The Pattern and Timing of Diversification of Philippine Endemic Rodents: Evidence from Mitochondrial and Nuclear Gene Sequences

SHARON A. JANSÅ,1,2 F. KEITH BARKER,1 AND LAWRENCE R. HEANEY3

1Bell Museum of Natural History, University of Minnesota, 1987 Upper Buford Circle, St. Paul, Minnesota 55108, USA; E-mail: jansa003@umn.edu (S.A.J.)
2Department of Ecology, Evolution, and Behavior, University of Minnesota, 1987 Upper Buford Circle, St. Paul, Minnesota 55108, USA;
3Field Museum of Natural History, Division of Mammals, 1400 S. Lake Shore Drive, Chicago, Illinois 60605, USA

Abstract.—The 22 genera and 64 species of rodents (Muridae: Murinae) distributed in the Philippine Islands provide a unique opportunity to study patterns and processes of diversification in island systems. Over 90% of these rodent species are endemic to the archipelago, but the relative importance of dispersal from the mainland, dispersal within the archipelago, and in situ differentiation as explanations of this diversity remains unclear, as no phylogenetic hypothesis for these species and relevant mainland forms is currently available. Here we report the results of phylogenetic analyses of the endemic Philippine murines and a wide sampling of murine diversity from outside the archipelago, based on the mitochondrial cytochrome b gene and the nuclear-encoded IRBP exon 1. Analysis of our combined gene data set consistently identified five clades comprising endemic Philippine genera, suggesting multiple invasions of the archipelago. Molecular dating analyses using parametric and semiparametric methods suggest that colonization occurred in at least two stages, one ca. 15 Mya, and another 8 to 12 million years later, consistent with the previous recognition of “Old” and “New” endemic rodent faunas. Ancestral area analysis suggests that the Old Endemics invaded landmasses that are now part of the island of Luzon, whereas the three New Endemic clades may have colonized through either Mindanao, Luzon, or both. Further, our results suggest that most of the diversification of Philippine murines took place within the archipelago. Despite heterogeneity between nuclear and mitochondrial genes in most model parameters, combined analysis of the two data sets using both parsimony and likelihood increased phylogenetic resolution; however, the effect of data combination on support for resolved nodes was method dependent. In contrast, our results suggest that combination of mitochondrial and nuclear data to estimate relatively ancient divergence times can severely compromise those estimates, even when specific methods that account for rate heterogeneity among genes are employed. [Biogeography; divergence date estimation; mitochondrial DNA; molecular systematics; Murinae; nuclear exon; Philippines; phylogeny.]

Studies of diversification in island archipelagos have held a central place in evolutionary biology for over 150 years and continue to be of great conceptual importance. Many current questions focus on the long-term dynamics of diversification and its impact on species richness (Whittaker, 1998; Heaney, 2000; Lomolino, 2000; Zink et al., 2000; Arbogast and Kenagy, 2001; Ricklefs and Bermingham, 2002, 2004; Riddle and Hafner, 2004), but the number of case studies is small, impeding development of new models (e.g., Fleischer et al., 1998;Losos and Schluter, 2000; Moritz et al., 2000; Price and Cagle, 2002; Steppan et al., 2003). In particular, testing current theories regarding the generation of species richness requires empirical studies in which there are robust phylogenies of multiple clades with reliable estimates for times of divergence events (Donoghue and Moore, 2003). Such phylogenies, when placed in a historical geological context, provide the necessary framework to examine the relative importance of various processes potentially responsible for diversification (e.g., dispersal, vicariance, and sympatric speciation; Arbogast and Kenagy, 2001; Heaney, 2000; Zink et al., 2000).

One potentially useful case study involves the murine rodents (“rats and mice”) of the Philippine Islands. The subfamily Murinae (Rodentia: Muridae) is one of the largest in Mammalia, with over 500 species currently recognized and a native range extending from Africa and western Europe to Australia, New Guinea, and nearby Pacific islands. The phylogeny of this diverse group is poorly known, with most genera not yet placed in a rigorous phylogenetic context (Michaux and Catzeflis, 2001; Jansa and Weksler, 2004; Steppan et al., 2004). The Philippine murine fauna consists of 22 indigenous genera and 64 species (Heaney et al., 1998; Rickart et al., 2005), over 90% of which are endemic to the archipelago. These rodents range in size from about 15 g to 2.3 kg, and occupy habitats from terrestrial burrowers that specialize in feeding on earthworms to arboreal folivores. Morphological and karyological studies summarized below have suggested that several clades of endemic murines may exist in the Philippines (Table 1; Musser and Heaney, 1992; Heaney and Rickart, 1990; Heaney, 2000; Rickart and Heaney, 2002; Rickart and Musser, 1993; Rickart et al., 2005); however, these studies did not propose rigorous phylogenies, and they had no means of estimating times of divergence. The first study to do so used mitochondrial sequences to examine phylogenetic relationships among species in the endemic genus Apomys, and concluded that diversification within this genus principally involved colonization between islands with some intraisland diversification over the past ca. 5 My (Steppan et al., 2003). Additional studies at deeper levels of divergence promise to yield further insights into the patterns and processes of diversification for island taxa.

The Philippine archipelago is an excellent theater for such studies; the geological history is complex but most primary aspects are now well documented (Hall, 1996, 1998, 2002). The islands have arisen by a combination...
of tectonic activities in widely separated parts of the western Pacific Ocean and have been undergoing progressive coalescence, especially during the last 25 million years. Nearly all of the subaerial islands have arisen de novo from below the sea, though some are remnants of continental fragments capped by marine sediments. The islands are now in close proximity to one another (from less than 1 km to 40 to 50 km), thus allowing some interisland colonization, but many have remained isolated by deep channels, which could drastically inhibit gene flow. Further, the islands vary from large (over 100,000 sq km) to small (a few hectare), and so are nearly ideal for measuring the effects of area and isolation, gene flow and genetic differentiation, and patterns of diversification and species richness (Heaney, 1986; Heaney and Rickart 1990; Steppan et al., 2003).

The purpose of this article is threefold: (1) to infer the phylogenetic relationships among genera and selected species of Philippine murines using mitochondrial and nuclear DNA sequences; (2) to estimate times of origin and diversification for the major clades of Philippine murines; and (3) to place this phylogeny in a historical geological context to gain insight into the origin and maintenance of murine diversity in the Philippines. Specifically, we used DNA sequences from the mitochondrial cytochrome _b_ gene and the nuclear-encoded IRBP exon 1 to infer relationships among Philippine murine genera using maximum parsimony, maximum likelihood, and Bayesian methods. To address specific historical biogeographic hypotheses surrounding the origin of murines in the Philippines, we estimated divergence times for several clades using methods that allow for the assumption of a molecular clock to be relaxed (Thorne and Kishino, 2002; Sanderson, 2002). The phylogenetic framework and temporal scale that we present provides a foundation to begin addressing processes of mammalian diversification in the Philippines.

## Materials and Methods

### Taxon Sampling

Our taxon sample includes representatives of all but 4 of the 22 native Philippine murine genera (see Appendix 1 for specimen information). We sampled species within each of these genera so as to encompass the morphological and geographic diversity of each and, when possible, we sequenced more than one individual of each species to verify the accuracy of our sequences. In addition, we included relevant murine genera from Asia and Africa to provide a broader phylogenetic context for the Philippine taxa. To provide appropriate outgroup taxa, we included examples of nearly every subfamily of murid rodent and included several genera of dipodids as an unequivocal root. Our final data set for phylogenetic analysis consisted of 60 species.

