DISTRIBUTION OF TARSIER HAPLOTYPES FOR SOME PARTS OF NORTHERN AND CENTRAL SULAWESI

Myron Shekelle[1], Juan Carlos Morales[2], Carsten Niemitz[3], Luluk Lely Soraya Ichwan[1], Don Melnick[2]

[1]Center for Biodiversity and Conservation Studies, Faculty of Mathematics and Natural Sciences University of Indonesia, Depok 16421, Indonesia, Email: shekelle@tarsier.org

[2]Center for Environmental Research and Conservation, Columbia University, New York, NY 10027, USA Email: jcm19@columbia.edu; djm7@columbia.edu

[3]Institute of Human Biology, Free University Berlin 14195 Berlin, Germany

ABSTRACT

DNA sequence data was taken from hair samples of wild Eastern tarsiers that were trapped-and-released. Twelve tarsier populations were sampled at approximately 100 km intervals along a transect that encircles Tomini Bay, running from Sangihe Island in the north, to the Togian Islands in the southeast. This study reports mtDNA haplotype data for 43 Eastern Tarsiers, and compares those with published sequences from Philippine and Western Tarsiers, as well as non-tarsier outgroups. A broad scale analysis, which included 21 unique tarsier haplotypes from 28 individuals, non-tarsier primates, and other mammals, found tarsiers to be a robustly supported monophyletic clade. The Eastern-Western-Philippine tarsier trichotomy was not resolved. Tarsier populations from Sangihe Island (*T. sangirensis*), which lies between Sulawesi and the Philippine island of Mindanao, and from the Togian Islands in Tomini Bay, were robustly supported monophyletic clades. Robust support was also provided for the basal position of *T. sangirensis* with respect to other Eastern tarsiers in this data set. A fine scale analysis, using only tarsiers and which included 26 unique haplotypes from 27 Eastern tarsiers samples (but used only the 3' end of the 12s gene), also found the Sangihe and Togian populations to be robustly supported monophyletic clades. The basal position of *T. sangirensis* had only weak support in the fine scale analysis, however. In both analyses, broad scale and fine scale, other populations of Eastern tarsiers in this data set were generally paraphyletic or polyphyletic. Hypothesis testing found the most-parsimonious tree to be significantly shorter than trees constrained by by the assumption of monophyletic clades in by the assumption of monophyletic clades in regions of macaque endemism, geological microplates that compose Sulawesi, but could not refute the null hypothesis of no difference in overall tree length when constrained by the assumption of monophyletic clades within tarsier acoustic groups.

Keywords: *Tarsius tarsier, T. spectrum, T. bancanus, T. syrichta, T. dianae, T. sangirensis, T. dentatus* molecular phylogeny, 12s, mtDNA

INTRODUCTION

While tarsiers are commonly represented in molecular phylogenetic studies of primates, rarely has more than a single taxon been represented. Dijan and Green (1991) sequenced the involucrin gene for both *T. bancanus* and *T. syrichta*, and Adkins and Honeycutt (1994) sequenced the cytochrome oxidase *c* subunit II mtDNA gene for the same two taxa. Meireles *et al.* (2003) analyzed nuclear DNA sequence for those same two taxa at the globin locus. Never before has DNA sequence data been published for any Eastern Tarsier. The lack of knowledge about DNA sequence variation among tarsiers raises questions about the appropriate analysis of Eastern tarsiers. What is the relationship among Eastern, Western, and Philippine tarsiers? are each of the three species groups monophyletic? What is the most appropriate outgroup for an analysis of Eastern tarsiers?

In order to analyze patterns of haplotype variation among Eastern tarsiers (= Hill's, 1955, *Tarsius spectrum*), it was necessary to conduct a broad scale analysis to address the larger questions about tarsier phylogenetics mentioned above. Since tarsiers are such a deep branch in the primate evolutionary tree, and since the outgroup to tarsiers has not been definitively determined despite the widespread acceptance of a monophyletic Haplorhini (see Morales *et al.* 1999, Yoder 2003), a fairly slowly evolving section of DNA was required. The 12s ribosomal RNA gene of the mtDNA genome has been used for addressing phylogenetic questions regarding Primates and superordinal relationships of mammals, and there is a substantial amount of comparative sequence data available (e.g. Springer and Douzery 1996, McNiff and Allard 1998). The 12s gene also has some hyper-variable regions in the stem-and-loop structure,

notably a long loop near the 3' end of the gene, that are valuable for population level analyses. Secondary structure of the gene product facilitates alignment of this variable-length gene (Springer and Douzery 1996). The 12s gene, therefore, is a practical compromise for this study, because it contains conservative regions with which to address issues of broad scale tarsier phylogeny as well as hyper-variable regions for a population level analysis of Eastern tarsiers.

Shekelle *et al*. (2001) reported results of a preliminary analysis of this data set. Eastern, Western, and Philippine tarsiers were found to be an unresolved trichotomy. Genetic distances among the three tarsier species were comparable to genetic distances among *Hylobates* and two other hominoid genera, *Pan* and *Homo*, measured at the same locus, indicating a relatively old split, conceivably dating to the middle Miocene. *Tarsius sangirensis* was found to be the outgroup of other Eastern tarsiers in the data set reported here: large areas of Sulawesi remain unsurveyed for tarsier genetic diversity, so it has not been verified that *T. sangirensis* is the outgroup of all Eastern tarsiers, nor even that Eastern Tarsiers, as a whole, are monophyletic. Togian tarsiers were an autapomorphic subset with a diagnostic 2 base pair deletion in the hyper-variable loop near the 3' end of the 12s gene.

The taxonomy of Eastern Tarsiers bears on the question of Sulawesi biogeography. There are two broad categories of hypotheses regarding this topic. One such category derives from empirical biological data, notably the Sulawesi macaques. MacKinnon and Mackinnon (1980) offered an implicit hypothesis that a unique taxon of tarsier would co-inhabit the distribution of each of seven macaque taxa, thus creating zones of primate endemism. They further observed that regions that are biogeographically linked to Sulawesi, which possess native tarsier populations, but which lack native macaque populations, represent distinct biogeographic regions where one could expect to find endemic tarsier taxa, e.g. the offshore island groups of Selayar, Banggai, and Sangihe, A similar hypothesis was implied by Niemitz *et al*. (1991).

A second category of biogeographic hypotheses for Sulawesi derives from empirical geological data. Prior to its current form, Sulawesi was an archipelago formed of numerous microplates of Asian, Australian, and oceanic origin. Hall (1996, 2001) reconstructed the geological evolution of Sulawesi by identifying these microplates and charting their movement over the past 50 million years.

Evans *et al*. (2003) used genetic surveys of two distantly related taxa, primates of the genus *Macaca* and toads of the genus *Bufo*, to address hypotheses of Sulawesi biogeography. They found concordant distributions of macaques and toads, which they interpreted to indicate a shared history of range fragmentation. The faunal boundaries in their study showed little correspondence with the microplates identified by Hall (1996).

Shekelle and Leksono (2004) used the distributions of tarsier acoustic forms to address the same topic. Using classic tools of biogeography, they layered the map of Evans *et al*. onto the map of Hall. Then they plotted the distributions of tarsier acoustic forms on the composite map. They found a nearly one-to-one correspondence between the distributions of tarsier acoustic forms and the composite map that combined the biological data and geological data. They reasoned that macaques were relatively recent immigrants to Sulawesi with much of their evolution occurring during the Pleistocene, after tectonic activity had already formed Sulawesi in its modern state. Macaque biogeography was, therefore, likely to have been shaped by Pleistocene range fragmentation/vicariance events. Indeed, many of the faunal boundaries in Evans *et al*.'s study appear to be consistent with geographic boundaries that are influenced by ocean level (e.g. the isthmus of Gorontalo, the Tempe depression). Tarsiers, on the other hand, were an older radiation that probably immigrated to Sulawesi in the Miocene. Tarsier biogeography would, therefore, be influenced by the geologic history of the microplates to a much greater extent than would that of the macaques, which may have colonized Sulawesi after the microplates had already coalesced. The tips of the tarsier branches would, nonetheless, be reshaped by the Pleistocene events that shaped the biogeography of macaques, and thus, tarsier biogeography also has elements of

macaque biogeography. Shekelle and Leksono called this the "hybrid biogeographic hypothesis" for Sulawesi, because it combined empirical data from biology and geology and made explicit the observation that the time of dispersal to Sulawesi was one critical component that would affect biogeography.

