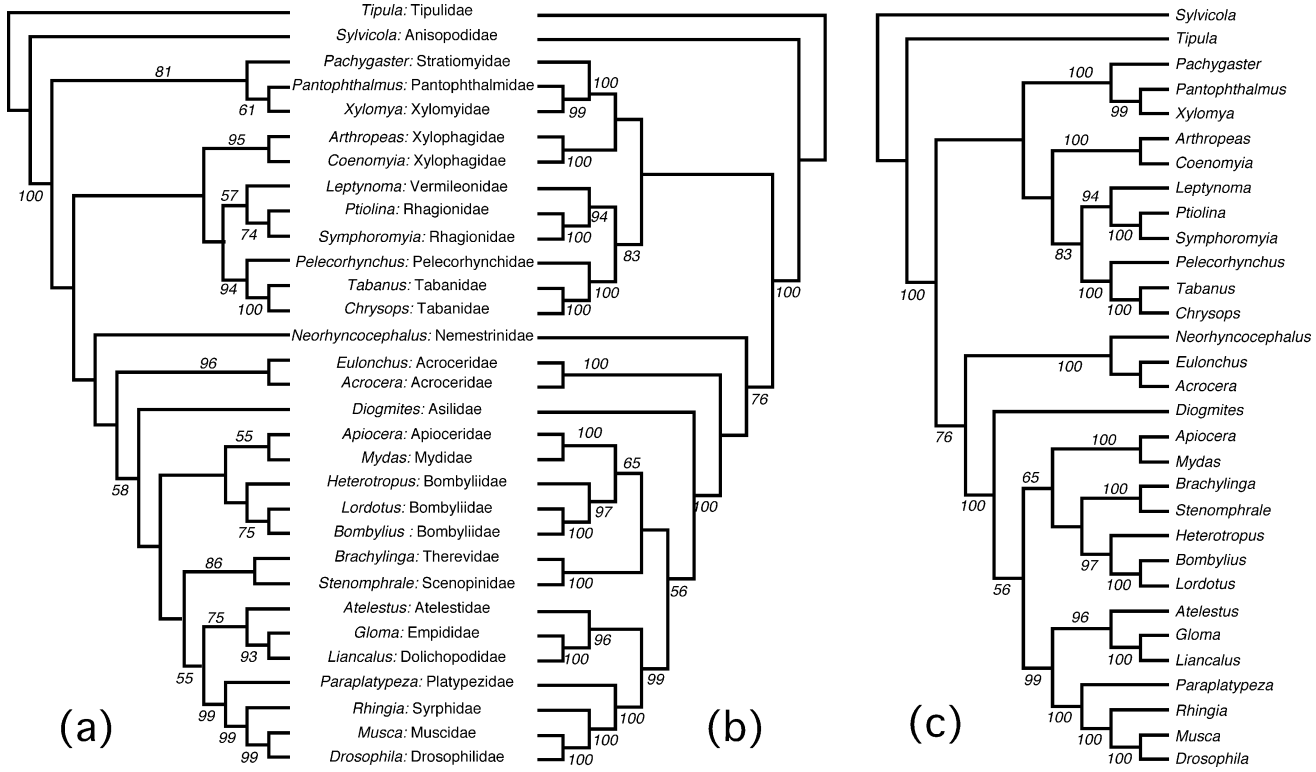

FIGURE 2. Alternative phylogenetic arrangements of the Brachycera based on analysis of the aligned 28S rDNA data set alone. (a) The single most-parsimonious tree found by heuristic search with TBR branch swapping and 20 random additions (1,566 steps, consistency index = 0.57; retention index = 0.56). Node values are nonparametric bootstrap percentages based upon 1,000 replicates. (b, c) Two topologies of equally highest mean posterior probability (p = 0.017) found by MCMC tree search in MrBayes 3.01 for the aligned 28S rDNA data under the HKY + gamma model of nucleotide substitution with estimated transfomation parameters and equal base frequencies. Node values are posterior probabilities based on a 50% majority rule consensus trees sampled after burn-in.

Comparison of the approximate posterior distributions for the different MCMC analyses indicated that convergence of the cold Markov chain to its stationary distribution had been achieved.

RESULTS AND DISCUSSION

Basal Relationships of the Brachycera

Parsimony analysis of the combined molecular and morphological data set yielded a single most-parsimonious, minimum-length tree (Fig. 1). Relationships inferred from this tree are consistent with those from recent morphological (Yeates, 2002) and molecular (Wiegmann et al., 2000) analyses and agree in large part with expected relationships from traditional morphology-based classifications (Griffiths, 1994; Sinclair et al., 1994; Yeates and Wiegmann, 1999). There is >80% bootstrap support for the monophyly of the Brachycera and for the monophyly of the four brachyceran infraorders: Stratiomyomorpha, Xylophagomorpha, Tabanomorpha, and Muscomorpha. Support is also strong for most of the higher level groupings, including the Xylophagomorpha + Tabanomorpha, and four morphologically well-supported clades in brachyceran classification: Heterodactyla, Eremoneura, Cyclorrhapha, and Schizophora (Woodley, 1989; Yeates and Wiegmann, 1999; Yeates, 2002) (Fig. 1). The Asiloidea (Bombyliidae, Asilidae, Apioceridae, Mydidae, Therevidae, and Scenopinidae) is not monophyletic, and support for relationships among these families is low (Yeates, 2002). Analysis of the combined molecular and morphological data weakly favors a monophyletic basal lineage of (Stratiomyomorpha(Xylophagomorpha + Tabanomorpha)), hereafter called the SXT clade. This arrangement was also found in a recently published morphology-based analysis and is supported by a fusion of the thoracic ganglia (Yeates et al., 2002). The limited support for this grouping in both molecular and morphological data cannot rule out potential alternatives such as Stratiomyomorpha or Xylphagomorpha + Tabanomorpha as the basal brachyceran lineage. For example, parsimony analysis of the aligned 28S rDNA data yielded the single most-parsimonious topology of Figure 2a. This tree shows weak support for a basal position for the Stratiomyomorpha and also differs in the arrangement of asiloid families (Fig. 2a). The Bayesian analysis of the 28S data alone indicates that these data do not contain strong support for any single fully resolved topology. The two Bayesian trees that we used to perform divergence time estimates (Figs. 2b, 2c) each had posterior probabilities of