### DNA Amplification and Sequencing

DNA was extracted from all tissues using a QiaAmp extraction kit (Qiagen Inc.). The entire cytochrome _b_ gene was amplified using primers MVZ05 and UMMZ04 (Smith and Patton, 1991; Jansa et al., 1999). To generate products of a suitable size for sequencing, the resulting PCR product was used as a template in two subsequent reamplification reactions, one using primer MVZ05 paired with UMMZ12 and one using UMMZ13 paired with UMMZ04 (Jansa et al., 1999). A portion of IRBP exon 1 was amplified using primers IRBPA and IRBPB as published in Stanhope et al. (1996). This PCR product was used in two subsequent PCR reactions, one using primer IRBPA paired with IRBPF and one using primers IRBPE and IRBPP (see Jansa and Voss, 2000, for primer sequences).

Initial PCRs that used genomic DNA as a template were performed as 20-µl reactions using Ampli-Taq Gold DNA polymerase (Perkin-Elmer) and recommended concentrations of primers, unincorporated nucleotides, buffer, and MgCl₂. Initial PCRs were performed using a four-stage touchdown protocol as described in Jansa and Weksler (2004). Reamplification reactions were performed using Taq DNA polymerase (Promega Corp.) in 30-µl reactions for 35 cycles using an annealing temperature of 50°C. The resulting products were sequenced in both directions using amplification primers and dye-terminator chemistry on an ABI 3700 automated sequencer. Sequences were edited and compiled using Sequencher 4.1 (GeneCodes, Inc.). All sequences have been deposited in GenBank with accession numbers DQ191467 to DQ191491 (cytochrome _b_) and DQ191492 to DQ191517 (IRBP) and have been deposited in TreeBASE.

### Phylogenetic and Ancestral Area Analysis

The resulting cytochrome _b_ and IRBP sequences were each aligned with reference to their respective translated amino acid sequences using MacClade 4.03 (Maddison and Maddison, 2001). Aligned sequences were subjected to phylogenetic analysis using maximum parsimony (MP) and maximum likelihood (ML) as implemented in

### Table 1. Classification of Philippine murines according to Musser and Heaney (1992).

| Division I (Old Endemics) | 1. Phloeomys Group: Phloeomys |
| 2. Crateromys Group: Crateromys, Carpomys, Batomys |
| 3. Apomys Group: Apomys |
| 4. Craniomys Group: Craniomys, Archvoldomys |
| 5. Chrotomys Group: Chrotomys, Celaomys (now part of Chrotomys; Rickart et al., 2005) |
| 6. Rhynchomys Group: Rhynchomys |
| Division II (Anonymomys) |
| Division III (New Endemics) |
| 1. Tryphomys and Abditomys |
| 2. Bullinus |
| 3. Tarsomys |
| 4. Limnomys |
| 5. Rattus |

*The genera listed here are those that occur on the oceanic Philippine islands. An additional five genera (Chirpomys, Haeromys, Maxomys, Palasamomys, and Sundamomys) occur on Palawan, which is geographically part of the Sunda Shelf; these taxa are not considered in this report.*
PAUP* ver. 4.0b10 (Swofford, 1999) and Bayesian phylogenetic analysis as implemented in MrBayes ver. 3.0 (Huelsenbeck and Ronquist, 2001). For the MP analysis, all characters were treated as unordered and equally weighted; all parsimony tree searches were heuristic with at least 20 replicates of random taxon addition followed by TBR branch swapping. To determine the best model for use in the ML and Bayesian analyses, we examined the relative fit of various models of nucleotide substitution for the cytochrome b and IRBP data separately and in combination using both hierarchical likelihood-ratio tests (hLRTs) and the Akaike Information Criterion (AIC) as implemented in ModelTest 3.6 (Posada and Crandall, 1998). Log-likelihood scores and estimated parameter values for the various models were calculated based on a neighbor-joining tree of Jukes-Cantor-corrected distances. Once the best-fit model of nucleotide substitution was chosen, we used additional comparisons (based on both hLRTs and the AIC) to evaluate whether a molecular clock could be enforced. Although the two approaches to model selection generally chose the same model, in the one case where they disagreed we performed full likelihood analyses using the model selected by the AIC for reasons outlined in Posada and Buckley (2004). The parameters describing this best-fit model were used in a heuristic tree search using a neighbor-joining tree as a start point for TBR branch swapping as implemented in PAUP*.

Analysis using gene-specific models is not currently implemented in a full likelihood evaluation; therefore, we conducted a Bayesian analysis in which parameters were allowed to vary independently across the two genes. For the Bayesian analysis, we conducted five independent runs of Metropolis-coupled Markov chain Monte Carlo (with four incrementally heated chains each). For each run, we specified a model with six categories of base substitution, a \( \Gamma \)-distributed rate parameter, and a proportion of invariant sites. Uniform interval priors were assumed for all parameters except base composition, which assumed a Dirichlet prior. We decoupled parameter estimation across the data set, allowing substitution parameters to be estimated for the cytochrome \( b \) and IRBP data sets independently. Runs were allowed to proceed for 2 million generations each, and trees were sampled every 100 generations. To evaluate the burn-in for each run, we plotted the log-likelihood values against generation time for each. Based on these results, we discarded at least the first 400,000 generations (4000 trees) from each run and pooled the remaining trees to calculate estimated parameter distributions and posterior probabilities for each node. The final parameter distributions and posterior probabilities for this mixed-model analysis are based on a total of 70,000 trees. To assess consistency of nodal posterior probability estimates across the five runs, we also estimated the correlation amongst probability values from each.

To evaluate whether the mixed-model Bayesian analysis returned different posterior probability estimates than an analysis that assumed a single model across both genes, we conducted a Bayesian analysis recognizing only a single partition. Prior probability values were assigned as above, and a single run of four incrementally heated chains was allowed to proceed for 2 million generations, sampling trees every 100 generations. The initial 300,000 generations (3000 trees) were discarded as burn-in, and the remaining trees were used to calculate estimated parameter values and posterior probabilities for each node.

Bootstrap values (Felsenstein, 1985) were calculated for all three data sets under the parsimony criterion using 1000 pseudoreplicates (heuristic searches with five random addition replicates, TBR branch swapping). For maximum likelihood analysis, bootstrap values were calculated by generating 100 bootstrap replicates using the seqboot program in PHYLIP 3.6 (Felsenstein, 1993), finding the best-fit tree for each under the GTR+I+\( \Gamma \) model in PhyML (Guindon and Gascuel, 2003) and calculating the 50% majority rule tree from the resulting trees in PAUP*.

To infer the ancestral area for the Philippine taxa, each of the five clades of Philippine endemics (A to E; see below) was treated in a separate ancestral area analysis using DIVA (Ronquist, 1996, 1997). We coded terminal taxa in each of these clades as occurring on Mindanao, Luzon, Sibuyan, or as being widespread across the Philippine archipelago. In the case of Apomys, where we lacked substantial species-level sampling, we coded the ancestral area for the genus as Luzon based on prior biogeographic studies (Steppan et al., 2003). We did not attempt ancestral area reconstruction for Crumomyys, as we lack the single species of the genus that occurs on Sulawesi.