METHODS

DNA sequence data were collected from hair samples from 101 wild-caught Eastern tarsiers (Figure 1). Geographic representation included: Sangihe (n = 5), Tangkoko (n = 21), Basaan (=Ratatotok) (n = 5), Molibagu (n = 14), Suwawa (n = 7), Libuo (n = 15), Sejoli (n = 6), Tinombo (n = 8), Marantale (n = 6), Kamarora (n = 7), Malenge (n = 5), Batudaka (=Wakai) (n = 2). This was supplemented by published sequence for *T. bancanus borneanus*, *T. syrichta syrichta*, and other mammals (from gen bank). Incomplete sequence data required many specimens to be excluded from the analysis presented here.

Extraction of total genomic DNA (tDNA) from samples was accomplished using Qiagen extraction kits (#29304 and #29306). Tissue samples from wild-caught tarsiers were from plucked hair. DNA concentration varied, and was adjusted accordingly using an estimate of the gel visualization to produce template DNA of nearly the same concentration.

The target DNA was amplified with the polymerase chain reaction (PCR). A 50 microliter (μL) PCR reaction consisted of the following volumes: 30 μL deionized water (dH_20), 10 μL 5X buffer solution 8.5 pH, 5 μL dNTP, 1 μL primer one (20 pM/μL), 1 μL primer two (20 pM/μL), 1 μL Taq polymerase, 3.5 mM Mg^{++} (1 heat-released bead), 1 μL template DNA. The 12s gene is about 950 base pairs (bp) in length. The following primers were used in various combinations to amplify the 12s gene in 1, 2, or 3 segments: 651f, 891f, 1247f, 933r, 1259r, and 1559r (Table 1). The primer names correspond roughly to the nucleotide number of the human sequence of the primer's 5' end.

The reaction was placed in a thermal cycle machine set to the following cycle parameters: "hot start" = 94° (15 sec.) (first cycle only); 35 cycles of denaturation = 94° (30 sec.), annealing = 58° (30 sec.), and extension = 72° (60 sec.); final extension = 72° (7 min.) (last cycle only); and hold = 4° (infinity). The annealing temperature was sometimes varied to improve amplification with various primer pair combinations. Numerous other permutations of

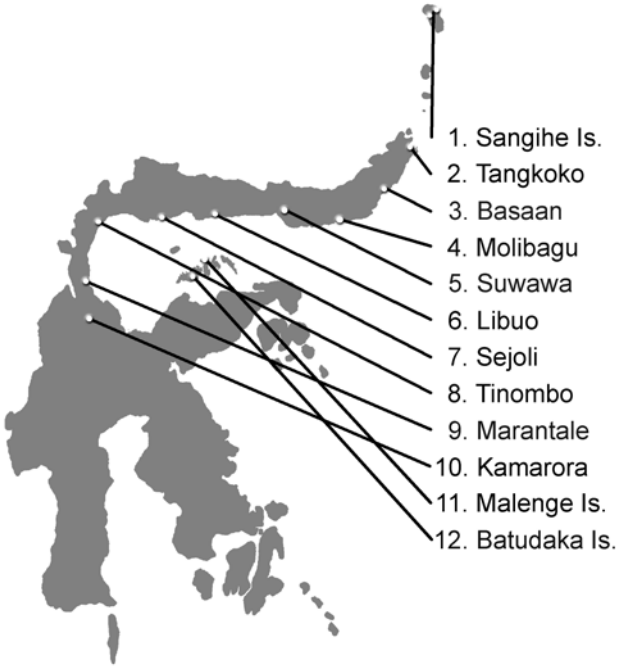

Figure 1: Sampling localities for this study.

Table 1a. PCR and sequencing primers

Primer Name	Length	Sequence
651f (= 378)	18-mer	AGG TTT GGT CCT AGC CTT
891f (tarsier)	21-mer	**A**[1]GG GTT GGT CAA TTT CGT GCC
925r-T.spec	17-mer	GCT TTA CGC CGT GCT TT
930r-T.ban	18-mer	CGC TTT ACG CCG GAT ATT
933r	20-mer	ATC TAA AAC ACT CTT TAC GC
1169r (tarsier)	23-mer	GGG **A**[2]TG TGA AGC ACC GCC AAG TC
1247f (tarsier)	24-mer	CCC GAT **A**[3]AA CCT TAC CAC CCC TTG
1259r	21-mer	GGT TTG CTG **AA**[4]G ATG GCG GTA
1559r (=550r)	24-mer	CCA GTA CAC TTA CCA TGT TAC GAC

[1] CA in *T. syrichta*
[2] G in *T. syrichta*
[3] T,C also present
[4] GG in DUPC 6343

Table 1b. Some primer products and annealing temperatures (from Shekelle 2003)

Primer 1	Primer 2	Length	Optimal Annealing Temperature
651f	925r-T.spec	294 bp	53.5º
"	930r-T.ban	299 bp	53.7º
"	933r	302 bp	51.6º
891f	1259r	409 bp	56.3º
1247f	1559r	359 bp	54.4º
1247f	1169r	16,538 bp	58.3º

thermal cycle parameters were experimented with, but the above setting was deemed to be the best.

Purification of the PCR product was achieved with the Qiagen PCR purification kit (#28106). Products were gel visualized with "DNA quant ladder" to estimate DNA concentration, evaporated on a speed vac, and then resuspended in a quantity of dH_20 sufficient to make the concentration of all of the purified samples approximately equal.

Sequencing of the purified PCR product used the Big Dye kit. A 10 mL sequencing reaction used the following volumes: 4 mL dH_20, 4 mL "Big Dye" mix, 1 mL primer (20 pM/mL concentration), 1 mL purified PCR product. The same primers that were used for PCR were also used for sequencing. The reaction was placed in a thermal cycle machine set to the following cycle parameters: 35 cycles of denaturation = 96° (05 sec.), annealing = 55° (10 sec.), extension = 60° (4 min.).

Sequencing reactions were cleaned of impurities using sephadex. Five grams of sephadex and 80 mL of dH_20 were combined and stirred until thoroughly mixed. An amount equal to 750-800 mL of mixture was aliquoted into spin columns. The columns were spun for 1 min. at 3000 rpm to remove excess moisture. The reactions were added to the top of the spin columns and were spun for 2 min. at 3500 RPM. The purified reactions were captured in 1.5 mL Eppendorf tubes, dried in a speed vac, and resuspended in 3 mL of loading dye (formamide dye mixed with 70 mL of loading solution).

The samples were electrophoresed on a polyacrylamide gel (29:1) and were scored by an ABI 377 PRISM automated DNA sequencer. Both

complementary strands were sequenced in order to double-check the reliability of the sequence data. Raw data were processed and pieced together using Autoassembler (ABI, Perkin Elmer). Alignment was made by eye using comparative data from GENBANK and assumptions about the secondary structure of the 12s ribosomal RNA (Springer and Douzery 1996).

A) Broad Scale Analysis of Tarsier Phylogenetics

A broad scale analysis of tarsier phylogenetics was performed using DNA sequence data from the 12s ribosomal RNA region of the mitochondrial DNA genome. A data matrix was constructed of 900+ b.p. for 36 haplotypes. The matrix included unique haplotypes of 21 tarsiers, 1 strepsirhine, 10 anthropoids, 1 tree shrew, 1 flying lemur, 1 megabat, and 1 microbat.

Eastern tarsiers were represented by 28 individuals. Individuals with identical haplotypes were grouped. Geographic representation was as follows: Sangihe (three haplotypes: ET048, ET049, ET050), Tangkoko (two haplotypes: ET001-002-005-014-082, ET083); Basaan(=Ratatotok) (two haplotypes: ET084, ET085), Molibagu (three haplotypes: ET018-020-023-024-027-029, ET019, ET026), Suwawa (not represented in the broad scale analysis), Libuo (two haplotypes: ET038, ET041), Sejoli (two haplotypes: ET096, ET100), Tinombo (one haplotype: ET074), Marantale (not represented in the broad scale analysis), Kamarora (one haplotype: ET062), Togian (three haplotypes: ET052, ET056, ET057).

