Supervised learning approaches to classifying morphological viariation: picking amongst taxonomic hypotheses for the Pacific Pond Turtle (*Emys marmorata*)

Peter D Smits[1], Kenneth D Angielczyk[1,2], James F Parham[3], and Bryan Stuart[4]

[1]Committee on Evolutionary Biology, University of Chicago
[2]Integrative Research Center, Field Museum of Natural History
[3]Department of Geological Sciences, California State University – Fullerton
[4]Section of Research and Collections, North Carolina Museum of Sciences

June 1, 2016

Corresponding author: Peter D Smits, Committee on Evolutionary Biology, University of Chicago, 1025 E. 57th Street, Culver Hall 402, Chicago, IL, 60637, USA; E-mail: psmits@uchicago.edu

Abstract

We investigate the morphometric identification of cryptic species using machine learning approaches by examining their implications for a recently proposed cryptic turtle species (*Emys pallida*). We collected landmark-based morphometric data from 354 adult *E. marmorata/"pallida"* museum specimens. We assigned a classification to each specimen for six different binning schemes based on geographic occurrence data recorded in museum collection archives. We used multiple machine learning methods, both unsupervised and supervised, to compare different classification hypotheses and asked whether it is possible to determine which amongst a set of classification hypotheses is best. In addition, we applied the above approach to two additional datasets: two well-supported subspecies of *Trachemys scripta* and a set of seven unambiguously distinct species closely related to *E. marmorata*. The results of this study indicate that there is no clear grouping of *E. marmorata/"pallida"* based on plastron shape. In contrast, the analysis of the other datasets demonstrate a near perfect classification, which demonstrates that application of the methods to plastron shape data can recover correct results in similar situations. Explanations for the lack of grouping in *E. marmorata* include that possibility that genetic differentiation is not associated with plastron

1

shape variation in this species complex 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 differentiate other emydid species and subspecies suggest that its lack of diagnosability most likely reflects the non-distinctiveness of this proposed taxon.

Introduction

Molecular systematics has repeatedly demonstrated the existence of cryptic species that can only be diagnosed using genetic data (Stuart et al. 2006; Bickford et al. 2007; Schilck-Steiner et al. 2007; Pfenninger and Schwenk 2007; Clare 2011; Funk et al. 2012). In attempts to streamline the documentation of biodiversity, several methods of species delimitation that rely almost entirely on genetic data have recently been proposed (Pons et al. 2006; Carstens and Dewey 2010; Hausdorf and Hennig 2010; O'Meara 2010; Yang and Rannala 2010; Huelsenbeck et al. 2011). Although strong caveats on the utility of these methods have been raised (Bauer et al. 2000; Carstens et al. 2013), they are nevertheless being used to name species (Leaché and Fujita 2010; Spinks et al. 2014).

The majority of extant taxa, and almost all extinct taxa, are delimited by morphology alone. This disjunction complicates interpretations of variation and diversity in deep time, as apparent morphological stasis may not reflect the true underlying diversity (Eldredge and Gould 1972; Gould and Eldredge 1977; Hunt 2008; Van Bocxlaer and Hunt 2013). Similarly, for many museum specimens of extant taxa (e.g. those preserved in formalin), it is difficult to acquire the genetic data needed for non-morphological species delimitation methods.

These considerations have sparked interest in whether geometric morphometric analyses can capture fine-scale variation that can be used for identifying cryptic species. This would make the task of identifying and maintaining endangered or conserved groups much easier and could contribute to improved classifications of extinct taxa and populations. Most such studies focus on using morphometrics to discover differences between taxa that were identified by other means (Polly 2003; Zelditch et al. 2004; Gaubert et al. 2005; Gündüz et al. 2007; Polly 2007; Demandt and Bergek 2009). Additionally, there has been work on automated taxon identification and classification of taxa into groups (Baylac et al. 2003; Dobigny et al. 2003; MacLeod 2007; van den Brink and Bokma 2011).