Estimating Dates of Divergence

Ever since the molecular clock theory was first proposed (Zuckerkandl and Pauling, 1965), there has been great interest in using biochemical data to estimate absolute divergence times among taxa. Despite accumulating DNA sequence evidence that demonstrates substantial rate variation both among genes (Yang and Nielsen, 1998; Rausher et al., 1999; Baker et al., 2001; Lin and Danforth, 2004;) and across lineages (Wu and Li, 1985; Britten, 1986), this interest has not diminished; rather, it has spurred the development of methods that account for rate heterogeneity among lineages (e.g., Sanderson, 1997, 2002; Thorne et al., 1998; Yoder and Yang, 2000; Kishino et al., 2001) and across multiple genes or gene partitions (Thorne and Kishino, 2002; Yang and Yoder, 2003). These methods differ in whether they require an explicit model of sequence evolution (Thorne et al., 1998; Thorne and Kishino, 2002; Yoder and Yang, 2000), are nonparametric (Sanderson, 1997), or combine parametric and nonparametric methods through a semiparametric approach (Sanderson, 2002). Of the parametric methods, some allow rates to vary from branch to branch (Thorne and Kishino, 2002), and others require that local rates be subjectively assigned to different parts of the tree (Yang and Yoder, 2003).
Our data set consists of two genes with different rates of evolution and substitution dynamics; therefore, we utilized the parametric, multigene, Bayesian approach of Thorne and Kishino (2002) executed in the programs estbranches and multiditime to estimate divergence times among lineages. We compared the results of this method with those from the semiparametric, penalized-likelihood (PL) approach of Sanderson (2002) to explore how different assumptions may affect divergence time estimates. For both methods, we used the topology resulting from the maximum likelihood analysis of the combined gene data set with additional placental taxa (Equus, Tapirus, Homo, Macaca) and a marsupial outgroup (Monodelphis) added to the base of the tree according to the phylogeny published by Springer et al. (2003).

For the PL analysis, we separately optimized branch lengths for IRBP, cytochrome b, and both genes together and performed dating analyses using each of these three trees. We established the best value for the smoothing parameter for each data set by using the cross-validation procedure implemented in the program r8s with log10 values ranging from −5 to 10. All PL analyses used the truncated Newton (TN) algorithm (Sanderson, 2002) with 10 random restarts and 10 subsequent perturbations. To estimate the sampling error of divergence time estimates, we constructed 100 bootstrapped data sets using the seqboot program from PHYLIP 3.6 (Felsenstein, 1993), and estimated branch lengths from each bootstrapped data set on the combined-gene likelihood topology using PAUP*. We then performed the dating analysis on each of these data sets and summarized the node statistics using the profile command in r8s.

For the parametric Bayesian analysis, we first estimated the branch lengths and their variance-covariance matrix for each data set on the combined-gene likelihood topology using the program estbranches. This program requires that parameter values for the substitution model be specified, and is limited to the F84+Γ model; therefore, we estimated parameter values for each data set with the baseml program of PAML 3.13 (Yang, 1997) using the F84 substitution model and a five-category Γ distribution. The variance-covariance matrices produced by estbranches were used as input to the program multiditime, which implements Markov Chain Monte Carlo (MCMC) sampling to estimate the posterior distribution of rates and divergence time estimates. Bayesian analyses of the two genes in combination were not based on averaged branch lengths from a uniform-model analysis (as in the PL analyses) but were accomplished by simultaneous input of the two separate variance-covariance matrices derived from each gene.

Multiditime requires that a prior distribution be specified for the age of the ingroup node and the rate of evolution at the ingroup node. We set 95 million years as the mean and standard deviation for the prior distribution for the time separating the ingroup (placental mammals) from the outgroup (Monodelphis) based on results reported by Springer et al. (2003). To establish the mean and standard deviation for the prior distribution of the rate of evolution at the ingroup node, we calculated the path lengths from the ingroup node to each tip on the ML tree and divided these values by the estimated time of divergence for the ingroup (95 Mya). We used both the median of these rates and the average rate with rodent lineages downweighted by 1/15th (the ratio of rodents to nonrodent outgroups). We calculated these values for each gene independently and used the weighted (by number of sites) average of these estimates as the value for both genes in combination. We sampled each Markov chain 10000 times and allowed 100 cycles between each sample; we omitted the first 1000 samples as burn-in. Runs were performed without conditioning on the data (prior runs) in order to verify performance of the MCMC sampler.

We chose calibration points both inside and outside the rodent radiation (see below for discussion of the divergence dates we selected). For primates, we set the divergence between Old World monkeys and apes between 21 and 29 Mya. We based the minimum estimate on the earliest appearance of hominoids at 21 Mya (Gebo et al., 1997) and the maximum estimate on a combination of fossil and molecular data (Pilbeam and Young, 2004). For perissodactyls, we set the range of divergence dates between horses and tapirs from 50 to 58 Mya based on fossil data (Prothero and Schoch, 1989; Janis et al., 1998). Within rodents, we used 12 Mya as the first unambiguous appearance of murines in the fossil record and set this as the minimum date of origin for the murine crown group (Jacobs, 1977; Jacobs and Downs, 1994). We used an additional minimum divergence date within murines of 6 Mya for the appearance of Otomys (Pocock, 1976). We constrained the origin of the crown clade Myomorpha, corresponding to the split between dipodids and muroids, between 35 and 70 Mya. The minimum of this range is based on the appearance of late Eocene fossils that represent early dipodids and muroids (reviewed in Flynn et al., 1985). The older end of this range cannot be unambiguously defined, but 70 Mya doubles the date of the first appearance of primitive muroids and has been used in other studies of rodent divergence time estimation (Steppan et al., 2004).

RESULTS

Phylogenetic Results

Parsimony analysis of the 568 informative cytochrome b characters resulted in three minimum-length trees with length = 6844, CI = 0.16, and RI = 0.34. The strict consensus of these trees shows no resolution among muroid subfamilies, and 41 of a possible 58 (70%) nodes are resolved (Fig. 1A). Parsimony analysis of the 499 informative IRBP characters resulted in 687 trees with length = 2142, CI = 0.42, and RI = 0.64. A strict consensus of these trees shows 49 of 58 (84%) nodes resolved (Fig. 1B). In contrast to the tree based on cytochrome b sequences, the IRBP tree shows well-supported resolution among subfamilies of muroid rodents, but relatively little resolution among genera and species. When all informative characters were analyzed in combination, we obtained a
FIGURE 1. Trees resulting from parsimony analysis of the cytochrome $b$ data set (A, strict consensus of three minimum-length trees), the IRBP data set (B, strict consensus of 687 minimum-length trees), and the combined-gene data set (C, single minimum length tree). Branch thickness and shape indicate nodal support as measured by bootstrap percentage (BP): A thin line indicates BP < 50%; a thickened line indicates 50% < BP < 75%; a thickened line with a terminal buttress indicates BP ≥ 75%. Numbers on the branches refer to node numbers in Table 3; these numbers are continued from Figure 2 and indicate nodes that do not appear on the maximum likelihood tree shown in that figure. Taxa that are native to the Philippines are shown in bold.

A. Cytochrome $b$

B. IRBP

C. Combined

non-murines and outgroup taxa

non-murines and outgroup taxa

non-murines and outgroup taxa

single, fully resolved tree with length = 8838, CI = 0.22, and RI = 0.42 (Fig. 1C).

For the cytochrome $b$ data, both hLRTs and the AIC identified the parameter-rich GTR+$\Gamma$ with no clock model as the best-fit model of nucleotide substitution. However, the two approaches to model selection arrived at different best-fit models for the IRBP data. Hierarchical likelihood-ratio testing chose a model with equal base frequencies, the same transition rate, and four different transversion rates, plus a proportion of invariant sites and a gamma-distributed rate parameter (TVMef+$\Gamma$ with no clock), whereas the AIC resulted in the same model as applied to the cytochrome $b$ data (GTR+$\Gamma$ with no clock). For reasons outlined in Posada and Buckley (2004), we conducted full-likelihood evaluation using the model chosen using the AIC (parameter values reported in Table 2). In general, nodes that are well supported in the separate parsimony analysis of each gene also receive high support in the likelihood analysis of that gene; however, differences resulting from analytical method are more apparent for the cytochrome $b$ data than for IRBP (Table 3). For both genes, likelihood bootstrap values are generally higher than those recovered under parsimony analysis; however, this pattern affects more nodes and is of greater magnitude for the cytochrome $b$ data than the IRBP data (Table 3). Trees resulting from separate likelihood analysis of each gene are largely congruent, but the two trees do have topological differences (results not shown); with the exception of one node describing relationships among species of *Apomys*, none of these conflicting nodes have strong (>75%) bootstrap support from either gene.