Also represented in the data matrix were other sequences taken from the literature: 1 Philippine tarsier (*T. syrichta syrichta*), 1 Western tarsier (*T. bancanus borneanus*), 1 strepsirhine (*Lemur catta*), 10 anthropoids (*Homo sapiens*, *Pan troglodytes*—two individuals, *Pan paniscus*, *Gorilla gorilla*, *Pongo pymaeus*—two individuals, *Hylobates lar*—two individuals, *Papio hamadryas*, and primate outgroups, the flying lemur (*Cynocephalus variegatus*), a tree shrew (*Tupaia glis)*, a megabat (*Donsonia mollucensis)*), and a microbat (*Eptesicus fuscus*).

Forty-five base pairs were trimmed from the 5' end of the gene, and 43 b.p. were trimmed from the 3' end of the gene to accommodate for the PCR primers and areas where missing data predominated. This left a data matrix with 935 characters including gaps. The data set was rooted with the four non-primate taxa (the tree shrew, flying lemur and the two bats) using the "assume outgroup to be paraphyletic" option in PAUP.

A parsimony analysis using PAUP [version 4.0b10 for Macintosh (PPC)] was used to produce strict consensus and bootstrap trees using two separate sequential approximation analyses. A sequential approximation analysis uses successive heuristic analyses, with characters being reweighted based upon the rescaled consistency index after each heuristic search. Heuristic searches, followed by character reweightings, are performed until successive analyses produce identical results. The rationale for this method is to allow the data themselves to adjust the weighting of relatively consistent characters versus relatively inconsistent ones. This method is particularly applicable for the 12s gene, since it includes both highly conserved and highly variable sites and saturation of some sites reduces their utility for older phylogenetic questions (Springer and Douzery 1996).

The first heuristic search used equal weights for each character. Of the 935 total characters, 477 characters were constant and 118 characters were parsimony uninformative. This left 340 parsimony informative characters. The stepwise addition option was used, with 10 random replicates. Other options employed included: gaps are treated as "missing", multistate taxa interpreted as uncertainty, branch-swapping algorithm = tree-bisection-reconnection (TBR), steepest descent option not in effect, initial 'MaxTrees' setting = 100 (will be auto-increased by 100), branches collapsed (creating polytomies) if maximum branch length is zero, 'MulTrees' option in effect, topological constraints not enforced, trees are unrooted. A total of 108 most-parsimonious trees were found, each with a tree length of 1337.

Following the first heuristic search, characters were reweighted based upon the rescaled consistency index using a base score of 100.

Characters were thus assigned a weight between 0 and 100, the latter number indicating a character that was completely consistent across the tree. A second heuristic search was performed using the same settings as the first, the only change being the reweighted characters. Four most parsimonious trees were found, each with a tree length of 50438 (because the reweighted characters use a base score of 100, the second heuristic search had a tree length that was 1 to 2 orders of magnitude greater than the first search).

Following the second heuristic search, characters were again reweighted based upon the rescaled consistency index, and a third heuristic search was performed. Results of the third heuristic search were identical to the results of the second heuristic search and the sequential approximation analysis was deemed to be complete.

B) Fine Scale Analysis of Eastern Tarsier Phylogenetics

A second sequential approximation analysis was performed on a data set that used only the 3' half of the 12s gene (~500 b.p.). Twenty-six unique haplotypes from 27 individuals were included in the analysis. Although the data matrix was comprised of only about half as many characters as in the broad scale analysis, it include the complete hyper-variable loop near the 3' end of the 12s gene, which is more applicable to fine scale analyses. The geographic representation of Eastern tarsiers was as follows: Sangihe (two haplotypes: ET048, ET049), Tangkoko (three haplotypes: ET001, ET003, ET010); Basaan (=Ratatotok) (two haplotypes: ET 084, ET085), Molibagu (three haplotypes: ET018, ET019, ET025), Suwawa (two haplotypes: ET089, ET090), Libuo (two haplotypes: ET 032, ET034), Sejoli (three haplotypes: ET096, ET097, ET100), Tinombo (two haplotypes: ET072, ET077), Marantale (two haplotypes: ET066, ET068), Kamarora (two haplotypes: ET062, ET063), Togian (three haplotypes: ET052, ET054, ET056-058). The Philippine and Western tarsiers were used to root the analysis. The "root as basal polytomy" option in PAUP was employed.

The same PAUP settings were used, as were used in the broad scale analysis. The fine scale analysis had 518 characters, of which, 413 were constant. Fifty-eight characters were parsimony uninformative, leaving 47 informative characters. A total of 131,967 most-parsimonious trees were found, each with a tree length of 150. After reweighting of characters based upon the rescaled consistency index, 222 most-parsimonious trees, each with a tree length of 10931 were found. In the third and fourth heuristic searches, following reweighting of the characters after each search, 221 most-parsimonious trees were found, each with a tree length of 10931.

RESULTS

A) Broad Scale Analysis of Tarsier Phylogenetics

Major elements of the strict consensus tree (Figure 2) included the erroneous placement of the flying lemur within Primates, a monophyletic Prosimii, and a monophyletic Philippine-Western tarsier clade. Within Eastern tarsiers, *T. sangirensis* was a basal outgroup, and Togian tarsiers and Tangkoko tarsiers were autapomorphic subsets. Tinombo and Kamarora tarsiers were represented by a single specimen and necessarily monophyletic. Sejoli, Libuo, and Molibagu tarsiers were each paraphyletic assemblages. Basaan (=Ratatotok) tarsiers were polyphyletic, one haplotype being basal to the Tangkoko clade and the other haplotype nested within the Molibagu clade.

A bootstrap analysis using 1000 replicates was performed using the character weights that resulted from the sequential approximation analysis (Figure 3). All other options were the same as in the heuristic parsimony analyses. Highly supported elements of the bootstrap tree included 100% bootstrap value support for the monophyly of the following clades: *Tarsius*, Eastern tarsiers, *T. sangirensis*, and Togian tarsiers. Other phylogenetic structure that appeared in the strict consensus tree either collapsed or was supported by lower bootstrap values, in the range of 50-83%.

B) Fine Scale Analysis of Eastern Tarsier Phylogenetics

Major elements of the strict consensus tree (Figure 4) included a monophyletic clade of Sangihe tarsiers (*T. sangirensis*) being basal to other Eastern tarsiers in the data set. The Togian tarsier formed a monophyletic clade that was basal to the remaining tarsiers (i.e. non-Sangihe Eastern tarsiers). The remaining tarsiers clustered in a series of clades where, generally, primitive-to-derived haplotypes followed a south-to-north pattern (i.e. the most derived haplotypes are on the extreme northern end of Sulawesi), but where haplotypes from a given locality were not monophyletic. Haplotypes from three localities, Marantale, Sejoli, and Ratatotok were polyphyletic. Haplotypes from Kamarora, Tinombo, Libuo, Suwawa, and Molibagu, and Tangkoko were paraphyletic.

The data set for the fine scale analysis was computationally intensive, probably because there were too few characters relative to taxa, and the bootstrap analysis was stopped after 230 replicates (Figure 5). The bootstrap analysis of the fine scale data set provided robust support for the monophyly of Eastern tarsiers, *T. sangirensis*, the Togian Island tarsier population, and *T. dentatus*, represented by a clade that consisted of two haplotypes found at Kamarora and one haplotype found at Marantale. Each of these clades was supported by bootstrap values between 91-100%. There was one major change in topology, with the bootstrap analysis finding the Togian Island tarsier population to be basal to other Eastern tarsiers in the data set, but this result was supported by a very low bootstrap value of only 54%. Other phylogenetic structure that appeared in the strict consensus tree either collapsed or was supported by lower bootstrap values, in the range of 66-81%.

C) Other Analyses

A final sequential approximation analysis was attempted using *T. sangirensis* to root the remaining Eastern tarsiers in the data set (not figured). This analysis was abandoned when it was found that there were only 31 parsimony informative characters for 26 taxa, many of those characters preferentially located on the branch that defines *T. sangirensis*.

D) Hypothesis Testing

The mtDNA phylogeny produced in the fine-scale analysis was used to directly address questions about tarsier biogeography.