approximately 0.017, and no topologies were recovered that had higher estimated posterior probabilities. Both topologies contain the SXT clade, but overall support for this grouping is low.

Posterior Distributions of Node Times

Figure 3a shows the prior distribution of divergence times as approximated by our MCMC method. The prior distribution for node times was intentionally set to be diffuse. The fact that the posterior distributions of divergence times for the constant rate analysis (Fig. 3b) and the variable rate analysis (Fig. 3c) are less variable than the prior distribution of divergence times can be attributed to the information contained within the brachyceran 28S data set. As shown in Figure 3, the width of credibility intervals for node times is influenced by the proximity in the tree to constraints. Nodes that are themselves constrained or that are near constrained nodes tend to have more narrow credibility intervals than do other nodes. This relationship is expected because the ability to constrain node times in one part of the phylogeny indicates that more information regarding these node times exists in these areas than in parts of the tree where constraints are absent.

The time estimates ranged between 48 and 216 MYA for the nonclock Bayesian divergence time estimation and between 81 and 214 MYA for the Bayesian divergence time estimation with a clock (Figs. 3b, 3c; Table 1). Although the null hypothesis of a constant rate of evolution for these 28S rDNA data is rejected ($P < 0.001$) by conventional likelihood-based tests (see Felsenstein, 1981; Muse and Weir, 1992), the date estimates for most nodes do not substantially differ under clock and nonclock assumptions. Only for a couple of nodes (e.g., *Drosophila/Musca*, 48 MYA nonclock vs. 86 MYA clock) does the assumption of a constant rate throughout the tree have a recognizable effect on divergence time estimation. We expect that our application of multiple constraints across the tree may have dampened any recognizable effect an assumption of rate homogeneity would have on time estimates in unconstrained branches.

As a first step in estimating divergence times from molecular sequence data and fossil information, it may be tempting to apply conventional sequence-based tests of the null hypothesis that rates are constant over time. If the null hypothesis is not rejected, the next step would be to estimate divergence times via methods that assume a constant rate (e.g., Kumar and Hedges, 1998). We do not recommend such a procedure because failure to reject the null hypothesis does not mean the null hypothesis is true and because the power and reliability of commonly applied methods, such as the relative rates test, have been questioned (e.g., Bromham et al., 2000). Moreover, conventional sequence-based tests of the constant rate hypothesis are not designed to indicate when the constant rate assumption conflicts with fossil information.

Our divergence time estimates agree in large part with the fossil record of flies but generally push age estimates older than previous inferences (Table 1). There has been some debate over the inferred age of the Brachycera. Krzeminski (1992) considered the extinct late Triassic family Alinkidae (208 MYA) to be the oldest known true representative of the suborder, but Grimaldi and Cumming (1999) were more skeptical, favoring instead the tabanomorphan fossil *Paleobolbomyia* (187 MYA) as a more clearly interpretable early brachyceran. The 28S data support an older date for Brachycera, well within the Triassic (Table 1), perhaps lending plausibility to pre-Jurassic forms such as Alinkidae. Additionally, the most-parsimonious topology and date estimates suggest that the initial diversification of Brachycera was a late Triassic split into two separate monophyletic lineages, the SXT clade and the Muscomorpha, with the latter group eventually containing the bulk of all further fly diversification. As mentioned above, support for the SXT grouping is weak in both morphology and molecules, and so estimated dates of origin for the three infraorders (SXT) based on the current data are

TABLE 1. Divergence time estimates[a] for brachyceran fly lineages based on 28S rDNA sequences and the phylogenetic topologies of Figures 1 and 2a–c.

