PHYLOGENETIC ANALYSES OF PAINTED SPINY POCKET MICE (*HETEROMYS PICTUS*) AND JALISCAN SPINY POCKET MICE (*HETEROMYS SPECTABILIS*)

Joanna R. Bateman[1], Duke S. Rogers[2] and Victoria L. Vance
[1]Department of Biology, Texas Tech University, [2]College of Life Sciences, Brigham Young University
(Joanna.R.Bateman@ttu.edu)

Background

Heteromys pictus and *H. spectabilis* comprise a species complex of spiny pocket mice distributed across the western and southern Mexican coast (Fig. 1). However, *H. spectabilis*, an endangered species restricted to SE Jalisco, forms a monophyletic group within *H. pictus*, causing paraphyly. The goal of this research is to determine how many species-level taxa are present. So far we have constructed phylogenetic trees based on individual genes, as well as a combined gene tree. Future research will focus on reduced genome sequencing (exome sequencing) and incorporating the geographic relationships between populations.

Methods

Individuals were organized into *Cytb* "haplogroups" (N#), which are based on "haplotype network" groupings defined by Victoria Vance in 2006. These haplogroups consistently form distinct monophyletic (*Cytb*) clades despite geographic overlap of habitat, which indicates that they are distinct non-crossing lineages.

Phylogenetic trees were constructed using Maximum Likelihood software (RAxML) and Bayesian software (BEAST). Gene trees for each gene (*Cytb* [Fig. 2], Beta-Fibrinogen Intron 7 [Bfib], Protein Kinase C Iota [PRKCI], and Interphotoreceptor retinol-binding protein 3 [IRBP]) were constructed in RAxML and BEAST, using ~193 ingroup specimens, depending on sequencing success. An additional Combined gene tree (Fig. 3) was constructed in BEAST using the previously mentioned genes and 346 ingroup individuals.

Kimura 2-Parameter (K2P) values using *Cytb* sequences were calculated (in MEGA) to predict divergence levels within the species complex (Table 1). Individuals from 1 outgroup species (*H. irroratus*) were also used to provide a comparison baseline for values found within the species complex. For reference, K2P values greater than 7.3% (on average) indicate separate **species** within Rodentia, and values greater than 13% (on average) indicate different **genera**.

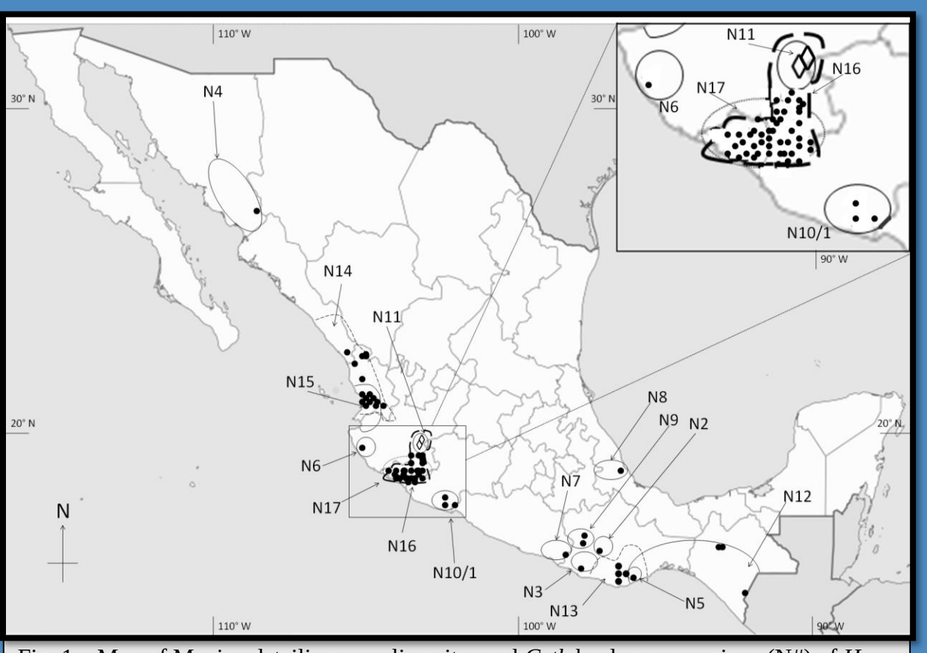


Fig. 1—Map of Mexico detailing sampling sites and *Cytb* haplogroup regions (N#) of *H. pictus-spectabilis.* Black dots (●) represent *H. pictus* sampling sites, with diamond marks (◊, Haplogroup 11) for *H. spectabilis* sampling sites.