1. Do tarsiers co-inhabit regions of primate endemism with Sulawesi macaques, such as was hypothesized by MacKinnon and MacKinnon (1980) and Niemitz *et al.* (1991), and which would be consistent with the biogeography of macaques and toads found by Evans *et al*. (2003)?
2. Do tarsiers inhabit regions of endemism identified by the microplates of Sulawesi, such as might be predicted by Sulawesi's geologic history as an archipelago? (Hall 2001)
3. Are tarsier acoustic and genetic groups statistically congruent and do tarsier acoustic groups diagnose discrete taxa, such as might be predicted by the mate recognition species concept of Paterson (1985), and MacKinnon and MacKinnon (1980), Niemitz *et al*. (1991), several papers by Nietsch (e.g. Nietsch and Niemitz 1993, Nietsch and Kopp 1998, Nietsch 1999), and therefore consistent with the hybrid biogeographic hypothesis for Sulawesi by Shekelle and Leksono (2004)?

These three hypotheses, abbreviated as (1) "macaque", (2) "microplates", and (3) "acoustic form", were tested by constructing constraint trees in MacClade and loading them into PAUP. In each case the constraint tree assumed a trichotomy among tarsiers, and monophyletic tarsier clades that were arranged in a star phylogeny (Figures 6, 7, 8). Thus, the acoustic form hypothesis test enforced a constraint tree in which each acoustic form had a monophyletic clade of haplotypes, but no other constraints on the topology were enforced, either within or among acoustic forms. A parsimony analysis to produce the most-parsimonious constrained tree was conducted for each hypothesis using the same PAUP settings that were used in the fine scale analysis, the only difference being the topological constraints. The macaque test found 48 most-parsimonious trees each with a tree length of 11,516. The acoustic form test found 34 most-parsimonious trees each with a tree length of 11,330. The microplate

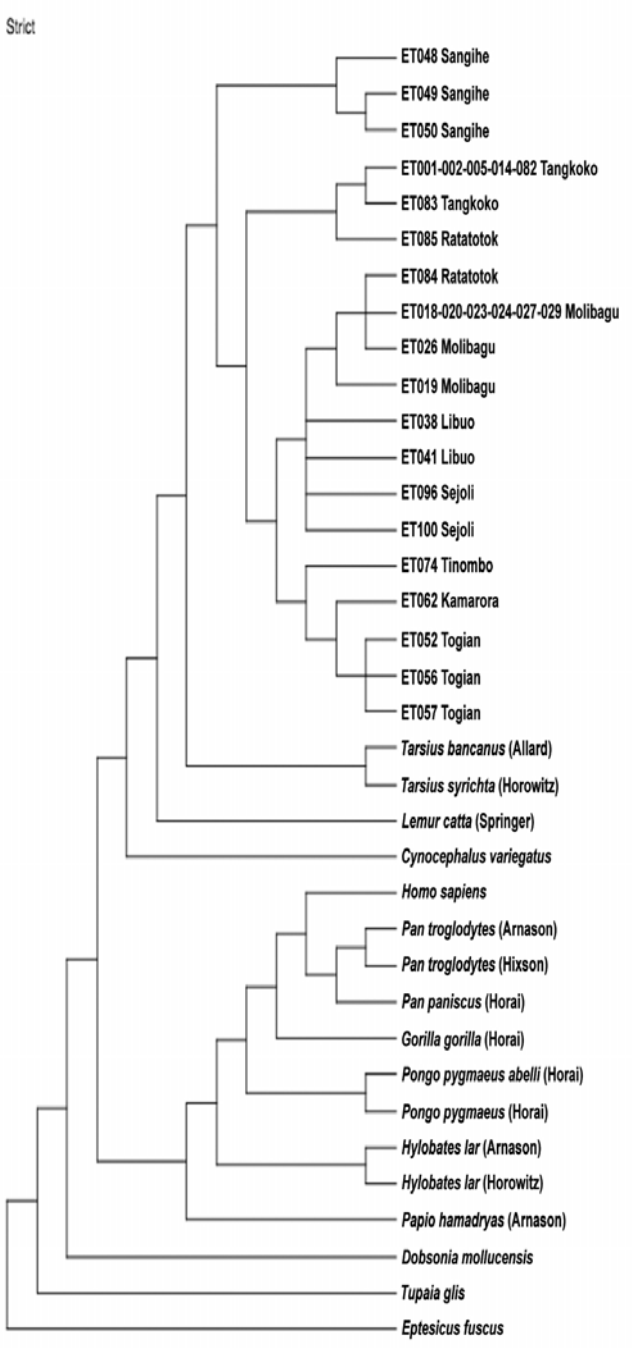

Figure 2. Broad scale phylogenetic analysis—strict consensus tree.

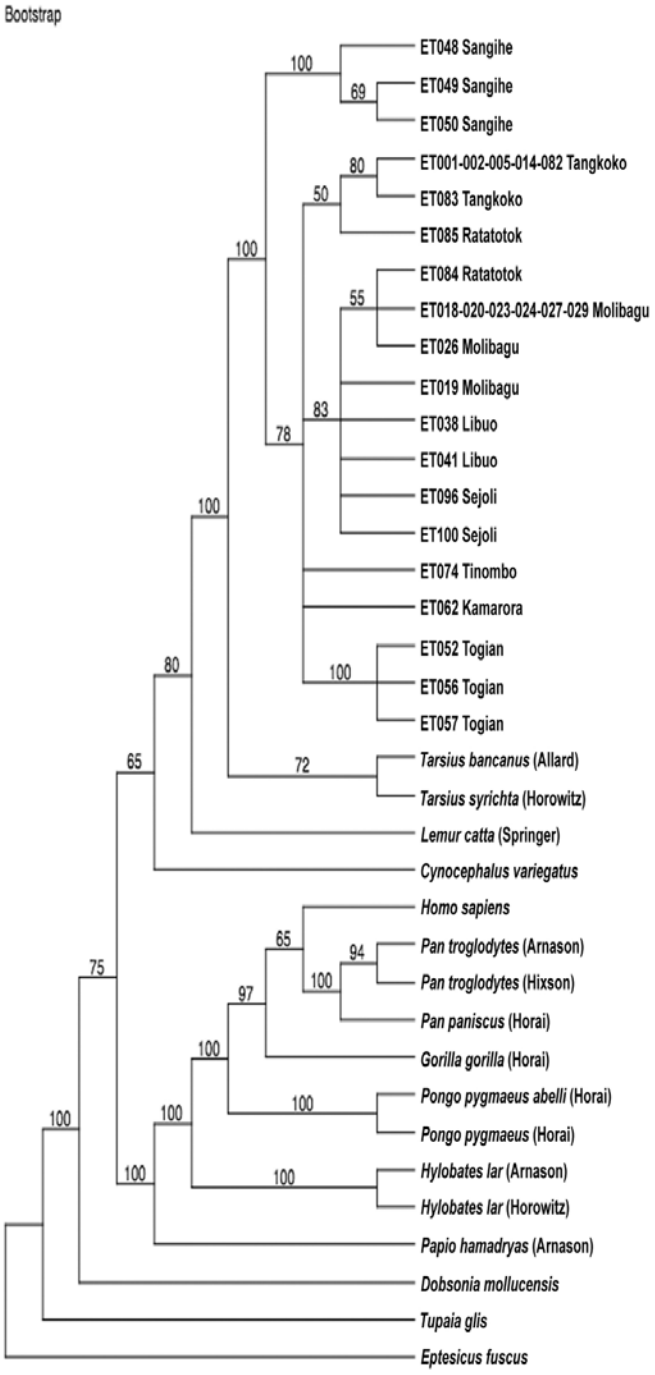

Figure 3. Broad scale analysis—bootstrap tree.