Taxon	Figure 1		Figure 2a	Estimates without a clock Figure 2b	Figure 2c	Earliest known fossil (MYA)	Hypothesized group age prior to study[b] (MYA)
	Estimates without a clock	Estimates with a clock					
Brachycera	216 (194, 241)	214 (192, 238)	222 (199, 245)	218 (195, 242)	216 (194, 240)	187 (208)	200
Muscomorpha	216 (194, 241)	214 (192, 238)	216 (193. 239)	218 (195, 242)	216 (194, 240)	144	198
Stratiomyomorpha	204 (176, 232)	199 (172, 228)	222 (199, 245)	192 (159, 223)	202 (173, 230)	187	198
Xylophagomorpha	192 (160, 224)	180 (149, 212)	204 (175, 232)	192 (159, 223)	193 (163, 223)	187	200
Tabanomorpha	192 (160, 224)	180 (149, 212)	204 (175, 232)	203 (174, 232)	193 (163, 223)	187	200
Heterodactyla	197 (177, 223)	195 (176, 221)	198 (179, 222)	202 (181, 226)	205 (184, 231)	144	185
Eremoneura	166 (143, 192)	165 (144, 189)	166 (147, 188)	179 (170, 199)	181 (170, 203)	150[c]	165
Empidoidea	163 (143, 189)	154 (128, 181)	144 (123, 170)	155 (129, 183)	156 (129, 185)	130[d]	150
Cyclorrhapha	142 (122, 169)	154 (128, 181)	144 (123, 170)	155 (129, 183)	156 (129, 185)	130[d]	150
Schizophora	84 (70, 113)	107 (75, 142)	85 (71, 119)	87 (71, 119)	87 (71, 1220)	80[d]	88
Drosophila/Musca	48 (29, 76)	81 (39, 121)	48 (29, 75)	50 (30, 80)	51 (30, 80)	70[d]	75–100

[a]Posterior means for Bayesian divergence time estimates are followed in parentheses by 95% credibility intervals.
[b]Evenhuis, 1994.
[c]Nagatomi and Yang, 1998.
[d]Grimaldi, 1999; Grimaldi and Cumming, 1999.