Here, we investigate the morphometric identification of cryptic species using machine learning approaches. In particular, we ask whether it is possible to determine which amongst a set of classification hypotheses is best and examine the implications of the results for a recently proposed set of cryptic turtle species.

Background and study system

Machine learning is an extension of known statistical methodology (Hastie et al. 2009) that emphasizes high predictive accuracy and generality at the expense of the interpretability of individual parameters. The basic statistical mechanics are supplemented by randomization, sorting, and partitioning algorithms, along with the maximization or minimization of summary statistics, in order to best estimate a general model for all data, both sampled and unsampled (Hastie et al. 2009). Machine learning approaches have found use in medical research, epidemiology, economics and automated identification of images such as handwritten zip codes (Hastie et al. 2009). The two major classes of machine learning methods are unsupervised and supervised learning. Unsupervised learning methods are used with unlabeled data where the underlying structure is estimated; they are analogous to clustering and density estimation methods (Kaufman and Rousseeuw 1990). Supervised learning methods are used with labeled data where the final output of data is known and the rules for going from input to output are inferred. These are analogous to classification and regression models (Breiman et al. 1984). Our application of the approaches used in this study illustrates only a sampling of the various methods available for clustering observations and fitting classification models.

Geometric morphometric approaches to identifying differences in morphological variation between different classes, including cryptic species, mostly have used methods like linear discriminate analysis and canonical variates analysis (Polly 2003; Zelditch et al. 2004; Gaubert et al. 2005; Gündüz et al. 2007; Polly 2007; Francoy et al. 2009; Sztencel-Jabonka et al. 2009; Mitrovski-Bogdanovic et al. 2013). Because of their similarity to multivariate approaches like principal components analysis (PCA), these methods are comparatively straightforward ways of understanding the differences in morphology between classes. They also benefit from producing results that can be easily visualized, which aids in the interpretation and presentation of data and results. Most previous morphometric studies did not assess which amongst a set of alternative classification hypotheses was optimal. For example, studies such as those of Caumul and Polly (2005) and Polly (2007) focused on comparing different aspects of morphology and their fidelity to a classification scheme instead of comparing the fidelity of one aspect of morphology to multiple classification schemes. In this context, the study of Cardini et al. (2009) is noteworthy because they compared morphological variation in marmots at the population, regional, and species level and determined the fidelity of shape to divisions at each of these levels.

Here, we used multiple machine learning methods, both unsupervised and supervised, to compare different classification hypotheses. These methods provide different advantages for understanding how to classify taxa, as well as the accuracy of the resulting classifications. Although machine learning methods such as neural networks have been applied to studying shape variation (Baylac et al. 2003; Dobigny et al. 2003; MacLeod 2007; van den Brink and Bokma 2011), including in the context of automated taxon identification and classification of groups, the number of cases remains limited. In the current study, we not only consider pure classification accuracy but also use a statistic of classification strength that reflects the rate at which taxa are both accurately and inaccurately classified.

We analyzed the problem of whether there are distinct subspecies or cryptic species within

the western pond turtle, *Emys marmorata* (Baird and Girard 1852) (formerly *Clemmys*
96 *marmorata*; see Feldman and Parham 2002). *Emys marmorata* is distributed from northern
Washington State, USA to Baja California, Mexico. Traditionally, *E. marmorata* was classified
98 into two named subspecies: the northern *E. marmorata marmorata* and the southern *Emys*
marmorata pallida (Seeliger 1945), with a central Californian intergrade zone in between.
100 *Emys marmorata marmorata* is differentiated from *E. marmorata pallida* by the presence of
a pair of triangular inguinal scales and darker neck markings. The triangular inguinal plates
102 can sometimes be present in *E. marmorata pallida* although they are considerably smaller.
Seeliger (1945) did not formally include the Baja California populations of *E. marmorata* in
104 either taxon, implying the existence of a third distinct but unnamed subspecies.