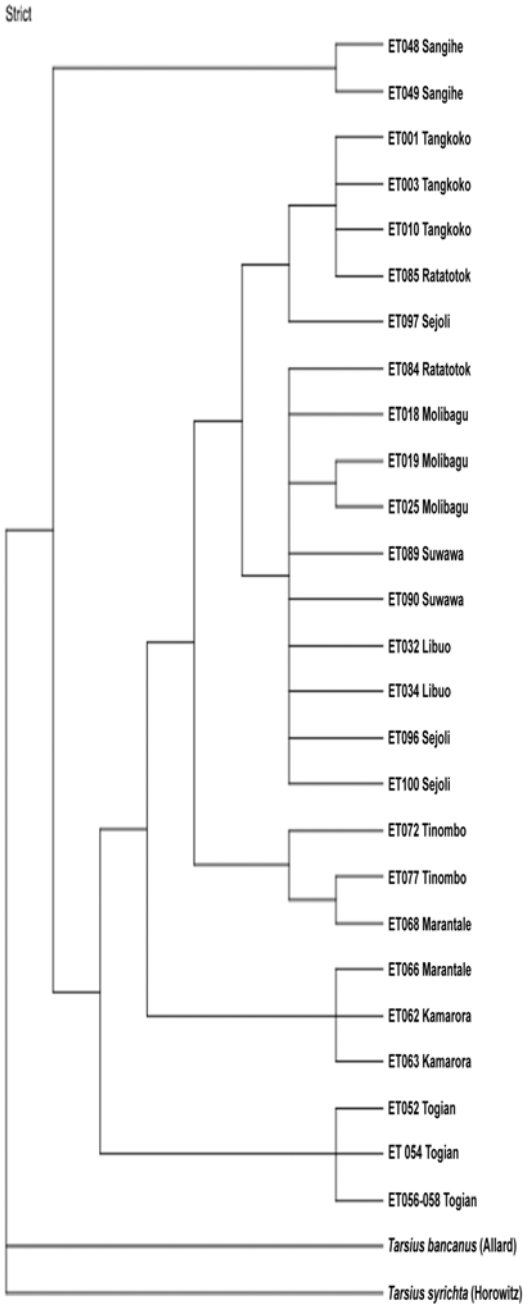

Figure 4. Fine scale analysis—strict consensus tree.

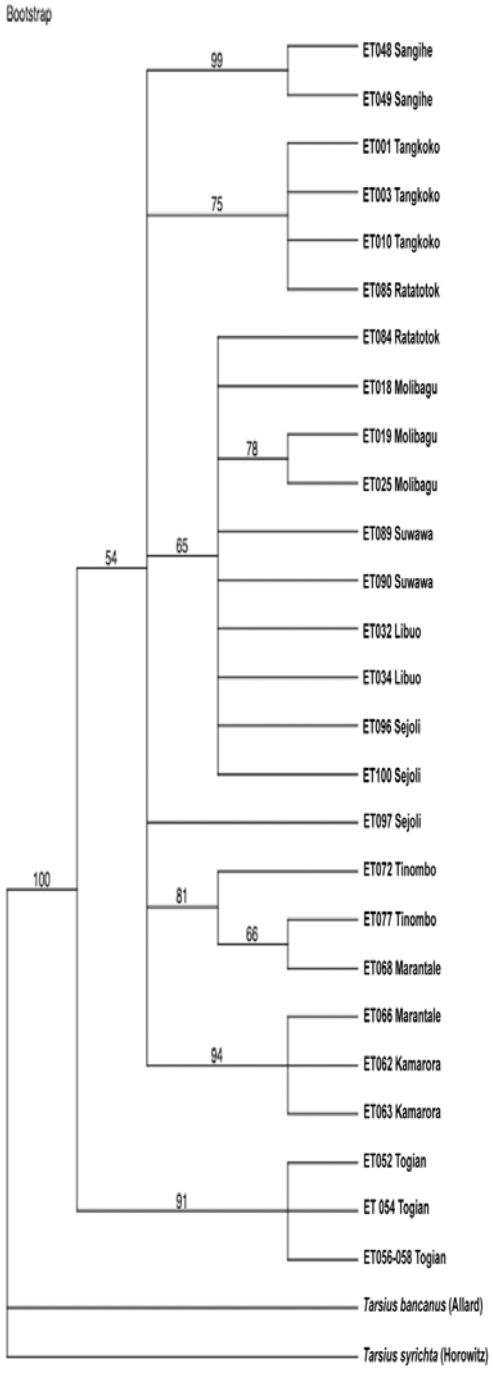

Figure 5. Fine scale analysis—bootstrap tree.

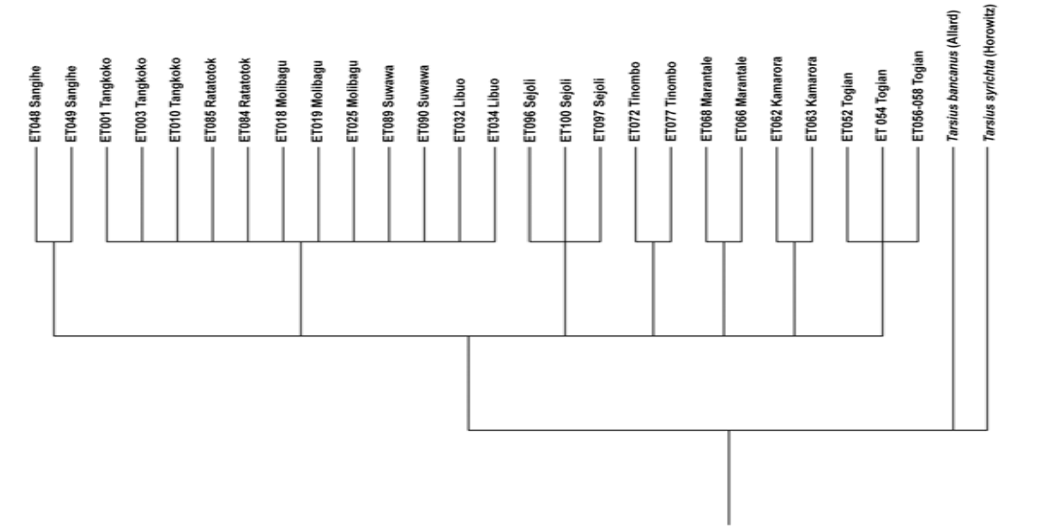

Figure 8. Microplate constraint tree. This tree enforces monophyletic tarsier clades from those localities that lie within each region of endemism identified by the microplates that form Sulawesi.

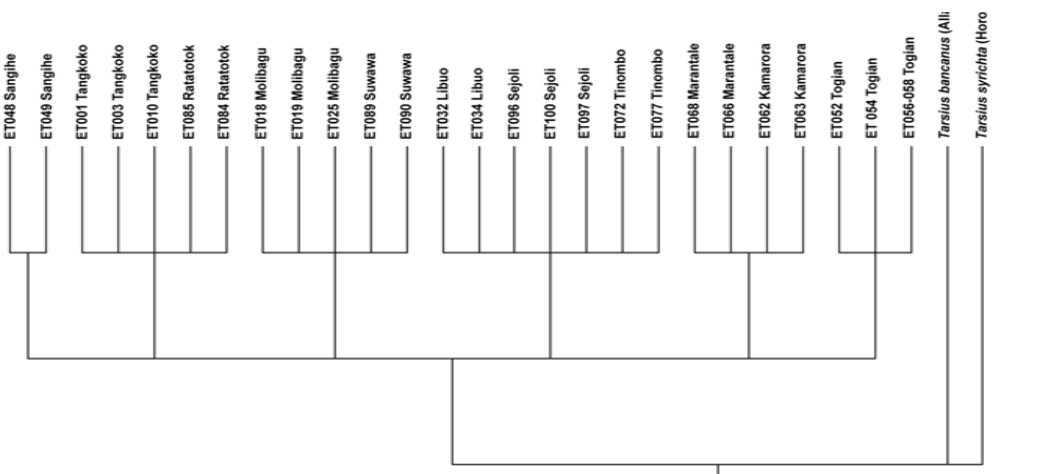

Figure 7. Macaque constraint tree. This tree enforces monophyletic tarsier clades from those localities that lie within each region of endemism identified by macaque distributions on Sulawesi.