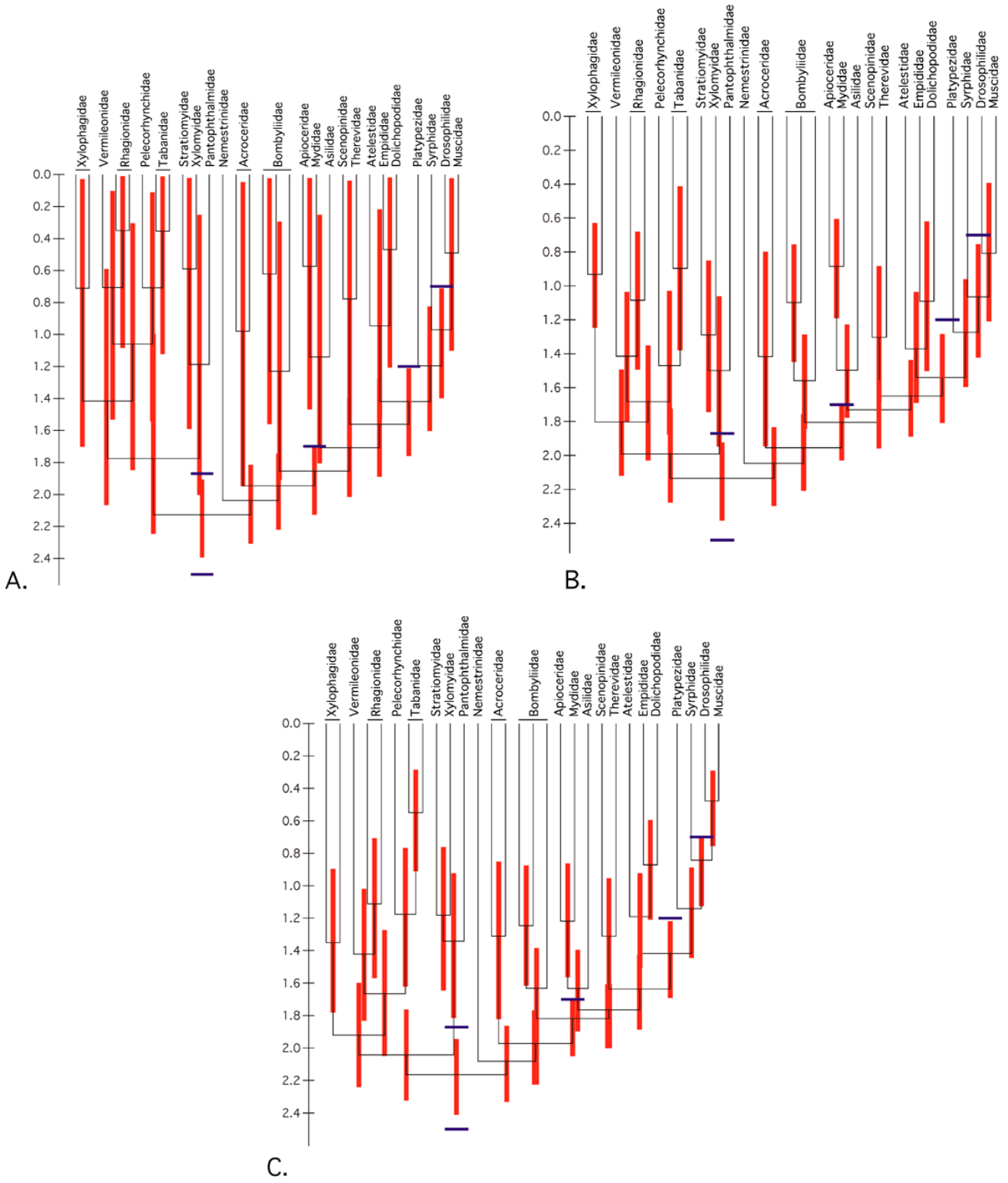

FIGURE 3. Estimated divergence times for lower brachyceran clades. Time units are hundreds of millions of years before present. The red vertical line through each node shows the time interval that contains 95% of the probability for the node age. Horizontal node positions represent the means of the probability distributions for node ages. Thick blue horizontal lines represent constraints on node times. These lines are centered around the node being constrained, and their vertical placement represents the age value of the constraint. A blue line above a node is its minimum age constraint, and a blue line below a node represents a maximum age constraint. (a) The approximate prior distribution of node times. (b) The approximate posterior distribution of node times when 28S rDNA data are analyzed by assuming a constant rate of evolution. (c) The approximate posterior distribution of node times when 28S rDNA data are analyzed by allowing evolutionary rates to change over time.

essentially contemporaneous even under different scenarios for the exact basal resolution of the tree (Table 1).

Within the Muscomorpha, nearly all of the major lineages are estimated to have appeared between 180 and 140 MYA. Given the large taxonomic, morphological, ecological, and behavioral diversity of these groups, it seems clear that the Jurassic was the major period of innovation for Diptera. Our dates are generally congruent with those inferred from the phylogenetic hypotheses of Grimaldi and Cumming (1999) and are consistent with fossil evidence (Evenhuis, 1994; Nagatomi and Yang, 1998; Krzminski and Evenhuis, 2000). Beverley and Wilson (1984) provided a Cretaceous age for the Schizophora (65–135 MYA) based on larval haemolymph protein evolution. Our new method generally agrees with dates obtained by Beverley and Wilson (1984) but results in narrower bounds on the origin of the Schizophora within the Upper Cretaceous (74–98 MYA). Relationships among schizophoran families are still too poorly resolved to fully evaluate and sharpen the date for the *Drosophila/Musca* split (Yeates and Wiegmann, 1999).

Robustness of Divergence Time Estimates

As indicated by the entries of Table 1, the divergence time estimates of key clades were not substantially affected by minor changes in the topology derived from various analyses of the data sets. This robustness was expected because the regions of topological uncertainty in the phylogeny tend also to be regions where branch length estimates are short and where time durations separating successive nodes are probably short. As long as even a little rate variation over time is permitted, the posterior distribution of node times is not very sensitive to the specific combination of the prior for the ingroup root rate and rate variation parameter (Table 2). However, when a constant rate of evolution is enforced, divergence time estimates can be more substantially affected (Bromham and Hendy, 2000; Bromham et al., 2000). The insensitivity to the prior distribution for the other cases is reassuring because it is sometimes difficult to quantitatively summarize prior expectations about ingroup root rate and the rate variation parameter.

TABLE 2. Effect of prior distributions for rate variation parameter and ingroup root rate on posterior distributions for node times and rate variation parameters. The tree of Figure 1 was analyzed with a variety of prior distributions for the rate variation parameter and the ingroup root rate. Each group of four entries corresponds to a specific combination of prior for the rate variation parameter and the ingroup root rate. For each group of four lines, the estimated posterior mean is followed in parentheses by the estimated 95% credibility interval. The four values presented for each treatment are respectively the estimated divergence times for *Musca/Drosophila*, Tabanidae/Pelecorhynchidae, Nemestrinidae/all other Muscomorpha, and the rate variation parameter.

Rate variation parameter[a]	Ingroup root rate[b]		
	0.002 (0.0015)	0.02 (0.015)	0.2 (0.15)
Clock	83 (40, 125)	81 (39, 121)	81 (41, 120)
	150 (104, 192)	147 (103, 188)	147 (102, 187)
	208 (185, 234)	205 (183, 230)	205 (184, 230)
	—	—	—
0.05 (0.05)	52 (30, 82)	51 (31, 80)	51 (31, 79)
	120 (79, 168)	120 (80, 165)	119 (80, 165)
	213 (189, 239)	209 (186, 234)	209 (186, 234)
	0.22 (0.11, 0.39)	0.23 (0.11, 0.38)	0.23 (0.12, 0.39)
0.5 (0.5)	48 (29, 76)	48 (29, 76)	47 (29, 74)
	115 (73, 162)	118 (77, 162)	118 (78, 164)
	211 (188, 237)	208 (186, 233)	207 (185, 233)
	0.42 (0.18, 0.85)	0.43 (0.19, 0.86)	0.44 (0.19, 0.89)
2.0 (2.0)	48 (29, 76)	47 (29, 74)	47 (29, 73)
	114 (73, 161)	118 (77, 163)	117 (77, 163)
	211 (187, 236)	208 (186, 233)	207 (186, 232)
	0.48 (0.20, 1.00)	0.49 (0.20, 1.04)	0.50 (0.21, 1.06)

[a]The row labels show the prior distribution for the rate variation parameter. The first row (Clock) is the case where rates do not vary over time. The remaining three rows show the mean (SD) of the gamma distributed prior for the rate variation parameter.

[b]Column heads are the gamma distributed prior for the ingroup root rate with the mean followed by the SD in parentheses.