Previous work on morphological variation in *E. marmorata* has focused primarily on differen-
106 tiation between populations over a portion of the species' total range (Lubcke and Wilson
2007; Germano and Rathbun 2008; Germano and Bury 2009; Bury et al. 2010); comparatively
108 few studies have included specimens from across the entire range (Holland 1992). Most of
these studies considered how local biotic and abiotic factors may contribute to differences in
110 carapace length and found that size can vary greatly between different populations (Lubcke
and Wilson 2007; Germano and Rathbun 2008; Germano and Bury 2009). There also has
112 been interest in size-based sexual dimorphism in *E. marmorata* (Holland 1992; Lubcke and
Wilson 2007; Germano and Bury 2009), with males being on average larger than females
114 based on total carapace length and other linear measurements. However, the quality of size
as a classifier of sex can vary greatly between populations (Holland 1992) because of the
116 magnitude of size differences among populations (Lubcke and Wilson 2007; Germano and
Bury 2009). The effect of sexual dimorphism on shape, *sensu* Kendall (1977), has not been
118 assessed (Holland 1992; Lubcke and Wilson 2007; Germano and Rathbun 2008).

Of particular relevance in the context of cryptic diversity in *E. marmorata* is the morphometric
120 analysis of carapace shape carried out by Holland (1992), who compared populations of
E. marmorata from three areas of the species range. This study concluded that geographic
122 distance was a poor indicator of morphological differentiation, and instead hypothesized
that geographic features such as breaks between different drainage basis are probably more
124 important barriers to dispersal and interbreeding. Additionally, Holland (1992) suggested that
morphological differences were more pronounced as the magnitude of barriers and distance
126 increased, but this variation required many variables to adequately capture, implying only
very subtle morphological differentiation between putatively distinct populations. Finally,
128 Holland concluded that *E. marmorata* is best classified as three distinct species: a northern
species, a southern species, and a Columbia Basin species. This classification is similar to that
130 of Seeliger (1945), except elevated to the species level and without recognition of a distinct
Baja species.

132 More recently, the phylogeography of *E. marmorata* and the possibility of cryptic diversity
was investigated using molecular data (Spinks and Shaffer 2005; Spinks et al. 2010, 2014).
134 Based on mitochondrial DNA, Spinks and Shaffer (2005) recognized four subclades within
E. marmorata, a northern clade, a San Joaquin Valley clade, a Santa Barbara clade, and

4

a southern clade. Analyses with nuclear DNA (Spinks et al. 2010) and single-nucleotide polymorphism (SNP) data suggest a primarily north–south division in *E. marmorata*, although these datasets differed from that of Spinks and Shaffer (2005) in the location of the break point. All three studies discussed the potential taxonomic implications of their results, with Spinks et al. (2014) going so far as to strongly advocate for the recognition of at least two species (*E. marmorata* and *E. pallida*), and a possible third based on populations in Baja California. However, they did not discuss in detail the morphological characters that would help to diagnose these species beyond those specified by Seeliger (1945). Given that these characters are variable within the proposed species, and that Holland (1992) described shell shape variation that might be consistent with this taxonomy, a geometric morphometric analysis of shell shape might provide a reliable way to diagnose groups (whether species or subspecies) within *E. marmorata*.

In this study, we attempt to estimate the best classification scheme of *E. marmorata* based on variation in plastron (ventral shell) shape in order to determine whether this character is consistent with any of the past divisions based on other morphological features or molecular data. We are particularly interested in whether any aspect of plastron shape can be used to reliably diagnose the proposed species of Spinks et al. (2014), and if so, the nature of that shape variation.

WHY DID WE ANALYZE PLASTRON SHAPE? PETER DOESN'T HAVE A GOOD ANSWER.