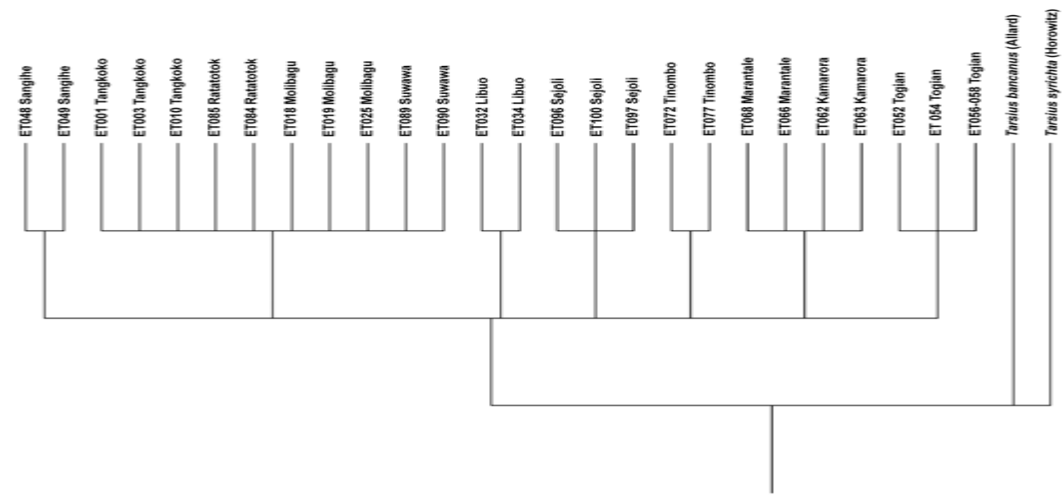

Figure 6. Acoustic constraint tree. This tree enforces monophyletic tarsier clades from those localities that share an acoustic form.

test found 235 most-parsimonious trees each with a tree length of 11,411.

Most-parsimonious constrained trees generated by each hypothesis were tested against the most-parsimonious unconstrained tree with the non-parametric test (=Templeton test) in PAUP, assuming a one-tailed test. In the macaque and microplate analyses, the null hypothesis of no significant difference in tree length between constrained tree and unconstrained tree was very confidently rejected ($P<0.005$). In other words, the unconstrained tree is significantly shorter than the constrained trees generated by the macaque and microplate hypotheses described above. The null hypothesis for the acoustic form hypothesis, however, could not be rejected at the 1% confidence interval (Table 2). There is no significant difference between the length of the most parsimonious tree, and the most parsimonious tree constrained by the hypothesis of monophyletic clades identified by tarsier acoustic form.

DISCUSSION

Phylogenetic Analysis

The most robust portions of the tree topology supported the monophyly of the Eastern tarsiers in this data set, and the monophyly of two isolated island populations, Sangihe and Togian tarsiers. The principal weaknesses were that the broad scale analysis could not confidently resolve the Eastern-Western-Philippine tarsier trichotomy, and the fine scale analysis had poor resolution for haplotypes from within insular Sulawesi. A Western-Philippine tarsier clade was supported by a bootstrap value of 72%, but experience shows that bootstrap values in this range are subject to instability in subsequent analyses.

Given the unexpectedly large degree of genetic variation among tarsiers in general, and Eastern tarsiers, in particular, it was not practical to build a DNA sequence database that was sufficient to definitively address both the broad scale and fine scale taxonomic questions. Indeed, one of the most notable results of this study also highlights a key drawback, that is, neither Philippine nor Western tarsiers are particularly closely related to Eastern tarsiers. Nevertheless, Philippine and Western tarsiers were required to root the fine scale analysis of Eastern tarsiers. In hindsight, the problem was roughly analogous to assembling a data set to answer phylogenetic relationships within Hominoidea from scratch using a lemur as the outgroup, while simultaneously addressing phylogeographic questions within the *Hylobates lar* group using an orangutan as the outgroup.

Current evidence indicates that Western and Philippine Tarsiers are relatively distant cousins of the Eastern Tarsiers that diverged from the latter between 5.6 mya and 17 mya. Meireles *et al*. (2003) estimated the divergence of Philippine and Western Tarsiers at about 5.6 mya using a molecular clock based upon nDNA. It is logical to assume that the divergence of Eastern Tarsiers is at least as old as the Western and Philippine Tarsiers based on, 1) the morphologic data in Musser and Dagosto (1987) and Groves (1998) that supports a Philippine-Western Tarsier clade, and 2) the genetic data in this study that finds an unresolved trichotomy with weak support for a Western-Philippine tarsier clade.

Several lines of evidence suggest that the divergence of Eastern Tarsiers is likely to be at least 9.5 mya. Morley (1998) offered palynological evidence from ocean core samples that showed evidence of biotic exchange across the Makassar Straits at 17, 14, 9.5, 3.5 and 1 mya. The two most recent of those dates, 3.5 and 1 mya, do not appear to be consistent with the tarsier genetic data of Meireles *et al*. (2003) or from this study. Hall (2001) estimated the most likely time for faunal exchange between Asia and Sulawesi as being about 10 mya, not necessarily via the Makassar Straits, but possibly via Java. Preliminary evidence from this study (Shekelle *et al*. 2001) offered a very rough molecular clock estimate of the divergence of Eastern tarsiers at about 13 mya, given that, 1) the average genetic distance between the Eastern tarsiers versus Western and Philippine tarsiers was found to be about 93% as great as the average genetic distance between *Hylobates* versus

Homo and *Pan*, and 2) Goodman *et al*. (1998) estimated the lesser ape-great ape split to be at least 14 mya. Additionally, Mercer and Roth (2003) gave a molecular clock estimate of 11.5 mya ago for the origins of the Sulawesi squirrels. There are, therefore, several independent lines of evidence that indicate that some of Sulawesi's older endemic taxa may have origins dating from 9.5-17 mya. To put this problem in perspective with another analogy, it would be as though Evans *et al*. (2003) had been obliged to root their analysis of Sulawesi macaques with a baboon and a mandrill, or possibly even a langur and a colobus monkey. Future analyses will certainly benefit from using a more appropriate outgroup, such as the Sangihe tarsier or perhaps an as yet unsequenced Eastern tarsier that proves to be basal to the whole radiation (e.g. possibly *T. pumilus*), but this will have to be determined by further experimentation.

Two other notable shortcomings of this study are that, 1) this data set analyzes DNA sequence from only a single gene and it is well known that different genes have different tree topologies, and 2) the gene in question is in the mtDNA genome, a genome that is known to be affected by the dispersal pattern of the organism in question. Matrilocal taxa and short-distance dispersers show more phylogeographic structure in their mtDNA gene tree than do patrilocal taxa and long-distance dispersers (Melnick and Hoelzer 1992, 1993). There are no data for dispersal among tarsiers (see Sussman 1999), so we cannot predict how dispersal patterns will affect these results.

Hypothesis Tests and Sulawesi Biogeography

The hypothesis for Sulawesi biogeography that derives from empirical biological data, such as Evans *et al*. (2003), and the one based upon empirical geologic data (e.g. Hall 2001), have numerous areas of incongruence. Given the quality of the evidence that supports each of the two hypotheses above, this seems puzzling at first glance, i.e. why do the empirical biological data not fit the model based upon the geological history of the island? Adding further to the mystery, the tarsier genetic data in this study seem incompatible with either hypothesis, as results of this study found that constrained trees based upon those two hypotheses were very significantly longer than the unconstrained most-parsimonious tree.

The distribution of tarsier acoustic forms appears to offer a novel solution to the issue of Sulawesi biogeography (Shekelle 2003, Shekelle and Leksono 2004). The distributions of tarsier acoustic forms share some boundaries with macaques, but also share some similarities with Sulawesi's microplates. For instance, tarsiers and macaques share faunal boundaries at the isthmus of Gorontalo and the isthmus of Palu. But between the isthmus of Gorontalo and the isthmus of Palu, in the region of *Macaca hecki*, there are three tarsier acoustic forms—an area that not coincidentally has three microplates (Figure 9). Given the *a priori* predictions that acoustic forms are distinct taxa, it is remarkable that a unique acoustic form is present in almost every biogeographic region predicted by the hybrid biogeographic hypothesis.