For the combined data set, the GTR+$\Gamma$ model with no clock was the best-fit model according to both hLRTs and AIC. The estimated substitution parameter values from maximum-likelihood analysis of each gene separately (cytochrome $b$ and IRBP, respectively) and of the combined-gene data set are shown in Table 2.
and the AIC (parameter values given in Table 2). However, the mixed-model Bayesian analysis indicates that IRBP and cytochrome b differ significantly in all but three estimates of parameter values for the substitution model (Table 4). Therefore, it is reasonable to ask whether a likelihood analysis that employs a single model across both genes will result in a reasonable phylogeny and nodal support estimates. One way to address this question is to compare the estimated Bayesian posterior probabilities for each node as calculated under a single model versus those calculated under a model that allows parameters to vary between the two genes. Comparison of posterior probability estimates for each node between these two analyses indicates that partitioning leads to differing interpretations of phylogeny at only two nodes (9 and 16; Table 3), suggesting that assuming uniformity does not markedly affect our phylogenetic conclusions (we note that nodal estimates are highly correlated across runs within analyses; \( r^2 \approx 1.0 \) for each). However, it is also reasonable to ask whether the single-model Bayesian analysis is an appropriate comparison to a full maximum likelihood evaluation of these data. Comparing parameter estimates from the single-model Bayesian analysis and the single-model likelihood analysis reveals that only two of the maximum likelihood estimated values are marginally outside the 95% confidence interval of those calculated from the Bayesian analysis (Table 4). Moreover, the maximum likelihood topology falls within the 95% credibility interval of trees recovered from the Bayesian analysis: the posterior probability of the ML tree is 0.2%. Thus, even though model parameter estimates differ between the two genes, these differences have relatively little effect on the resulting topology, and from a phylogenetic standpoint the combined-gene data set can be reasonably analyzed using a single substitution model.

The tree resulting from a combined-data likelihood analysis (Fig. 2) is highly congruent with that resulting from the combined-data parsimony analysis, and nodes that are well supported in one analysis are also well supported in the other (Table 3). Topologies resulting from these two analyses differ in only two nodes: (1) the position of Rattus preator, and (2) the position of Archboldomys...
relative to *Chrotomys* and *Rhynchomys*. Neither of the alternative placements for these taxa is well supported with either bootstrap percentages (<50%) or estimated Bayesian posterior probabilities (<0.95; Table 3).

Murine monophyly is well supported by IRBP and by the combined data when analyzed using parsimony, maximum likelihood, or Bayesian approaches (node 35; Table 3); however, the cytochrome *b* data fail to recover murine monophyly in any analysis, nor do they provide a strongly supported alternative arrangement among murines and other muroid subfamilies. The IRBP and combined data analyses recover five major clades of rats and mice that are native to the Philippines; these are labeled A to E in Figure 3: all but one of these (clade A) are recovered by the cytochrome *b* data as well. Relationships within and among these clades are discussed below.

**Divergence Time Estimates**

We used the topology resulting from the likelihood analysis of the combined-gene data set as the basis for all divergence time analyses; however, we compared results using branch length estimates from each gene and from the combined-gene data set (Table 5). The two methods of divergence time estimation (PL and Bayesian) returned similar date estimates for all nodes (Table 5). In addition, Bayesian estimates using a prior rate of evolution based on the median of rates across taxa did not differ significantly from those that were based on a prior rate using the average rate with rodent lineages downweighted (results not shown). In contrast, divergence time estimates differed dramatically depending on whether branch lengths were based on cytochrome *b*, IRBP, or the combined data (Table 5). The earliest divergence dates were estimated from the cytochrome *b* data, the most recent from IRBP, with estimates from the combined-gene data set falling somewhere between the two. An initial PL analysis using cytochrome *b* branch lengths resulted in a basal divergence time estimate that was unreasonably old (root node = 400 Mya); therefore, we constrained all PL estimates from cytochrome *b* to have a maximum root node divergence of 100 Mya (Springer et al., 2003). Analyses based on IRBP or the combined-gene data set did not require this additional constraint.

We doubt the accuracy of most divergence date estimates based on trees with branch lengths estimated from the cytochrome *b* data and the combined-gene data set for the following reasons. First, rate heterogeneity across sites in the cytochrome *b* and combined data sets is high (*α* = 0.444; Table 2), and variable sites are saturated for most inter-generic comparisons (results not shown). Second, the optimal PL smoothing values for cytochrome *b* (smoothing = 0.01) and the combined data (smoothing = 100) were considerably lower than that from IRBP alone (smoothing = 106), suggesting that cytochrome *b* severely violates the molecular clock among higher taxa. Likewise, the posterior distribution of the autocorrelation parameter *v* has a mean of 0.04 (0.02–0.07, 95% credibility interval) for cytochrome *b*, and 0.29 (0.15–0.51) for IRBP, reflecting the extreme rate heterogeneity of the mitochondrial gene relative to the nuclear. Finally, as noted above, the unconstrained PL analysis of the cytochrome

---

**TABLE 5. Comparison of stem-group divergence time estimates from nonparametric penalized likelihood (PL) and parametric Bayesian (PB) methods using all calibration points (see Materials and Methods and Fig. 3 for calibration dates used). Dating analyses were performed on the tree resulting from likelihood analysis of the combined data with branch lengths estimated from each of three data sets (cytochrome *b*, IRBP, and combined). Values shown are the average and standard deviation (in parentheses) in millions of years. Note that clade D corresponds to the divergence between *Mus* and *Rattus*.**

<table>
<thead>
<tr>
<th>Node</th>
<th>Cytochrome <em>b</em></th>
<th>IRBP</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PL PB PL PB PL PB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clade A</td>
<td>16.4 (2.9) 18.4 (3.0) 4.0 (1.2) 3.1 (1.3) 10.1 (0.7) 12.9 (2.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clade B</td>
<td>21.0 (3.6) 22.4 (3.5) 3.2 (0.7) 3.1 (1.3) 11.7 (0.8) 15.6 (2.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clade C</td>
<td>17.0 (3.5) 19.8 (3.5) 8.4 (1.3) 7.1 (2.2) 11.7 (1.1) 15.6 (2.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clade D</td>
<td>42.4 (13.8) 42.9 (4.0) 20.5 (1.6) 15.8 (3.4) 29.2 (1.4) 34.6 (3.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clade E</td>
<td>44.6 (10.8) 46.3 (3.9) 23.2 (1.7) 18.7 (3.8) 32.7 (1.5) 37.0 (3.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus</td>
<td>47.9 (3.1) 50.2 (4.0) 29.0 (2.2) 22.6 (4.3) 39.6 (1.9) 41.1 (4.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myodonta</td>
<td>70.0 (0.0) 65.8 (3.4) 65.1 (5.4) 52.6 (6.7) 70.0 (1.8) 61.9 (4.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horse-Tapir</td>
<td>50.0 (1.9) 52.5 (2.0) 58.0 (1.2) 54.1 (2.3) 52.8 (3.0) 53.4 (2.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ape-Monkey</td>
<td>29.0 (2.8) 25.1 (2.2) 21.0 (0.0) 24.3 (2.2) 28.2 (1.8) 25.2 (2.2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Figure 3.** The rate-smoothed tree resulting from parametric Bayesian estimation of divergence dates based on the IRBP data set. Scale bar is in millions of years. Numbers at nodes indicate assumed calibration points in millions of years, with numbers in circles representing assumed minimum dates and oblongs spanning the ranges between constrained minimum and maximum dates. Letters L, M, and S after each taxon name indicate geographic provenance for that taxon (Luzon, Mindanao, or Sibuyan, respectively); the same letters on nodes indicate reconstructed ancestral areas. Clades shaded in grey are endemic to the Philippines, and are labeled A, B, C, D, and E as referred to in the text. Inset: Histogram of estimated Bayesian posterior distributions for divergences of Clades A to E. The shading of each bar indicates the proportion of generations from the MCMC analysis each clade was assigned a given age.