Because of unclear geographic boundaries between subgroups of *E. marmorata*, we compare multiple hypotheses of morphologically– and molecularly–based classification. We hypothesize that if morphological variation corresponds to class assignment, it should be possible to determine the best classification hypothesis of *E. marmorata* from amongst multiple candidate hypotheses. However, if morphological variation does not correspond to any of the standing hypothesis, then supervised learning model generalization performance will be poor.

Materials and Methods

Specimens, sampling, morphometrics

Three different landmark-based morphometric datasets describing plastron variation were assembled for this analysis: specimens from seven distinct emydine species, *T. scripta* specimens from both subspecies, and *E. marmorata* specimens from across its geographic range. We chose to focus on adults because significant changes in plastron shape occur over the course of ontogeny in *E. marmorata* and other emydines (Angielczyk and Feldman 2013).

The first dataset is a compilation of 101 specimens of *T. scripta*: 51 specimens of *T. scripta scripta* and 50 specimens of *T. scripta scripta*. These landmark data are new to this study.

The second dataset, we analyzed 578 total specimens from the following species: *Emys blandigii, Terrapene coahuila, Clemmys guttata, Glyptemys insculpta, Glyptemys muhlenbergii,*

Emys orbicularis, and *Terrapene ornata*. Like the first data set, these specimens are a subset of those used in Angielczyk et al. (2011) and Angielczyk and Feldman (2013).

The final dataset dataset included 354 adult *E. marmorata* museum specimens; a subset of those included in Angielczyk and Sheets (2007), Angielczyk et al. (2011), and Angielczyk and Feldman (2013). We assigned a classification to each specimen for the different binning schemes based on geographic occurrence data recorded in museum collection archives. When precise latitude and longitude information were not available we estimated them from locality information. Because Spinks and Shaffer (2005), Spinks et al. (2010), and Spinks et al. (2014) did not use vouchered specimens we were not able to directly sample the individuals in their studies. Therefore our specimen classifications were based solely on the geographic information, not explicit assignment using molecular data. Because the exact barriers between different biogeographic regions are unknown and unclear, we represented some hypothesis with two schemes for a total of six different schemes. These schemes differed based on where geographic boundaries were assigned. This changes the classification of certain individuals near the boundaries between groups, providing a test of the robustness of the classification schemes. Sex information was only know for a subset of the total dataset and was not included as a predictor of classification. Sex information was used to determine if observations cluster by sex or not. The scheme names are as follows: Mito 1 and 2 correspond to Spinks and Shaffer (2005), Mito 3 corresponds to Spinks et al. (2010), Morph 1 and Morph 2 correspond to Holland (1992), and Nuclear corresponds to Spinks et al. (2014).

Following previous work on plastron shape (Angielczyk and Sheets 2007; Angielczyk et al. 2011; Angielczyk and Feldman 2013), we used TpsDig 2.04 (Rohlf 2005) to digitize 19 landmarks (Fig. 1). Seventeen of the landmarks are at the endpoints or intersection of the keratinous plastral scutes that cover the plastron. Twelve of the landmarks were symmetrical across the axis of symmetry. Because damage prevented the digitization of all the symmetric landmarks in some specimens, we reflected landmarks across the axis of symmetry (i.e. midline) prior to analysis and used the average position of each symmetrical pair. In cases where damage or incompleteness prevented symmetric landmarks from being determined, we used only the single member of the pair. We conducted all subsequent analyses on the resulting "half" plastra. We superimposed the plastral landmark configurations using generalized Procrustes analysis (Dryden and Mardia 1998), after which, we calculated the principal components (PC) of shape using the `shapes` package for R (**?**Dryden 2013).