Table 2: Results of hypothesis tests. The null hypothesis of no significant difference between tree lengths was rejected in all three tests using the Templeton test. Two of the results, the macaque and microplate tests, were very highly significant. The acoustic test was not rejected at the 1% confidence interval.

tree	tree length	N	z	t	critical value 0.01	p
most parsimonious	10,931	-	-	-	-	-
acoustic	11,330	10	1.8869	9	5	>0.01
macaque	11,516	12	2.7477	4	10	<<0.005
microplate	11,411	9	2.6656	0	3	<<0.005

In retrospect, the implicit hypothesis that tarsiers might share regions of endemism with macaques, e.g. MacKinnon and MacKinnon (1980), Niemitz *et al*. (1991) was influenced by the somewhat limited understanding of Sulawesi's geologic history that existed prior to the geological reconstructions of Hall in the mid 1990's. Also, the implicit assumption that the Eastern tarsier radiation took place at approximately at the same age as the Sulawesi macaque radiation had not been examined. This assumption, however, is almost certainly false. It can be inferred that the Sulawesi macaque radiation is likely to have occurred mostly or entirely during the Pleistocene and Holocene (Delson 1980, Goodman *et al*. 1998, Evans *et al*. 1999). Thus, much or all of macaque evolution on Sulawesi occurred after the coalescence of the microplates into the modern Sulawesi (Hall 2001) and vicariance/range fragmentation is expected to greatly outweigh geological history as the primary biogeographic factor affecting differentiation. Indeed, macaque distributions appear to have faunal boundaries that are associated with Pleistocene geographic barriers, such as the isthmus of Gorontalo and the Tempe depression, neither of which are associated with a microplate boundary.

Sulawesi tarsiers, on the other hand, are a much older radiation than Sulawesi macaques with roots in the Miocene, and their arrival to Sulawesi almost certainly predates the formation of Sulawesi in its present form. It is logical to predict, therefore, that Eastern tarsiers have a pattern of distribution that was first shaped by colonization of the proto-Sulawesi archipelago during the Miocene and Pliocene, which could have included sweepstakes dispersal and ancient vicariance events. Subsequent to the coalescence of the microplates into Sulawesi, tarsier distributions were then reshaped by Pleistocene vicariance events, with the tips of the tarsier branches bearing the effects of the forces that shaped the distributions of macaques and toads. This observation is the core of the hybrid biogeographic

Figure 9. (from Shekelle and Leksono 2004).

Left: a biogeographic map of Sulawesi and surrounding islands based upon empirical data from the distribution of genetic variability in *Macaca* and *Bufo* (Evans et al. 2003), plus regions that lack endemic macaques and are presumably biogeographically distinct (MacKinnon and MacKinnon 1980).

Center: a biogeographic map of the same area based upon the geological reconstructions of Hall (2001) concerning the tectonic activity of the microplates of the proto-Sulawesi archipelago.

Right: a composite map of the distribution of tarsier acoustic forms layered on top of the *right* and *center* maps.

hypothesis for Sulawesi, which synthesizes empirical data from biology and geology, and explicitly acknowledges a time component that is a critical factor shaping biogeography in the region.

Results from the hypothesis tests presented here show that the genetic data in this study cannot refute the hybrid biogeographic hypothesis and, indeed, the phylogenetic analyses are broadly consistent with it. The deepest evolutionary splits predicted by the hybrid biogeographic hypothesis receive strong support from the tarsier genetic data, while the more recent evolutionary splits are consistent with the genetic data and traces of phylogeographic structure are apparent in the consensus tree. The regions identified by MacKinnon and MacKinnon (1980) that have tarsiers, but which lack native macaque populations (i.e. the island chains of Banggai, Sangihe, Selayar, and Togian), can be logically argued to be more ancient than those areas that possess both taxa. Two such regions appear in this study, Sangihe and Togian, and in both cases, the tarsiers are identified as robustly supported monophyletic clades that are distantly related to other tarsiers of Sulawesi. The relatively recent evolutionary events, i.e. those within insular Sulawesi, are not perfectly supported by congruence of tarsier acoustic forms and monophyletic genetic clades, but that is not necessarily what would be predicted. Even so, it is interesting to speculate about phylogeographic structure in the consensus tree from the fine scale analysis. The Marantale population (*T. dentatus*), for instance, lies within the known hybridization zone between *M. tonkeana* and *M. hecki* (Bynum *et al.* 1997). The gene tree shows Marantale tarsier haplotypes to be polyphyletic, one haplotype clusters with Kamarora to the south, while the other clusters with Tinombo to the north, which are themselves nested within a northern clade. This pattern might possibly indicate that *T. dentatus* males occasionally hybridize with females from the Tinombo acoustic form. The two other populations that are polyphyletic in this tree, Sejoli and Ratatotok, also sit close to faunal boundaries identified in the hybrid biogeographic hypothesis.

In a study that was very similar to this one, Shaw (1993) examined biogeography and taxonomy in the Hawaiian cricket genus *Laupala* using acoustics and mtDNA. Shaw (2002) revisited conclusions in her 1993 study in light of nDNA and found that the mtDNA phylogenies had provided "extensively misleading" results, probably because of interspecific hybridization. Shaw (2002) cautioned against basing evolutionary interpretations among closely related species groups on mtDNA phylogenies, and found that nDNA provided results that were more consistent with other factors, such as acoustics, biogeography, and morphology. Indeed, for some time it has been known that lineage sorting and hybridization can produce data sets wherein taxa are not defined by monophyletic groups (e.g. Melnick and Hoelzer 1992, 1993).

Several lines of evidence are broadly consistent with the hybrid biogeographic hypothesis including: the distributions of macaques and toads, the microplates that form the island of Sulawesi, the self-evident observation that the time of dispersal to Sulawesi is critical for biogeography, the tarsier acoustic data, and the tarsier genetic data. Some puzzling issues remain, however. For one, the arrival of macaques on Sulawesi is certainly recent compared to tarsiers, but there is less evidence that *Bufo* shares a similarly recent arrival to the region. *Bufo* is an ancient genus, and we can speculate that if, perhaps, *Bufo* arrived long before *Macaca*, then why should those two taxa share congruent distributions? Another puzzle is why faunal boundaries should remain congruent with microplate boundaries after millions of years. Is there, perhaps, some relationship between the underlying bedrock of the microplates and the ecology of the flora and fauna on the surface? Additionally it is worth mentioning that the pattern of dispersal and range fragmentation will vary among taxa, so the 15 biogeographic subregions predicted by tarsier acoustic forms should be considered a minimum estimate for Sulawesi. With additional taxa, the overall picture is likely to be more complex with more subregions.

C) Summary

Two phylogenetic analyses (i.e. broad scale analysis and fine scale analysis) used sequential approximation to wring as much phylogenetic information as possible from the available DNA sequence data. As expected, the first analysis offered decisive support for the monophyly of tarsiers. It also offered robust support for the monophyly of Eastern tarsiers in the data set. It could not convincingly resolve the Eastern-Western-Philippine tarsier trichotomy, however. This question may be resolvable, but will always suffer somewhat from lack of a suitable outgroup. Results of the broad scale analysis indicated that all non-tarsier taxa could be pruned from the data matrix for the fine scale analysis, but that both Western and Philippine tarsiers were required to root the analysis. Eastern tarsiers are not closely related to Western or Philippine tarsiers, and future fine scale analyses will benefit by identifying a more suitable outgroup from within the Eastern tarsiers. The fine scale analysis confirmed the monophyly of *T. sangirensis* and offered support for its basal position among the Eastern tarsiers in this data set. Subsequent pruning of the Philippine and Western tarsier and rooting the remaining data set with the *T. sangirensis*, however, left a data set with too few informative characters to be worthwhile. From this it was concluded that no further phylogenetic information could be wrung from the DNA sequence data, and conclusions about evolution among Eastern tarsiers were based on the analysis where Eastern tarsiers were rooted with Philippine and Western tarsiers.

The most parsimonious tree from the fine scale analysis was used to reject biogeographic hypotheses based on the distributions of macaques and toads as well as the microplates that form the geological history of Sulawesi. A third hypothesis of congruence between tarsier acoustic groups and genetic groups could not be rejected. Differences between macaque distributions and tarsier distributions are not unexpected given what we now know about the relative ages of these two radiations. The hybrid biogeographic hypothesis for Sulawesi (Shekelle and Leksono 2004), a comprehensive hypothesis which combines empirical biological and geologic data and explicitly considers the time of immigration to Sulawesi, was examined in light of tarsier acoustic data, which shows a remarkable fit with the aforementioned hypothesis. Tarsier genetic data in this study support predictions of that hypothesis. The implication is that each of the 15 tarsier acoustic forms thus surveyed is a distinct taxon, the validity of which can be examined more rigorously with additional DNA sequence data. More field surveys will very likely result in the discovery of more acoustic forms.