*b* data yielded unreasonably old estimates of basal divergences (400 million years at the root), suggesting that the smoothing algorithm was maximizing fit by imposing unreasonably low evolutionary rates for basal branches of the tree. For these reasons, we report divergence time estimates from IRBP branch lengths in Figure 3 as our best estimate of divergence dates among these higher taxa.

**DISCUSSION**

**Phylogenetic Relationships**

Although many of the Philippine murines have been known since early expeditions (Thomas, 1898), Musser and Heaney (1992; hereafter referred to as MH) established the first working hypothesis for relationships among them. MH identified a number of morphological
characteristics that they used to diagnose and cluster genera of Philippine murines; however, they stopped short of a complete phylogenetic analysis of these characters, and only a generalized phylogram of the results has been presented (Heaney and Rickart, 1990; Heaney, 2000). Thus, the analyses presented in this report provide the first explicitly phylogenetic test of MH’s ideas.

MH’s scheme splits the endemic Philippine rodents into three divisions with various groupings of taxa within each division (Table 1). Their Division I contains those taxa “in which the species retain many basic features primitive for rodents yet also exhibit unique combinations of specializations, some of them spectacular” (Musser and Heaney, 1992:56). MH did not diagnose these taxa as a natural group but simply set them apart as distinct from the other Philippine murines. Within Division I, MH recognized several distinct groups, including the arboreal, folivorous cloud (Phloeomys and Crateromys) and tree (Carposomys) rats; the insectivorous and vermivorous shrew rats (Archboldomys, Crunomys, Rhynchomys, Chromomys); and the forest mouse Apomys (Table 1). MH could not find obvious close relatives to these taxa among the living Australasian murines and used this observation plus the apparent antiquity of the group as the basis for naming Division I the Philippine “Old Endemics.” MH suggested that these Old Endemics originated in the late Miocene or early Pliocene and proposed that their phylogenetic ties would be found among the Old Endemic hydromyines and conilurines of Australia and New Guinea.

MH echoed earlier accounts (Thomas, 1898) that described Phloeomys as a unique and isolated lineage of murine rodent; however, they suggested that of the other native Philippine murines, Phloeomys was probably most closely related to species in the Crateromys Group. Nonetheless, subsequent studies of karyotypes showed that Chrotomys and Phloeomys share a number of derived chromosomal features, but that Archboldomys (part of the Crunomys group) is karyotypically unique among Philippine murines. We find a clade of shrew rats similar to that originally suggested by MH (node 22; Fig. 2), except we exclude Crunomys from this clade. Although MH were hesitant about associating Apomys with this group, our results clearly place Apomys as the sister taxon to the shrew rats. Chromosomal variation among species of Apomys is impressive, but some species share derived chromosomal elements with Rhynchomys and Chrotomys, suggesting an affiliation between the three (Rickart and Musser, 1993; Rickart and Heaney, 2002). Additionally, these three genera plus Archboldomys show a number of derived morphological traits presumably associated with ground-dwelling and vermivory. The rostrum and dentary are elongated (greatly so in Rhynchomys); the dentition has been simplified, with either a reduced (Apomys, Archboldomys, and most Chrotomys species) or missing (Rhynchomys and some Chrotomys species) third molar; the tail is short (except in some Apomys) and thinly haired; and the hind foot is long and narrow.

The most striking departure of our tree from MH’s hypothesis is in the placement of the shrew rat Crunomys (clade C; Fig. 3). Musser and Heaney considered Crunomys to be most closely allied with the shrew rat Archboldomys erected the “Crunomys group” of Old Endemics to contain them, and listed a number of features of the dentition, cranial architecture, and foot morphology to unite the two, though they noted that many of the shared features were probably simplifications or losses. In contrast, Rickart and Heaney (2002) found substantial karyotypic differences between the two genera, and our phylogenetic hypotheses from mtDNA and nuclear IRBP sequences, as well as both combined, fail to place Crunomys as the sister taxon to Archboldomys. Rather, our phylogeny unites Crunomys with Maxomys as part of a large clade containing New Endemic taxa from the Philippines and Sunda Shelf. The difference between our molecular results and MH’s morphological hypothesis may stem from the fact that Crunomys exhibits many of the same morphological simplifications as Archboldomys, including characteristics of the molars, tail length, and hind foot shape. Our molecular results reveal these traits to be either synapomorphies or homoplastic for these taxa.

Among the Old Endemic taxa, Musser and Heaney erected a group of “shrew rats” to encompass those Philippine rats with a largely vermivorous diet (Rickart et al., 1991) and terrestrial habits, consisting of the Chrotomys, Rhynchomys, and Crunomys groups (Table 1). Chromosomal studies showed that Chrotomys and Rhynchomys share a number of derived chromosomal features, but that Archboldomys (part of the Crunomys group) is karyotypically unique among Philippine murines.

Our phylogeny unites Crunomys, Archboldomys, and most Chrotomys species as part of the Archboldomys group, our results place Crunomys as the sister taxon to Archboldomys. Rather, our phylogeny unites Crunomys with Maxomys as part of a large clade containing New Endemic taxa from the Philippines and Sunda Shelf. The difference between our molecular results and MH’s morphological hypothesis may stem from the fact that Crunomys exhibits many of the same morphological simplifications as Archboldomys, including characteristics of the molars, tail length, and hind foot shape. Our molecular results reveal these traits to be either synapomorphies or homoplastic for these taxa.

MH’s Division II contains the single genus Anonymomys, for which we were unable to obtain fresh tissues for our molecular work. MH could not find characters to link Anonymomys with any other Philippine murine, but in an earlier report, Musser and Newcomb (1983) tentatively identified the Sundaic genus Haeromys as its sister taxon. Clearly, molecular data from Anonymomys and additional murines from islands of the Sunda Shelf and
Indochina are critically needed to assess the phylogenetic position of this unique genus.

Finally, MH grouped several genera with “Rattus-like” features as Division III. Genera in this group have historically been associated with Rattus to various degrees (Ellerman, 1941; Misonne, 1969). Some had been described as species of this genus, thereby obscuring the diversity of the Philippine and Sundaic murine fauna. Careful diagnostic work (Musser and colleagues; see citations from Musser and Carleton, 1993) has remedied much of this confusion by erecting several new genera and narrowing the contents of the genus Rattus. MH described the Philippine New Endemic murines as a cluster of taxa that appear to share a number of derived characteristics of the skull and dentition, though they were hesitant to define this group as a lineage apart from the other New Endemic taxa that occur throughout Australasia. They speculated that these taxa appeared in the Philippines sometime after the origin of the Old Endemic taxa, perhaps arriving in the late Pliocene or early Pleistocene.