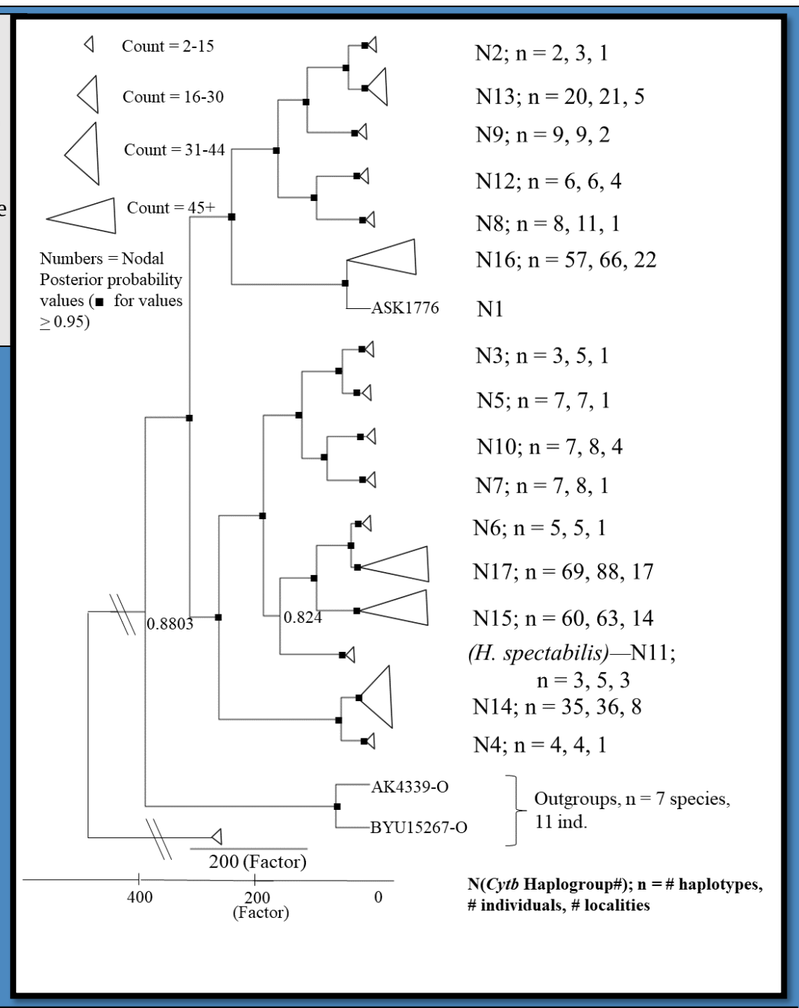
Results

The phylogenetic trees constructed via BEAST (Fig. 2-3) consistently had stronger nodal support than the RAxML trees, indicating more robust hypotheses for relationships within the species complex. Thus the BEAST trees were retained for further analysis. Out of the individual genes, *Cytb* had the most well-defined topology and strongest nodal support.

Within the species complex, the *Cytb* haplogroups (N1-N17) predicted two main clades in the *Cytb* tree (Fig. 2), and three main clades in the Combined gene tree (Fig. 3).

K2P analyses for *H. pictus* and *H. spectabilis* (Table 1) were successfully run between individuals and haplogroups. The K2P range within *H. pictus* overall was 0 to 17.7 (%), while the range within *H. spectabilis* was 0.1-3.1 (%). Comparing *H. pictus* and *H. spectabilis* individuals, K2P values ranged between 9.9-17.4 (%), and values between the species complex and its sister species *H. irroratus* ranged from 14.4-19.2 (%).

Fig. 2—Bayesian phylogenetic hypothesis for H. pictus and H. spectabilis based on *Cytb* sequence data in BEAST. Numbers to the right of terminals (i.e. N2, etc) are haplogroup designations based on haplotype networks defined by Victoria Vance.