Biasing effects

We estimated the possible effect of digitization error CITATIONS on our results by comparing within (replicated) specimen Procrustes distances to the distances between classification scheme centroids. 10 randomly selected specimen both for this study and an additional four times. These 50 landmark configurations were then Procrustes superimposed. A range of four Procrustes distances were then calculated as the average of the pairwise distances between each of the replicate configurations.
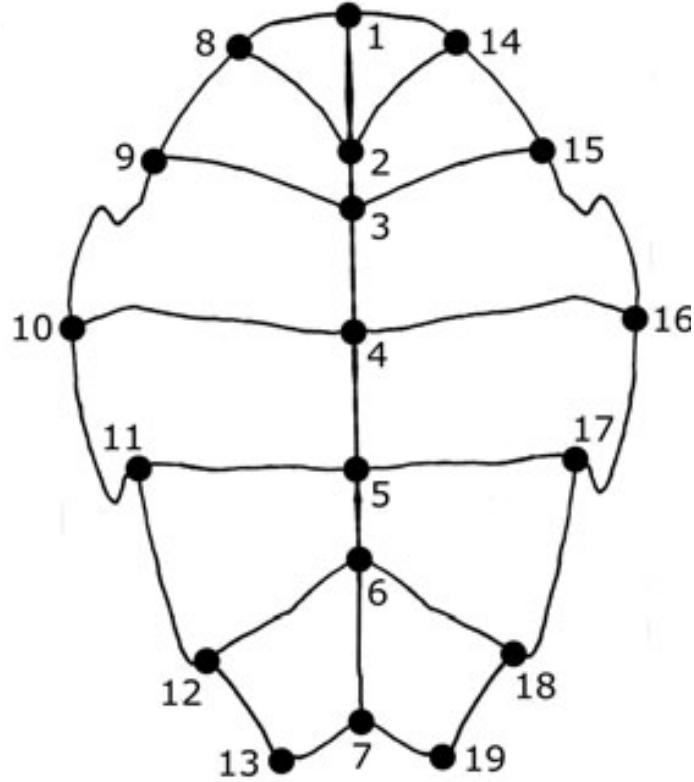
Figure 1: Depiction of general plastral shape of *E. marmorata* and position of the 19 landmark used in this study. Anterior is towards the top of the figure.

The relative magnitude of digitization error was calculated as the ratio between the average of the within-species replicate distances and the average distance between any two configurations in the replicate dataset. The goal of this ratio is to determine if the within group distances are on average smaller than the between individual distances; a value of 0 indicates perfect grouping, a value of 1 indicates no difference between grouping and no grouping, and a value of 1+ indicates that the grouping is counter-intuitive to the data.

Turtles are well known to demonstrate strong sexual dimorphism in carapace shape CITATION. To test for issues surrounding sexual dimorphism in our *E. marmorata* data set, we used a simple permutation test to determine if the distance between the mean female and male shapes is greater than expected when the sex labels are randomly shuffled. Because not all of our specimens have sex identifications associated with them, this analysis was done using a subset of the data.

Supervised learning approaches

The maximum set of possible predictors or features used for any model are the first 25 principal components (PCs), scaled centroid size, and the interaction between scaled centroid

7

size and PC 1. Size and the interaction between size and PC 1 were included as predictors in order to account for a possible interaction between size and shape over the duration of an individual as well as potential size differences between classes, even if this is unlikely (Seeliger 1945; Holland 1992). We say "maximum set" because the best or selected models based on 5-fold cross-validation does need not to, nor will they likely, include all predictors possible (see below).

This approach is in many ways analogous to PCA regresion. PCA regression takes advantage of two aspects of PCA for improving regression fit (Hastie et al. 2009). Because the PCs of shape are by definition orthogonal, allowing them to easily as independent predictors or features of class membership without fear of colinearity.

In classification studies, such as this one, a common metric of performance is the receiver operating characteristic (ROC) which is the relationship between the false and true positive rates (Hastie et al. 2009). The area under the ROC curce (AUC) is then the derived estimate of the model performance; AUC ranges from 0.5 to 1 which corresponding to performance similar to random guesses and perfect classificationrates, respectively (Hastie et al. 2009). Both ROC and AUC are preferrable to simple classificaiton accuracy when class membership is unbalanced, as it is in these analyses (Hastie et al. 2009). The standard ROC and AUC calculations are defined only for binary classifications, which is not the case for our seven species and *Emys* complex datasets. To generalize this approach for situations with mulitple response classes, we used an all-against-one strategy where the model AUC is the average of the AUC values from the multiple binary comparisons of one class compared to all others (Hand and Till 2001).