We lack molecular data from two genera, Tryphomys and Abdimomys, that MH assigned to this group. These two are most likely sister taxa (Musser and Newcomb, 1983), and MH list a number of characters that unite the two. Beyond this grouping, however, MH were “unable to satisfactorily cluster any of the other genera in Division III” (p. 123), and subsequent chromosomal data showed some similarity among all New Endemic genera without resolving intergeneric relationships (Rickart and Musser, 1993; Rickart and Heaney, 2002). Our tree fails to unite the Philippine New Endemics as a single clade, but intersperses these genera among taxa that have been characterized as New Endemics from elsewhere in Australasia (Musser and Newcomb, 1983). Within this large clade of widespread New Endemics, we recover a well-supported clade including the Philippine taxa Limnomys, Tarsomys, and Rattus everetti—one of three species of Rattus endemic to the Philippines. This clade is nested within other species of Rattus; however, neither our taxon sampling (the currently defined genus Rattus is large—containing nearly 60 described species—and complex, with several unresolved taxonomic problems; Musser and Carleton, 1993) nor present nodal support values allow us to place it with any confidence.

The remaining New Endemic genus Bullimus is clearly not part of this clade, nor is it particularly closely related to the species of Rattus that we included; rather, Bullimus appears as the sister taxon to the Sundaic genus Sundamys. This relationship is surprising. Although Misonne (1969) placed species now referred to Sundamys in the subgenus Bullimus of Rattus, subsequent revisionary work (Musser and Newcomb, 1983; Musser and Heaney, 1992) recognized both Sundamys and Bullimus as genera distinct from Rattus. Specifically, Musser and Newcomb concluded that, “there is no evidence that the Philippine Bullimus is phylogenetically closely allied to the Malaysian Sundamys” (p. 510). Our molecular data support a link between the two, but a much broader sampling of Sundaic murines will be needed to fully comprehend the phylogenetic position of Bullimus.

Divergence Dates

Divergence dates within rodents are extremely controversial and depend on a number of factors, including the type of data, methods of estimation, and calibration points used. For murine rodents, dates based on molecular data are typically much older (22 to 51 Mya for the Mus-Rattus divergence; Janke, 1994; Kumar and Hedges, 1998; Cao et al., 2000; Adkins et al., 2003) than those inferred from the fossil record (12 to 14 Mya; Jacobs and Downs, 1994; Flynn et al., 1985). This discrepancy has been attributed to application of a uniform molecular clock across mammals (Steppan et al., 2004), even though murine rodents appear to have an accelerated rate of molecular evolution (e.g., Wu and Li, 1985). Although this explanation certainly accounts for some of the extreme dates of murine diversification based on molecular data (e.g., 18 to 52 Mya, Janke et al., 1994; 39.8 to 41.6 Mya, Kumar and Hedges, 1998), it does not appear to explain the discrepancy entirely. Studies that use relaxed molecular clock methods, either Bayesian (Cao et al., 2000, Springer et al., 2003; Hasegawa et al., 2003) or penalized-likelihood methods (Adkins et al., 2003), and nonmuroid calibration points also recover relatively old dates for the origin of murines (e.g., 21 to 33 Mya, Cao et al., 2000; 12 to 20.7 Mya, Springer et al., 2003; 13.4 to 19 Mya, Hasegawa et al., 2003; 12 to 37 Mya; Adkins et al., 2003).

Discrepancies in estimated dates of divergence also appear to be attributable to how the various researchers treat calibration points, particularly the appearance of Progonomys (the first known fossil with unambiguously murine affinities; Jacobs and Downs, 1994) in the fossil record at 12 Mya. Studies that use relaxed molecular clocks but employ 12 Mya as an absolute date of divergence for murines (e.g., Steppan et al., 2004; Chevret and Dobigny, 2005) naturally infer this as the date of murine origin (or, incorrectly, as the date of the Mus-Rattus split if the taxon sample did not include more basal murine lineages). In contrast, studies that do not use this murine calibration point (Adkins et al., 2003), or that treat it as a minimum rather than an absolute date of divergence (Springer et al., 2003; this study), infer dates of murine origin that are up to 10 to 15 million years older.

The argument that 12 Mya represents the absolute date of divergence for crown murines relies on a literal interpretation of the muroid fossil record (e.g., Steppan et al., 2004) that rests on two assumptions. First, this interpretation assumes that Progonomys (Siwalik Group, Pakistan, 11.8 Mya; Jacobs and Downs, 1994) and not Antemus (Siwalik Group, Pakistan, 13.75 Mya; Jacobs and Downs, 1994) represents the first occurrence of crown murines in the fossil record (see Flynn, 1985; Jacobs, 1977 for an alternative interpretation). Second, and more problematic, this interpretation assumes that the fossil record of murines is adequately sampled worldwide to conclude that murines were not present before the
appearance of these taxa. Conflict between molecular clock estimates and the fossil record must be judged in light of the temporal and spatial sampling of the fossil record (Foote et al., 1999; Tavaré et al., 2002). We are unaware of any quantitative evaluation of the quality of the rodent fossil record, but given the spatial and temporal patchiness of the record for many mammalian groups, particularly in southern continents and the tropics (Tavaré et al., 2002), we find the discrepancy between recent molecular dates (including ours) and those implied by the literal interpretation of the fossil record to be unsurprising. Given the observation that none of these fossils have been placed phylogenetically, and that the dental characters used to identify them are ambiguous and possibly convergent (Denys et al., 1995; Jansa and Weksler, 2004), we prefer the conservative interpretation that the appearance of Progonomys at 12 Mya represents the minimum date of divergence for the murine radiation. Our estimated dates of divergence are in line with those inferred from recent molecular dating studies that employ relaxed-clock methods (see citations above) and predate those inferred by Steppan et al. (2004) by 7 to 10 My. Future analyses including additional rodent species, calibration points, and sequence data will be necessary to reconcile this discrepancy.

**Biogeography**

The Philippine archipelago first originated as land above sea level during the late Eocene or early Oligocene, but much of the growth of the islands was the result of tectonic and volcanic activity during the Miocene and Pliocene. All islands except Palawan are of oceanic origin and have never been connected to the Asian mainland. The geological history of the archipelago over the last 25 million years is extraordinarily complex (Hall, 1996, 1998, 2002). Volcanic activity and tectonic uplift have formed new islands and added to existing ones, tectonic motion has moved island fragments throughout the region, and sea level changes have caused the joining and fragmentation of land masses. Each of these geological factors may have played a role in restricting or facilitating dispersal among islands and in the autochthonous diversification of murines in the Philippines (Heaney, 1986, 2000; Heaney and Rickart, 1990; Steppan et al., 2003).

Musser and Heaney (1992) hypothesized that the Philippine murines colonized the islands from the Asian mainland in at least two waves, the first giving rise to the Old Endemic taxa with a subsequent wave founding the New Endemics. They suggested that the Old Endemics may have originated during the Miocene and early Pliocene, “during that time when clusters of land began to appear above sea level to form insular precursors of the present-day Archipelago” (Musser and Heaney, 1992:56), whereas the New Endemics arrived much later during the late Pliocene or early Pleistocene, though they cautioned that this estimate was “pure speculation based on known sea level lowerings” (Musser and Heaney, 1992-95). By adding a time dimension to our phylogenetic hypothesis, we can explicitly test these hypotheses and address whether certain geological events may have played a role in shaping the diversity of murine rodents in the Philippines.