Evolution of Flies and Angiosperms

Our analysis suggests that the major brachyceran lineages were likely already established by the time of the first appearance of flowering plants (angiosperms) but that the origin and diversification of the Cyclorrhapha in the Tertiary is likely contemporary with many of the major diversifications within angiosperm lineages (Magallón and Sanderson, 2001; Wikström et al., 2001). Fossil-based estimates suggest that angiosperms first appeared in the Lower Cretaceous (132 MYA; Hughes, 1994; Crane et al., 1995), but recent molecular phylogenetic estimates of divergence times indicate that the group could be much older (158–179 MYA; Wikström et al., 2001). Our results suggest that the earliest brachyeran Diptera (SXT clade, Fig. 1; ca. 7,000 species) appeared in the late Triassic (220 MYA) but with Muscomorpha (ca. 75,000 species) having its origin at about the same time as the oldest estimates for angiosperms. The greatest diversity in Muscomorpha begins with the Asiloidea (paraphyletic on our tree, 12,000 species) in the Jurassic (200–170 MYA) and peaks within the Schizophora in the Tertiary between 65 and 20 MYA, a period of major expansion for fly hosts and habitats, including angiosperms, grass biomes, and ungulates (Blagoderov et al., 2002). These recent hypotheses for the origin of the angiosperms (e.g., Wikström et al., 2001) also correspond with the origin and diversification of the Eremoneura, particularly the basal Cyclorrhapha. The hypothesis that coevolution of angiosperms and their insect pollinators is the engine of diversity for both groups has been strongly challenged in recent times (Labandeira and Sepkoski, 1993; Gorelick, 2001).

The earliest evidence of insect adaptation to flower feeding is in the Upper Jurassic (144–156 MYA; see Grimaldi, 1999). Our analysis puts the origin of the Nemestrinidae prior to this, in the Upper Triassic, 210 MYA. Grimaldi (1999) postulated that a long proboscis and pollen feeding behavior originated in the Upper Cretaceous and that nemestrinids were the first insects to be adapted to flowers. These early nemestrinids may have fed on flowerlike structures of Jurassic Bennettitales and gnetaleans (Grimaldi, 1999),

TABLE 3. Divergence time estimates for comparison of *Drosophila* and major dipteran groups.

Lineage	Example comparison with *Drosophila*	Divergence time (MYA)
Acalyptrate superfamilies	*Rhagoletis*	48–86[a]
Schizophora	*Musca*	29–80[b]
Eremoneura	*Empis*	143–203[b]
Heterodactyla	*Bombylius*	176–216[b]
Brachycera	*Tabanus*	176–216[b]
Diptera[c]	*Anopheles*	223–240[d]

[a]Divergence time range based on mean age estimates for schizophoran stem and crown groups (Beverley and Wilson, 1984; Grimaldi and Cumming, 1999).

[b]Dates based on Bayesian divergence time estimates reported in the current study and the topology of Figure 1.

[c]Phylogenetic position of *Anopheles* (Culicomorpha) is uncertain among basal nonbrachyceran lineages (Yeates and Wiegmann, 1999); estimate based on hypothesized age of common ancestry for Brachycera and basal diptera ("Nematocera").

[d]Krzeminski and Evenhuis, 2000; Yeates and Wiegmann, 1999.

preadapting them for their more recent associations with long-corolla flowers. Detailed phyologenetic analyses coupled with comparative investigation of key morphological features within flower-associated lineages of flies, such as Nemestrinidae, pangonine Tabanidae, and many Bombyliidae and Empididae, will ultimately be required to fully test assertions that ecological associations between insects and plants had a direct causative effect on the stunning macroevolutionary success of both.

Implications of Revised Dates for Genetic Comparisons Between Drosophila and Other Diptera

We have extracted from our results a set of divergence time estimates between fly lineages (Table 3). These estimates may serve as a guide to the antiquity of genetic comparisons between dipteran model organisms. This taxonomically extensive, revised time scale is an attempt to free dipteran age estimates from a tradition of conjecture and from the biologically dubious assumption of a molecular clock. This time scale represents our best efforts to provide dates for comparisons that are necessitated by the breadth of research done on Diptera. As dipteran genomes are completed, increased phylogenetic resolution and revised divergence time estimates will be critical for establishing a precise chronology upon which to base our understanding of the development, neurobiology, population genetics, and comparative genomics of *Drosophila* and other flies.

ACKNOWLEDGMENTS

We thank L. L. Deitz, D. Grimaldi, C. Mitter, M. B. Mostovski, and two anonymous reviewers for valuable critical readings of earlier versions of the manuscript. We also thank P. J. Chandler, M. E. Irwin, E. I. Schlinger, D. W. Webb, and M. Zumbado for providing important fly specimens for nucleotide sequencing and M. Trautwein and A. Hastings for artistic contributions and help with the illustrations. This research was supported by grants from the National Science Foundation (DEB-9521925, DEB-0089745, INT-990934), the Japanese Society for the Promotion of Science, and the Japanese Science and Technology Corporation.

REFERENCES

ADKINS, R. M., E. L. GELKE, D. ROWE, AND R. L. HONEYCUTT. 2001. Molecular phylogeny and divergence time estimates for major rodent groups: Evidence from multiple genes. Mol. Biol. Evol. 18:777–791.

AYALA, F. K. 2000. Neutralism and selectionism: The molecular clock. Gene 261:27–33.

BEVERLEY, S. M., AND, A. C. WILSON. 1984. Molecular evolution in *Drosophila* and the higher Diptera. II. A time scale for fly evolution. J. Mol. Evol. 21:1–13.