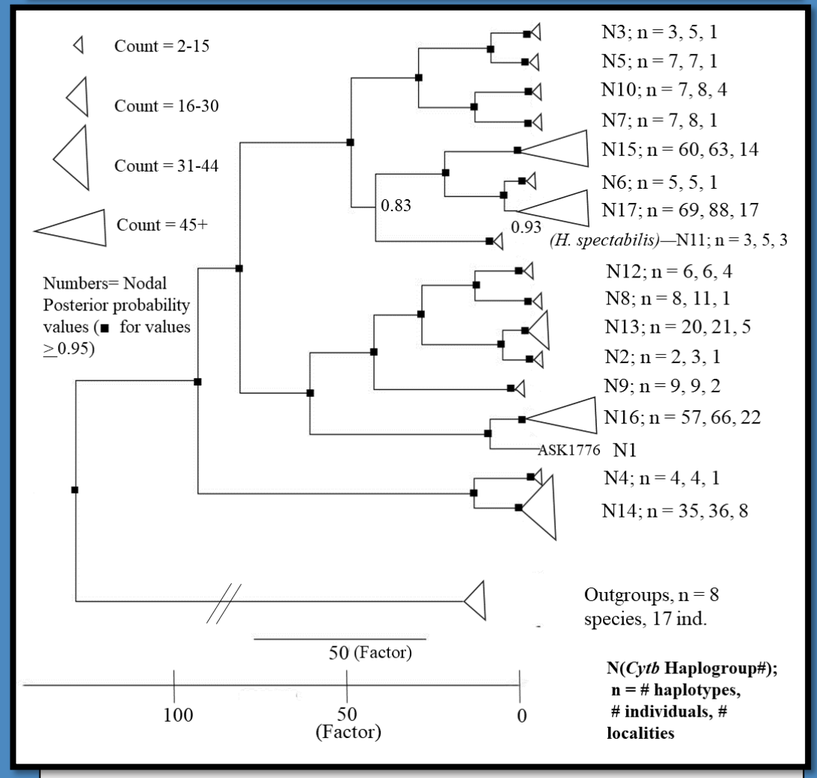


Fig. 3—Bayesian phylogenetic hypothesis for H. pictus and H. spectabilis based on Combined (Bfib, *Cytb*, IRBP, and PRKCI) sequence data in BEAST. Numbers to the right of terminals (i.e. N2, etc) are haplogroup designations based on haplotype networks defined by Victoria Vance.

Table 1—Kimura 2-Parameter (K2P) range values (%) both within selected *Heteromys* species and between species (based on *Cytb* sequences).

Species	K2P within-species values (%)	*H. irroratus* (outgroup)	*H. spectabilis*
***H. irroratus* (outgroup)**	4.5		
H. spectabilis	0.1-3.1	17.0-17.4	
H. pictus	0-17.7	14.4-19.2	9.9-17.4

Discussion

Regarding the phylogenetic trees, there is strong nodal support for the *Cytb* haplogroups being monophyletic lineages, and based on the *Cytb* and Combined gene trees, there appear to be 2-3 distinct clades within the species complex. However, the exact definition of these clades requires more resolution, as some haplogroups like N4/N14 and N9 do change position between the trees. As for the overall clade composition, the 1st clade appears to comprise the haplogroups along the western Mexican coast (N3, N5, N10, N7, N15, N6, N17, and N11), while the 2nd clade mainly consists of the haplogroups along the southern coast and going into Guatemala (N12, N8, N2, N13, N9, N16, and N1), though there is some geographic overlap between the clades along the southern coast. N4 and N14 either form their own clade or are part of the 1st clade.

Meanwhile, the K2P values strongly predict multiple cryptic species, and potentially genera, are present in the species complex. The K2P values across the overall *H. pictus* group (all haplogroups except N11) varied wildly, from 0% (identical) all the way up to 17.7%, which are values generally expected between sister *genera* in rodents. In comparison, *H. spectabilis* and sister species *H. irroratus* had within-species K2P ranges that didn't exceed 4.5%. This predicts that *H. spectabilis* should be retained as a species, but *H. pictus* has lineages that are significantly diverged from each other and this species should be split into multiple species.

Future Work

Moving forward, I am expanding this research as part of my PhD Dissertation at Texas Tech University, and I am planning to use reduced genome sequencing to more deeply evaluate the species complex. This will primarily be accomplished through exome sequencing, which targets the genome regions that produce the protein products. As there are likely multiple cryptic species here, we will need to rely more on the molecular data over the morphological data to find identifiable traits.

Additionally, I will aim to incorporate analyses for the geographic relationships between the haplogroups, as some of the overlapping populations (ex. N12, N13, and N5) are distantly related according to the BEAST trees and the K2P values. Ideally this will help us to predict why some sister groups are so geographically distant from each other while being surrounded by populations from different clades.

Acknowledgments

We wish to thank Dr. Robert Bradley, Emily Wright, Sarah Vrla, and Emma Roberts for help in preparing this presentation. We also wish to thank the Monte L. Bean Museum for providing DNA samples.