Our divergence time analysis suggests that murine rodents originated ca. 22.6 ± 4.3 Mya and that the earliest Old Endemic clade (clade E) arose on the Philippines shortly thereafter (ca. 18.7 ± 3.8 Mya; Table 5). According to our analysis, the second clade of Old Endemics (clade D) must have arisen subsequent to clade E (15.8 ± 3.4 Mya; Table 5); however, our analysis does not provide sufficient precision to distinguish between the times of origin for these two clades (Fig. 3, inset). Recent results (Steppan et al., 2005) suggest that a clade comprising murine taxa from New Guinea subdivides the branch leading to the Philippine taxa in clade D. Therefore, our estimate for the origin of this Philippine clade may be too old. However, these New Guinean taxa appear to bisect the branch leading to Philippine taxa in clade D (see Fig. 6 in Steppan et al., 2005), and based on our chronogram (Fig. 3) our estimate is likely no more than 2.4 My too old. We clearly identify a second wave of dispersal into the islands to found the genus Bullimus (clade B) and the taxa in Clade A (Fig. 3, inset). Although our tree indicates that these were separate invasion events, we cannot distinguish between the times of origin for clades A and B (Table 5); they both occurred around 3.1 ± 1.3 Mya. We date the origin of Crunomys (clade C) in the Philippines at 7.1 ± 2.2 Mya, but note that the confidence interval from the IRBP parametric Bayesian (PB) analysis for this estimate overlaps with those from the New Endemic clades A and B (Fig. 3, inset). Moreover, this may be an overestimate for this clade’s origin into the Philippines as we lack samples from two species of Crunomys, including the single species native to Sulawesi.

Our results confirm that there have been at least two periods when murines colonized the Philippines, one 15 to 20 Mya during which the ancestors of the Old Endemics arrived, and one approximately 3 Mya when the founders of the New Endemic lineages arrived (Fig. 3, inset). Our phylogenetic hypothesis suggests that there were multiple invasions during both of these time periods: at least two invasions were required to found the Old Endemic clades (clades D and E), and at least three to found the New Endemic lineages (clades A, B, and C). Further, our results confirm that most of the diversity of Philippine murines is associated with speciation that took place within the archipelago, with the two Old Endemic clades producing at least 15 and 25 species (the “folivore” and “vermivore” clades, respectively), and the three New Endemic clades producing 3 to 5 species each. Additionally, our results support earlier conclusions that the diversification took place over periods of time measured in millions of years, not during the middle and late Pleistocene but during the Pliocene and Miocene, and that the dynamics of species richness is best measured on that timescale (Heaney, 1986, 2000; Heaney and Rickart, 1990; Steppan et al., 2003).

The geography of the Philippine archipelago during the timespan corresponding to the colonization of the
Old Endemic clades (ca. 15 Mya) was strikingly different from its modern geography (Hall, 1998; Fig. 4A). Fifteen million years ago, the only subaerial land corresponding to modern Philippine islands occurred in northern Luzon and as a distant landmass that became part of modern Mindanao (Fig. 4A). The landmass of northern Luzon was well separated from the remaining proto-Philippine islands and other Australasian islands by large expanses of open water (Fig. 4A). Intriguingly, living members of the two Old Endemic clades occur primarily on Luzon and reconstruction of the ancestral area for each of these clades is Luzon (Fig. 3). One possible interpretation of our results is that taxa in these Old Endemic clades are the last living portion of a remarkably diverse radiation that was widespread across Australasia, and whose members on the Asian mainland are extinct. This interpretation is consistent with a recent study (Steppan et al., 2005) showing several New Guinean and Australian murines to be sister to taxa in the Old Endemic clade D.

The extent and position of the Philippine islands during the timespan corresponding to invasion of the New Endemic taxa (ca. 3 Mya) is very similar to their modern placement (Fig. 4B). Large, subaerial landmasses existed that correspond to the northern and southeastern arms of Luzon, the islands of Mindoro, eastern Mindanao, and several of the smaller Philippine islands. Although these landmasses had not yet reached their modern positions, they are close to each other and to the Sunda Shelf. Thus, several Philippine landmasses were available for colonization during this timespan, and several avenues of colonization existed, including pathways across shallow seas from Sulawesi, island-hopping across small islands between Borneo and the southern Philippines, as well as direct colonization across broad seas. Testing the role that Sulawesi played in providing a source population for Philippine murines will require including the single species of *Crunomys* and additional murine genera from that island. Similarly, testing Sunda Shelf associations will require far denser sampling of modern Sundai taxa. Each of the New Endemic clades has members on Mindanao and Luzon, in addition to several of the smaller Philippine Islands. Our present taxon sample does not allow us to unambiguously resolve the ancestral area for any of the New Endemic clades; to do so will require additional sampling at and below the species level.

**Combined Data in Phylogeny and Divergence Time Estimation**

Combination of molecular data with heterogeneous rates and patterns of evolution offers the potential to resolve relationships at multiple hierarchical levels, but also poses significant analytical challenges (Bull et al., 1993, Chippindale and Wiens, 1994, Sullivan, 1996; Nylander et al., 2004). In particular, integration of mitochondrial and nuclear genes in vertebrate phylogenetics is challenging, due to their extreme heterogeneity in several evolutionary characteristics (e.g., Springer et al. 1999). In agreement with previous studies, our analyses reflect the disparate evolutionary dynamics of the two genomes, as all but two estimated model parameters differ significantly between the nuclear and mitochondrial genes (Table 4). Despite this heterogeneity, we note that combined analysis of the two data sets under both parsimony and likelihood frameworks increased phylogenetic resolution. Reflecting this heterogeneity, the effect of data combination on support for resolved nodes was method dependent. Overall support was slightly reduced in the parsimony analysis of the combined data as compared to the analysis of IRBP alone, although some individual nodes received markedly increased bootstrap values; in contrast, uniform-model likelihood analysis reflected increased overall support for the combined-data analysis (Table 4). Likewise,
partitioned Bayesian analysis significantly increased support for a number of nodes relative to an unpartitioned analysis (Table 4). Thus, our observations support the notion that mitochondrial and nuclear data sets can provide extremely useful, complementary phylogenetic information, given the analytical tools to address underlying heterogeneity of evolutionary dynamics.

In contrast, our results suggest that combination of mitochondrial and nuclear data to estimate divergence times can severely compromise those estimates. We used relatively ancient calibration points both external and internal to the murid radiation, ranging in age from 6 to 70 million years. These dates are well within the range of sequence saturation for cytochrome \[b\] for the taxa involved (not shown), which causes rates of sequence evolution to be underestimated and divergence times consequently to be overestimated across the tree. This phenomenon is apparent even with the corrections afforded by the invariant sites plus \(\Gamma\)-distributed rates model used to analyze these data. Additionally, both our semiparametric and parametric relaxed molecular clock analyses yielded clear evidence that rate heterogeneity among lineages is much higher for cytochrome \[b\] than for IRBP. Thus, for the taxa and calibrations used here, cytochrome \[b\] appears to present the most severe problem with respect to changes in evolutionary rate, with the least prospect for estimating those changes, due to degradation of signal at high divergence levels. These effects lead, regardless of analytical method used (Table 5), to what we interpret as severe overestimates of divergence times among Philippine rodents. This includes the partitioned Bayesian method of Thorne and Kishino (2002), which integrates separate branch length estimates from multiple data sets (Table 5, combined PB analysis). Although relaxed molecular clock analyses provide powerful tools for estimating divergence times when evolutionary rates are heterogeneous, we caution that some data sets undoubtedly present problems of estimation beyond the ability of current methods to ameliorate. We suggest that these methods will perform best across a range of sequence divergences where saturation is not a problem, or where a more comprehensive set of recent and ancient calibration points is available from the fossil record.

