Department of Integrative Biology
University of California Berkeley
3040 Valley Life Sciences Bldg #5050
Berkeley, CA 94720
psmits@berkeley.edu

January 23, 2018

Editor
PeerJ

Dear Editor,

Please find enclosed the manuscript entitled "Ensemble approaches for estimating congruence between species delimitation and morphological variation: comparing taxonomic hypotheses for the Pacific Pond Turtle (*Emys marmorata*)" which we are submitting for consideration for the *PeerJ*. This manuscript is authored by myself (Peter D Smits), Kenneth D Angielczyk, James F Parham, and Bryan L Stuart. We believe our analysis and results are of interest to the community as we present a non-standard approach for evaluating the intensity of cryptic diversity and the efficacy of different classification schemes.

We investigated the morphometric identification of cryptic species using machine learning approaches by examining their implications for a recently proposed cryptic turtle species (*Emys pallida*). Our dataset consists of geometric morphometric data collected from 532 adult *E. marmorata/"E. pallida"* museum specimens. Each specimen was then assigned a classification for six different binning schemes based on geographic occurrence data recorded in museum collection archives. We analyzed our data using an ensemble of supervised machine learning approaches to determine which classification hypothesis was best supported by the data. In order to veryfiy our approach, we applied the same analysis to two alternate turtle morphology datasets, one consisting of eight unambiguously distinct species closely related to *E. marmorata*, and the other consisting of two subspecies of *Trachemys scripta*.

Our results indicate that there is no clear "best" grouping of *E. marmorata/"E. pallida"* based on plastron shape. In contrast, the analyses of the clear-cut examples produced near perfect classifications, demonstrating that the methods can recover correct results when an appropriate signal exists. Explanations for the lack of grouping in *E. marmorata* include the possibility that genetic differentiation is not associated with plastron shape variation below the species level and/or that local selective pressures (e.g., from hydrological regime) overwhelm morphological differentiation. A reconsideration of the methods used to delimit *"E. pallida,"* the lack of barriers to gene flow, the strong evidence for widespread admixture between lineages, and the fact that plastron shape can be

used to delineate other emydine species and sub-species suggest that its lack of diagnosability most likely reflects the non-distinctiveness of this proposed taxon.

Thank you for considering our work. Please send all correspondence regarding this manuscript to me via my email address (psmits@berkeley.edu).

<div style="text-align:center">Sincerely,</div>

<div style="text-align:center">Peter D Smits</div>

encl: Article; figures, tables.