How cryptic is cryptic diversity? Machine learning approaches to plastral variation in *Emys marmorata*.

Peter D Smits *[1], Kenneth D Angielczyk †[2], and James F Parham ‡[3]

[1]Committee on Evolution Biology, University of Chicago
[2]Department of Geology, Field Museum of Natural History
[3]Department of Geological Sciences, California State University – Fullerton

June 28, 2013

Abstract

1 Introduction

Cryptic diversity is when taxa were only first deliminated via molecular means and were not or cannot deliminated via morphological identification CITATION. The discovery of this previously unknown diversity has

Here, we address the question of how much of cryptic diversity may be a product of sample size as well as methodology used for classifying taxa based

*psmits@uchicago.edu
†kangielczyk@fieldmuseum.org
‡jparham@fullerton.edu

solely on morphology. Specifically, we ask if fine scale variation in morphology can provide corroboration for subspecific assignment, and if To analyze this question, we apply multiple machine learning approaches to estimate the best classification scheme of *E. marmorata* subspecies based on morphological variation in plastral shape.

2 Materials and Methods

2.1 Specimens

We collected morphometric data from 524 specimens. Geographic information was recorded from museum collection information. When precise latitude and longitude information was not known for a specimen, it was inferred from whatever locality information was presented.

Specimens were given a class assignment was based on geographic information. Because the exact geographic barriers between different class is unknown and fuzzy, two assignments for both morphological and molecular hypotheses of class were used.

2.2 Geometric morphometrics

Following Angielczyk et al. (2011), 19 landmarks were digitized using TpsDig 2.04 (Rohlf, 2005). 17 of these landmarks are at the endpoints or intersection of the keratinous plastral scutes that cover the platron. These landmarks were chosen to maximize the description of plastral variation. 12 of these landmarks are symmetrical across the axis of symmetry and in order to prevent degrees of freedom and other concerns (Klingenberg et al., 2007), these landmarks were reflected across the axis of symmetry and the average position of each symmetrical pair was used. Analysis was then conducted on the resulting “half” plastra.

“Half” plastra landmark configurations were superimposed using generalized Procrustes analysis (Dryden and Mardia, 1998) after which, the principal components of shape were calculated. This was done using the `shapes` package in R (Dryden, 2013; R Core Team, 2013).

2.3 Machine learning analyses

2.3.1 Unsupervised learning

2.3.2 Supervised learning

3 Results

3.1 Geometric morphometrics

3.2 Machine learning analyses

3.2.1 Unsupervised learning

3.2.2 Supervised learning

4 Discussion

Acknowledgements

PDS would like to thank David Bapst, Benjmin Frable, Michael Foote, and Dallas Krentzel for useful discussion which enhanced the quality of this study.

References

Angielczyk, K. D., C. R. Feldman, and G. R. Miller. 2011. Adaptive evolution of plastron shape in emydine turtles. Evolution 65:377–394.

Dryden, I. L. 2013. shapes: Statistical shape analysis. R package version 1.1-8.

Dryden, I. L., and K. Y. Mardia. 1998. Statistical shape analysis. Wiley, New York.

Klingenberg, C. P., M. Barluenga, and A. Meyer. 2007. Shape analysis of symetric structures: quantifying variation among individuals and asymmetry. Evolution 56:1909–1920.

R Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rohlf, F. J. 2005. TpsDig 2.04.