CONCLUSIONS

The pioneering work of Guy Musser and his colleagues (citations in Musser and Carleton, 1993) recognized the importance of dispersal in founding endemic rodent faunas throughout Southeast Asia. With respect to the Philippine murines, Musser and Heaney (1992) identified the “Old Endemic” and “New Endemic” groups and hypothesized at least two waves of immigration into the archipelago. The present study provides the first explicitly phylogenetic test of many of MH’s ideas and shows remarkable congruence with many of their conclusions, despite reliance on different data sources (morphology versus molecules) and different analytical methods (we employ an explicitly phylogenetic approach). Although the molecular data identify five probable dispersal events, our dating analyses suggest that these dispersals were clustered, with colonization of Old Endemic taxa occurring ca. 15 Mya, and the New Endemic clades and *Crunomys* arriving 8 to 12 million years later. Thus, our data are largely consistent with the “two fauna” hypothesis of MH, but our estimated divergence dates are considerably older than many of the dates suggested in that study. In the future, integration of molecular, morphological and other phenotypic data from these rodents will provide important insights into the ecological mechanisms and adaptive processes that have shaped this fauna within the biogeographic framework presented here.

ACKNOWLEDGEMENTS

We thank the curators and collections staff at the Field Museum of Natural History, the National Museum of Natural History, and the University of Michigan Museum of Zoology for their help in arranging and processing loans, particularly Richard Thorton, Michael Carleton, Jeremy Jacobs, and Linda Gordon (USNM); Bill Stanley and John Phelps (FNMNH); and Phil Myers, Steve Hinshaw, and Toni Gorog (UMMZ). Eric Rickart and Toni Gorog collected a number of the specimens used in this study and we thank them for generously providing tissue samples. Guy Musser provided critical guidance and advice in all stages of this project. We thank Scott Steppan, Larry Flynn, Eric Rickart, Brian Barber, and Andy Jones for useful comments on earlier drafts. This research was conducted with financial support from the National Science Foundation, the MacArthur Foundation, the Barbara Brown, Ellen Thorne Smith, and Marshall Field funds of the Field Museum of Natural History, and the University of Minnesota. We thank the Philippine Department of Environment and Natural Resources for granting permits and for their support.

REFERENCES


Ellerman, J. R. 1941. The families and genera of living rodents, Vol. II. British Museum (Natural History), London.

APPENDIX 1

SPECIMEN INFORMATION

Museum numbers of voucher specimens and GenBank accession numbers are listed below. Voucher numbers are listed first, followed by GenBank accession numbers for cytochrome b and IRBP. Except where noted, mitochondrial and nuclear genes were sequenced from the same specimen. Accession numbers without any voucher numbers preceding them indicate sequences that were taken directly from GenBank.


Outgroups for divergence time analysis

---

Museum: Apomys insiguis (FMNH 147083; DQ911497, DQ911492);
Apomys microdus (USNM 458907; DQ911468, DQ911493);
Apomys musculus (USNM 458923; DQ911469, DQ911494);
Archiboscodon luteozonensis (USNM 573834, AY687858, DQ911495);
Bolomys granti (USNM 458914; DQ911470, DQ911496);
Bolomys salomonensis (USNM 458784; DQ911471, DQ911497);
Bolomys sibyraeus (Cypr. 158618; DQ911498;
Bolomys gorona (FMNH 154823; DQ911473, DQ911499);
Bolomys luteozonensis (FMNH 169127, DQ911474, DQ911500);
Carpospilus phaeurus (FMNH 175565, DQ911475, DQ911501);
Ctenomys silaceus (FMNH 169133, AY687859; IRBP: FMNH 175725, DQ911502);
Ctenomys gonzalezii (USNM 458953, AY687860, DQ911503);
Ctenomys haezyii (CMC 776; DQ911476, DQ911505);
Crassomys melanius (FMNH 147105, DQ911477, DQ911506);
Cricetulus bunsoides (FMNH 147942, DQ911478, DQ911507);
Limnonmys “bryophylus” (FMNH 147970; DQ911479, DQ911508);
Limnonmys binaurus (FMNH 147947; DQ911480, DQ911509);
Maxomys whiteheadi (UMMZ 174492, DQ911481; IRBP: USNM 174435, DQ911510);
Nivieerter excelsior (USNM 574372, DQ911511);
Nivieerter rapit (UMMZ 174435, DQ911483, DQ911512);
Phloeomys cuniculus (USNM 537332, DQ911484, AY326103);
Rattus onca (FMNH 143250; DQ911485, DQ911513);
Rattus exulans (USNM 458836; DQ911486, AY326105);

---

---
Rattus preator (USNM 580077; DQ191487, DQ191514); Rattus tanezumi (FMNH 137032; DQ191488, DQ191515); Rhynchomys isareogensis (USNM 573900; DQ191489, AY326108); Sundamys muelleri (Cytb: UMMZ 174436; DQ191490; IRBP: AY326111); Tarsonys apoeensis (FMNH 148178; DQ191491, DQ191516); Hylomyscus allenii (Cytb: AY057817); Hylomyscus denniae (IRBP: AY326088); Praomys petteri (Cytb: AF518368); Praomys delectorum (IRBP: AY326104); Mastomys natalensis (Cytb: AY057818; IRBP: AY326093); Mus musculus (Cytb: AY326096); Aethomys namaquensis (Cytb: AF141215); Aethomys chrysophilus (IRBP: AY326075); Grammomys macmillani (Cytb: AF141218; IRBP: AY326086); Otomys irroratus (Cytb: AF141222); Otomys archiæae (IRBP: AY326101)

Deomyinae: Acomys spiroïdissimus (Cytb: AJ010559; IRBP: AY326074); Lophuromys sikapusi (Cytb: AJ012023) Lophuromys flavopunctatus (IRBP: AY326091)

Gerbillinae: Meriones unguiculatus (Cytb: AF119264; IRBP: AY326095); Tatera kempi (Cytb: AJ012024); Tatera robusta (IRBP: AY326113)

Calomyscinae: Calomyscus bailwardi (Cytb: AF160604; IRBP: AY163581)
Nesomyinae: Nesomys rufus (Cytb: AF160592; IRBP: AY326099)

Cricetinae: Cricetus longicaudatus (Cytb: AB033693; IRBP: AY326082); Phodopus sungorus (Cytb: AF119266; IRBP: AY163631)

Sigmodontinae: Tylomys microtis (Cytb: AF307839; IRBP: AY163643); Sigmodon hispidus (Cytb: AF108702); Sigmodon alstoni (IRBP: AY163640); Neotoma albigula (Cytb: AF108704); Neotoma lepida (IRBP: AY163599); Akodon toba (Cytb: U03527); Akodon azarae (IRBP: AY163578)

Arvicolinae: Microtus californicus (Cytb: AF163593)

Cricetomyinae: Cricetomys enini (Cytb: AF160613; IRBP: AY326081); Steatomys parvus (Cytb: AF160599; IRBP: AY326110)

Mystromyinae: Mystromys albicaudatus (Cytb: AF160607; IRBP: AY163594)

Petromyscinae: Petromysecus collinus (Cytb: AF160601; IRBP: TTU 55218; DQ191517)

Myospalacinae: Myospalax aspalax (Cytb: AF326272; IRBP: AY326097)

Spalacinae: Spalax ehrnegeri (Cytb: AJ389537); Spalax zemni (IRBP: SZU48589)

Rhizomyinae: Rhizomys pruinosus (Cytb: AF326274; IRBP: AY326107); Tachyoryctes splendidus (Cytb: AF160602; IRBP: AY326112)

Phloeomys pallidus (upper left) and Carpomys phaeurus (lower left), which are arboreal, herbivorous species, and Rhynchomys soricoides (upper right) and Chrotomys whiteheadi (lower right), which are predominantly vermivorous ground-living species, represent some of the extensive morphological diversification that has taken place within the two largest endemic clades of Philippine murines. Illustrations by Velizar Simeonovski, copyright Field Museum of Natural History.