ACKNOWLEDGEMENTS

This material is based on work supported by the National Science Foundation under Grant No. INT 0107277 to MS. Additional financial support was provided to MS by the National Science Foundation Predoctoral Fellowship, Primate Conservation, Inc., Washington University Department of Anthropology, L.S.B. Leakey Foundation, Explorer's Club, Wenner Gren Foundation, and the Margot Marsh Biodiversity Fund. MS wishes to acknowledge the support Jatna Supriatna and Noviar Andayani for sponsoring his work in Indonesia. Permits and other logistical support were supplied by The Indonesian Institute for Sciences (LIPI), The Department of Forestry, and Siti Prijono of the *Museum Zoologicum Bogoriense*. Preliminary genetic data that was not used in this study was collected in the labs of Allan Larson (Washington University), LIPI Pusat Bioteknologi (Center for Biotechnology, Indonesian Institute of Science, Cibinong), and Rajawali Hospital (Bandung). The genetics lab work was assisted by Tom Titus, Susan Jacobs, and Todd Jackman, and Luluk Lely Soraya Ichwan. Many people offered advice, comments, and/or assistance on this project including Tab Rasmussen, Allan Larson, Bob Sussman, Jim Cheverud, Jane Phillips-Conroy, Erik Trinkaus, Mitchell Sommers, and Alan Templeton. Portions of this appeared in the dissertation of MS.

REFERENCES

Adkins RM & RL. Honeycutt. 1994. Evolution of the primate cytochrome c oxidase subunit II gene. *J. Mol. Evol.* 38:215-231.

Bynum, EL, DZ. Bynum & J. Supriatna. 1997. Confirmation and location of the hybrid zone between wild populations of *Macaca tonkeana* and *Macaca hecki* in Central Sulawesi, Indonesia. *Amer. J. Primatol.* 43:181-209.

Delson, E. 1980. Fossil macaques, phyletic relationships and a scenario of deployment. In *The Macaques: Studies in Ecology, Behavior and Evolution*, Lindburg, DG. (ed) pp:10-30. New York: Van Nostrand Reinhold Company.

Dijan, P & H, Green. 1991. Involucrin gene of tarsioids and other primates: alternatives in evolution of the segment of repeats. *Proc. Natl. Acad. Sci. USA* 88:5321-5325.

Evans, BJ, JC. Morales, J. Supriatna & DJ. Melnick. 1999. Origin of the Sulawesi macaques (Cercopithecidae, *Macaca*) as inferred from a mitochondrial DNA phylogeny. *Biol. J. Linn. Soc.* 66:539-560.

Evans, BJ, J. Supriatna, N. Andayani & DJ. Melnick. 2003. Monkeys and toads define areas of endemism on Sulawesi. *Evol.* 57(6):1436-1443.

Goodman M, CA. Porter, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, & Groves CP, 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol. Phylo. Evol.* 9:585-598.

Groves, CP. 1998. Systematics of tarsiers and lorises. *Primates*, 39:13-27.

Hall, R. 1996. Reconstructing Cenozoic SE Asia. In. Tectonic Evolution of Southeast Asia. Hall R, Blundell D. (eds) Geol. Soc. Lond. Special Publication 106:153-184.

Hall, R. 2001. Cenozoic reconstructions of SE Asia and the SW Pacific: changing patterns of land and sea. In *Faunal and Floral Migrations and Evololution. SE Asia-Australia*, Metcalf I, Smith J, Morwood M, Davidson I. (eds) pp:35-56. Lisse: Swets and Zeitlinger Publishers.

Hill, WCO. 1955. *Primates: Comparative Anatomy and Taxonomy. II. Haplorhini: Tarsioidea.* Edinburgh: Edinburgh University Press.

MacKinnon, J & K. MacKinnon. 1980. The behavior of wild spectral tarsiers. *Int.J. Primat.* 1(4):361-379.

McNiff, BE & MW. Allard. 1998. A test of Archonta monophyly and the phylogenetic utility of the mitochondrial gene 12s rRNA. *Am. J. Phys. Anthropol.* 107:225-241.

Melnick, DJ & GA. Hoelzer. 1992. Differences in male and female macaque dispersal lead to contrasting distributions of nuclear and mitochondrial DNA variation. *Int. J. Primatol.* 13(4):379-393.

Melnick, DJ & GA. Hoelzer. 1993. What is mtDNA good for in the study of primate evolution? *Evol. Anthro* 2(1):2-10.

Meireles, CM, J. Czelusniak, SL. Page, DE. Wildman & M. Goodman. 2003. Phylogenetic position of tarsiers within the order Primates: evidence from g–globin DNA sequences. In *Tarsiers: Past, Present, and Future*. Wright PC, Simons EL, Gursky S. (eds) pp:145-160. New Brunswick: Rutgers UP.

Mercer, JM, & VL. Roth 2003. The effects of Cenozoic global change on squirrel phylogeny. *Science* 299:1568-1572.

Morales, JC, TR. Disotell & DJ. Melnick. 1999. Molecular phylogenetic studies of nunhuman primates. In *The Nonhuman Primates*. Dolhinow P, Fuentes A. (eds) pp:18-28. Mountain View CA: Mayfield Publishing Company.

Morley, RJ. 1998. Palynological evidence for tertiary plant dispersals in the SE Asian region in relation to plate tectonics and climate. In *Biogeography and Geological Evolution of SE Asia*. Hall R, Holloway JD (eds) pp:211-234. Leiden: Backhuys.

Musser, GG & M. Dagosto. 1987 The identity of *Tarsius pumilus*, a pygmy species endemic

to the montane mossy forests of Central Sulawesi. *Amer. Mus. Novit.* 2867:1-53.

Niemitz, C, A. Nietsch , S. Warter & Y. Rumpler. 1991 *Tarsius dianae*: A new primate species from Central Sulawesi(Indonesia). *Fol. Primatol.* 56:105-116.

Nietsch, A. 1999. Duet vocalizations among different populations of Sulawesi tarsiers. *Int. J. Primatol.* 20(4):567-583.

Nietsch, A & C. Niemitz, C. 1993. Diversity of Sulawesi tarsiers. *Deutsche Gesellschaft für Säugetierkunde* 67:45-46.

Nietsch, A & ML. Kopp . 1998. Role of vocalization in species differentiation of Sulawesi tarsiers. *Fol. Primatol.* 68(suppl.1):371-378.

Paterson, HEH. 1985. The recognition concept of species. In *Species and Speciation*. Vrba ES. (ed) pp:21-29. Monograph No. 4. Pretoria: Transvaal Museum.

Shaw, KL. 1993. The Evolution of Song Groups in the Hawaiian Cricket Genus *Laupala*. Doctoral Dissertation, Washington University, 1993.

Shaw, KL. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *PNAS* 99(25):16122-16127.

Shekelle, M. 2003. Taxonomy and Biogeography of Eastern Tarsiers. Doctoral thesis. Washington University, St. Louis.

Shekelle, M, JC. Morales & DM. Melnick. 2001. Genetic and acoustic evolution among Eastern Tarsiers of Northern and Central Sulawesi. Presented at the International Society of Primatologists, 14th Congress, Adelaide, Australia. January 7-12, 2001.

Shekelle, M & SM. Leksono. 2004. "*Rencana Konservasi di Pulau Sulawesi: Dengan Menggunakan* Tarsius *Sebagai* 'Flagship Taxon'. *Biota* IX(1):1-10.

Springer, MS & E. Douzery. 1996. Secondary structure and patterns of evolution among mammalian mitochondrial 12s rRNA molecules. *J. Mol. Biol.* 43:357-373.

Sussman, RW. 1999. *Primate Ecology and Social Structure, Volume 1: Lorises, Lemurs, and Tarsiers*. Needham Heights (MA): Pearson Custom Publishing.

Yoder, A. 2003. The phylogenetic position of genus *Tarsius*: whose side are you on? In *Tarsiers: Past, Present, Future*. Wright P, Simons E, Gursky S (eds) pp:161-175. New Brunswick, New Jersey: Rutgers UP.