BLAGODEROV, V. A., E. D. LUKASHEVICH, AND M. B. MOSTOVSKI. 2002. Order Diptera Linne, 1758. The true flies (=Muscida Laicharting, 1781). Pages 227–240 *in* History of insects (A. P. Rasnitsyn and D. Quick, eds.). Kluwer, Dordrecht, The Netherlands.

BOLSHAKOV, V. N., P. TOPALIS, C. BLASS, E. KOKOZA, A. DELLA TORRE, F. C. KAFATOS, AND C. LOUIS. 2002. A comparative genomic analysis of two distant Diptera, the fruit fly, *Drosophila melanogaster*, and the malaria mosquito, *Anopheles gambiae*. Gen. Res. 12:57–66.

BONNETON, F., P. J. SHAW, C. FAZAKERLEY, M. SHI, AND G. A. DOVER. 1997. Comparison of bicoid dependent regulation of hunchback between *Musca domestica* and *Drosophila melanogaster*. Mech. Dev. 66:143–156.

BROMHAM, L. D., AND M. D. HENDY. 2000. Can fast early rates reconcile molecular dates with the cambrian explosion? Proc. R. Soc. Lond. B 267:1041–1047.

BROMHAM, L. D., D. PENNY, A. RAMBAUT, AND M. D. HENDY. 2000. The power of relative rates tests depends on the data. J. Mol. Evol. 50:296–301.

CHVÁLA, M. 1983. The Empidoidea (Diptera) of Fennoscandia and Denmark. II. General Part. The families Hybotidae, Atelestidae and Microphoridae. Fauna Entomol. Scand. 12:1–297.

CLARK, D. V., AND S. HENIKOFF. 1992. Unusual organizational features of the *Drosphila-Gart* locus are not conserved within Diptera. J. Mol. Evol. 35:51–59.

COLLINS, K. P., AND B. M. WIEGMANN. 2002. Phylogenetic relationships of the Eremoneura (Diptera: Brachycera) based on 28S rDNA and EF-1α sequences. Insect Syst. Evol. 33:421–444.

CRANE, P. R., E. M. FRIIS, AND K. R. PEDERSEN. 1995.The origin and early diversification of angiosperms. Nature 374:27–33.

CUMMING, J. M., B. J. SINCLAIR, AND D. M. WOOD. 1995. Homology and phylogenetic implications of male genitalia in Diptera—Eremoneura. Entomol. Scand. 26:120–151.

EVENHUIS, N. L. 1994. Catalogue of the fossil flies of the world (Insecta: Diptera). Backhuys, Leiden, The Netherlands.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17:368–376.

FELSENSTEIN, J. 1985. Confidence limits on phylogeny: An approach using the bootstrap. Evolution 39:783–791.

FELSENSTEIN, J. 1989. Phylip—Phylogeny inference package (version 3.2). Cladistics 5:164–166.

FRIEDRICH, M., AND D. TAUTZ. 1997. Evolution and phylogeny of the Diptera: A molecular phylogenetic analysis using 28S rDNA. Mol. Biol. Evol. 14:644–653.

GILLESPIE, J. H. 1986. Variability of evolutionary rates of DNA. Genetics 113:1077–1091.

GILLESPIE, J. H. 1991 The causes of molecular evolution. Oxford Univ. Press, Oxford, U.K.

GORELICK, R. 2001. Did insect pollination cause increased seed plant diversity? Biol. J. Linn. Soc. 74:407–427.

GRIFFITHS, G. C. D. 1972. The phylogenetic classification of Diptera Cyclorrhapha, with special reference to the structure of the male postabdomen. Ser. Entomol. 8:1–340.

GRIFFITHS, G. C. D. 1994. Relationships among the major subgroups of Brachycera (Diptera): A critical review. Can. Entomol. 126:861–880.

GRIFFITHS, G. C. D. 1996. Review of papers on the male genitalia of Diptera by D. M. Wood and associates. Stud. Dipterol. 3:107–123.

GRIMALDI, D. 1999. The co-radiations of pollinating insects and angiosperms in the Cretaceous. Ann. Mo. Bot. Gard. 86:373–406.

GRIMALDI, D., AND J. CUMMING. 1999. Brachyceran Diptera in Cretaceous ambers and Mesozoic diversification of the Eremoneura. Bull. Am. Mus. Nat. Hist. 239:1–124.

HENNIG, W. 1981. Insect phylogeny. John Wiley and Sons, New York.
HUELSENBECK, J. P., B. LARGET, AND D. SWOFFORD. 2000. A compound Poisson process for relaxing the molecular clock. Genetics 154:1879–1892.
HUELSENBECK, J. P., AND F. RONQUIST. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754–755.
KALUGINA, N. S., AND V. G. KOVALEV. 1985. Dipterous insects of Jurassic Siberia. Palaeontological Institute, Akademia Nauk, Moscow (in Russian).
KISHINO, H., J. L. THORNE, AND W. J. BRUNO. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol. Biol. Evol. 18:352–361.
KRZEMINSKI, W. 1992. Triassic and Lower Jurassic stage of Diptera evolution. Mitt. Schweiz. Entomol. Ges. 65:39–59.
KRZEMINSKI, W., AND N. L. EVENHUIS. 2000. Review of Diptera paleontological records. Pages 535–564 *in* Contributions to a manual of Palaearctic Diptera, Volume 1 (L. Papp and B. Darvas, eds.). Science Herald, Budapest.
KRZWINSKI, J., R. C. WILKERSON, AND N. J. BESANSKY. 2001. Toward understanding anopheline (Diptera: Culicidae) phylogeny: Insights from nuclear single-copy genes and the weight of evidence. Syst. Biol. 50:540–556.
KUMAR, S., AND S. B. HEDGES. 1998. A molecular timescale for vertebrate evolution. Science 392:917–920.
LABANDEIRA, C. C. 1994. A compendium of fossil insect families. Milw. Public Mus. Contrib. Biol. Geol. 88:1–71.
LABANDEIRA, C. C., and J. J. SEPKOSKI, Jr. 1993. Insect diversity in the fossil record. Science 261:310–315.
MADDISON, D. R., M. D. BAKER, AND K. A. OBER. 1999. Phylogeny of carabid beetles as inferred from 18S ribosomal DNA (Coleoptera: Carabidae). Syst. Entomol. 24:103–138.
MAGALLÓN, S., AND M. J. SANDERSON. 2001. Absolute diversification rates in angiosperm clades. Evolution 55:1762–1780.
MCALPINE, J. F. 1970. First record of calyptrate flies in the Mesozoic era (Diptera: Calliphoridae). Can. Entomol. 105:105–111.
MOSTOVSKI, M. B. 2000. Contributions to the study of fossil snipeflies (Diptera: Rhagionidae). The genus *Palaeobolbomyia*. Paleontol. J. 34(suppl. 3):360–366.
MOSTOVSKI, M. B., AND E. A. JARZEMBOWSKI. 2000. The first brachycerous flies (Diptera: Rhagionidae) from the Lower Jurassic of Gondwana. Paleontol. J. 34(suppl. 3):367–369.
MUSE, S. V., AND B. S. WEIR. 1992. Testing for equality of evolutionary rates. Genetics 132:269–276.
NAGATOMI, A. 1977. Classification of Lower Brachycera (Diptera). J. Nat. Hist. 11:321–335.
NAGATOMI, A. 1992. Notes on the phylogeny of various taxa of the orthorrhaphous Brachycera (Insecta: Diptera). Zool. Sci. 9:843–857.
NAGATOMI, A. 1996. An essay on phylogeny of the orthrrhaphous Brachycera (Diptera). Entomol. Mon. Mag. 132:95–148.
NAGATOMI, A., AND D. YANG. 1998. A review of extinct Mesozoic genera and families of Brachycera (Insecta, Diptera, Orthorrhapha). Entomol. Mon. Mag. 134:95–192.
PITNICK, S., T. MARKOW, AND G. S. SPICER. 1999. Evolution of multiple kinds of female sperm-storage organs in *Drosophila*. Evolution 53:1804–1822.
REN, D. 1998a. Flower-associated brachyceran flies as fossil evidence for Jurassic angiosperm origins. Science 280:85–88.
REN, D. 1998b. Late Jurassic Brachycera from northeastern China (Insecta: Diptera). Acta Zootaxonomica Sin. 23:65–83.
SANDERSON, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218–1232.
SANDERSON, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. Mol. Biol. Evol. 19:101–109.
SHAW, P. J., A. SALAMEH, A. P. MCGREGOR, S. BALA, AND G. A. DOVER. 2001. Divergent structure and function of the bicoid gene in Muscoidea fly species. Evol. Dev. 3:251–262.
SINCLAIR, B. J., J. M. CUMMING, AND D. M. WOOD. 1994. Homology and phylogenetic implications of male genitalia in Diptera—Lower Brachycera. Entomol. Scand. 24(1993):407–432.
SMITH, S. W., R. OVERBEEK, C. R. WOESE, W. GILBERT, AND P. M. GILLEVET. 1994. The genetic data environment and GUI for multiple sequence analysis. Comput. Appl. Biosci. 10:671–675.
STUCKENBERG, B. R. 1999. Antennal evolution in the Brachycera (Diptera), with a reassessment of terminology relating to the flagellum. Stud. Dipterol. 6:33–48.
STUCKENBERG, B. R. 2001. Pruning the tree: A critical review of classifications of the Homeodactyla (Diptera, Brachycera), with new perspectives and an alternative classification. Stud. Dipterol. 8:3–41.
SWOFFORD, D. L. 2001. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer, Sunderland, Massachusetts.
THORNE, J. L., AND H. KISHINO. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. 51:689–702.
THORNE, J. L., H. KISHINO, AND I. S. PAINTER. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.
WANG, D. Y.-C., S. KUMAR, AND S. B. HEDGES. 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals, and fungi. Proc. R. Soc. Lond. B 266:163–171.
WIEGMANN, B. M., C. MITTER, AND F. C. THOMPSON. 1993. Evolutionary origin of the Cyclorrhapha (Diptera): Tests of alternative morphological hypotheses. Cladistics 9:41–81.
WIEGMANN, B. M., S. C., TSAUR, D. W. WEBB, D. K. YEATES, AND B. K. CASSEL. 2000. Monophyly and relationships of the Tabanomorpha (Diptera: Brachycera) based on 28S ribosomal gene sequences. Ann. Entomol. Soc. Am. 93:1031–1038.
WIKSTRÖM, N., V. SAVOLAINEN, AND M. W. CHASE. 2001. Evolution of the angiosperms: Calibrating the family tree. Proc. R. Soc. Lond. B 268:2211–2220.
WOODLEY, N. E. 1989. Phylogeny and classification of the 'Orthorrhaphous' Brachycera. Pages 1371–1396 *in* Manual of Nearctic Diptera, Volume 3 (J. F. McAlpine and D. M. Wood, eds.). Agriculture Canada, Ottawa.
YANG, L., B. M. WIEGMANN, D. K. YEATES, AND M. E. IRWIN. 2000. Higher-level phylogeny of the Therevidae (Diptera: Insecta) based on 28S ribosomal and elongation factor-1 alpha gene sequences. Mol. Phylogenet. Evol. 15:440–451.
YANG, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J. Mol. Evol 39:306–314.
YANG, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13:555–556.
YEATES, D. K. 2002. Relationships of extant lower Brachycera: A quantitative synthesis of morphological characters. Zool. Scripta 31:105–121.
YEATES, D. K., D. M. MERRITT, AND C. BAKER. 2002. The adult ventral nerve cord as a phylogenetic character in brachyceran Diptera. Org. Divers. Evol. 2:89–96.
YEATES, D. K., AND B. M. WIEGMANN. 1999. Congruence and controversy: Toward a higher-level phylogeny of the Diptera. Annu. Rev. Entomol. 44:397–428.
ZATWARNICKI, T. 1996. A new reconstruction of the origin of the eremoneuran hypopygium and its implications for classification. Genus 7:103–175.

First submitted 26 September 2002; reviews returned 29 December 2002; final acceptance 31 July 2003
Associate Editor: Karl Kjer

APPENDIX. Taxa sampled for sequencing, with GenBank accession numbers and specimen source localities. Vouchers and DNA stocks are stored in the North Carolina State University (NCSU) Insect Collection.

Taxon	Infraorder	Family	Species	GenBank accession no.	Source locality
Basal dipteran outgroups	Tipulomorpha	Tipulidae	*Tipula (Platytipula) paterifera* Alexander	AY456142, AY456152	Maryland
	Psychodomorpha	Anisopodidae	*Sylvicola alternatus* (Say)	AY456141, AY456151	North Carolina
Brachycera	Xylophagomorpha	Xylophagidae	*Coenomyia ferruginea* (Scopoli)	AF238504, AF238526, AF238547	Tennessee
			Arthropeas magnum Johnson	AF238503, AF238525, AF238549	Saskatchewan
	Tabanomorpha	Vermileonidae	*Leptynoma hessei* (Stuckenberg)	AF238506, AF238528, AF238552	South Africa
		Rhagionidae	*Ptiolina fasciata* Loew	AF238508, AF238530, AF238554	Saskatchewan
			Symphoromyia hirta Johnson	AF238512, AF238534, AF238558	Illinois
		Pelecorhynchidae	*Pelecorhynchus personatus* Walker	AF238520, AF238545, AF238569	Australia
		Tabanidae	*Tabanus rufofrater* Walker	AF238513, AF238537, AF238561	Georgia
			Chrysops carbonarius Walker	AF238514, AF238538, AF238562	North Carolina
	Stratiomyomorpha	Pantophthalmidae	*Pantophthalmus sp.*	AF238501, AF238523, AF238547	Costa Rica
		Xylomyidae	*Xylomya parens* (Williston)	AY456143, AY456153	Illinois
		Stratiomyidae	*Pachygaster leachii* (Curtis)	AF238502, AF238524, AF238558	England
	Muscomorpha	Nemestrinidae	*Neorhycocephalus volaticus* (Williston)	AY456145, AY456155	California
		Acroceridae	*Acrocera bulla* Westwood	AY456144, AY456154	California
			Eulonchus sp.	AY456146, AY456156	California
		Asilidae	*Diogmites* sp.	AY456148, AY456158, AY456161	North Carolina
		Apioceridae	*Apiocera haruspex* Osten Sacken	AF266249	California
		Mydidae	*Mydas* sp.	AY456147, AY456157	California
		Bombyliidae	*Heterotropus senex* Melander	AY456150, AY456160	Arizona
			Lordotus sp.	AF503071, AF503026	California
			Bombylius major Linné	AY456149, AY456159	North Carolina
		Therevidae	*Brachylinga* sp.	AF147849	Dominica
		Scenopinidae	*Stenomphrale teutankameni*	AF147824	Israel
		Atelestidae	*Atelestus pulicarius* Walker	AF503033, AF503005, AF502984, AF502963	England
		Empididae	*Gloma fuscipennis* Meigen	AF503047, AF503008, AF502987, AF502966	England
		Dolichopodidae	*Liancalus* sp.	AF503047, AF503011, AF502990, AF502969	California
		Platypezidae	*Paraplatypeza atra* (Meigen)	AF503014, AF502993, AF502972	England
		Syrphidae	*Rhingia nasica* Say	AF503019, AF502998, AF502977	North Carolina
		Muscidae	*Musca domestica* Linnaeus	AF503025, AF503004, AF502983	NCSU lab culture
		Drosophilidae	*Drosophila melanogaster* Meigen	M21017	NCSU lab culture