We adopted a training and testing paradigm for selecting parsimonious models and estimating their overall error rates (Hastie et al. 2009; Kuhn and Johnson 2013). Within-sample model performance is inherently biased upwards, so model evaluation requires overcoming this bias. With very large sample sizes, as in this study, part of the sample can be used as the "training set" and the remainder acts as the "testing set." In this approach, following all cleaning and vetting, the data is split into a training dataset and a testing dataset. The former is used for fitting the model where as the later is used for measuring model performance, a process called model generalization. In this analysis, we used 80% of samples as the training set while the remaining 20% were used as the testing set.

For a given supervised learning method, we compared the fit of 27 models as the average AUC from 10 rounds of 5-fold cross-validation. Cross-validation is an approach for estimating the average out-of-sample predictive error of a model by simulating out-of-sample data from the training data itself (Hastie et al. 2009). In a single round of k-fold cross-validation, the training data is divided into k blocks where the model is fit to $k-1$ blocks and the values of the kth block are predicted; this is then repeated for all combinations of blocks. Within each round, the predictive performance metrics is averaged across all folds. Finally, the predictive performance metric is the averaged across all rounds of k-fold cross-validation. This process was implemented using the R package `caret` CITATION.

Method name	abbreviation	R package	citation
multinomial logistic regression	MLR	nnet	citation
linear discriminate analysis	LDA	MASS	citation
penalized discrminiate analysis	PDA	mda	citation
single-hidden-layer neural network	NN	nnet	citation
random forests	RF	randomForest	citation

Table 1: table of the methods

For a given supervised learning method, the "best" trained model is that the highest mean AUC as estimated from 5-fold cross-validation. The selected or final model, however, is the next most parsimonious model that is within one standard error of the best model; this is a variant on the "one-standard error" rule from Hastie et al. (2009).

Instead of relying on a single supervised learning method, we chose to use an ensemble of mulitple approaches so that the congruence bewteen the them could be used as a means of "support" for one conclusion or another. The supervised learning methods used here are described in Table 1. Each of these methods makes different assumptions, treat data differently, and can produce different qualities of classification results depending on the nature of the data (Hastie et al. 2009).
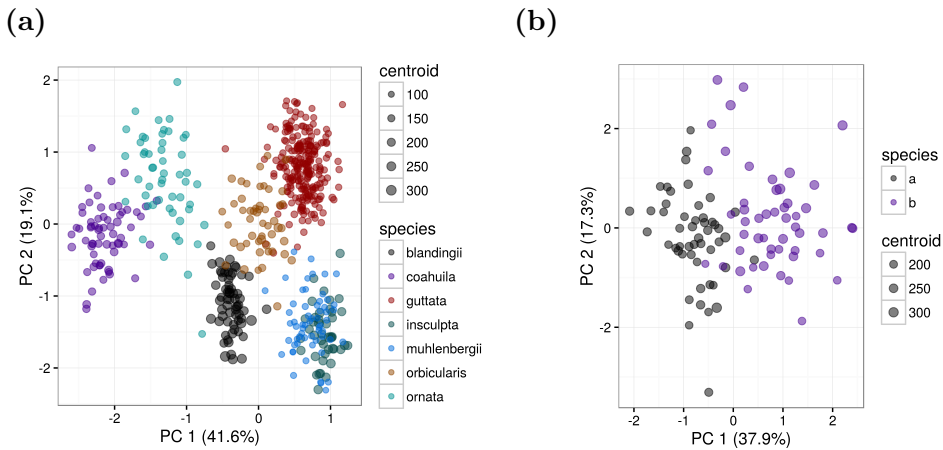
Results

Geometric morphometrics
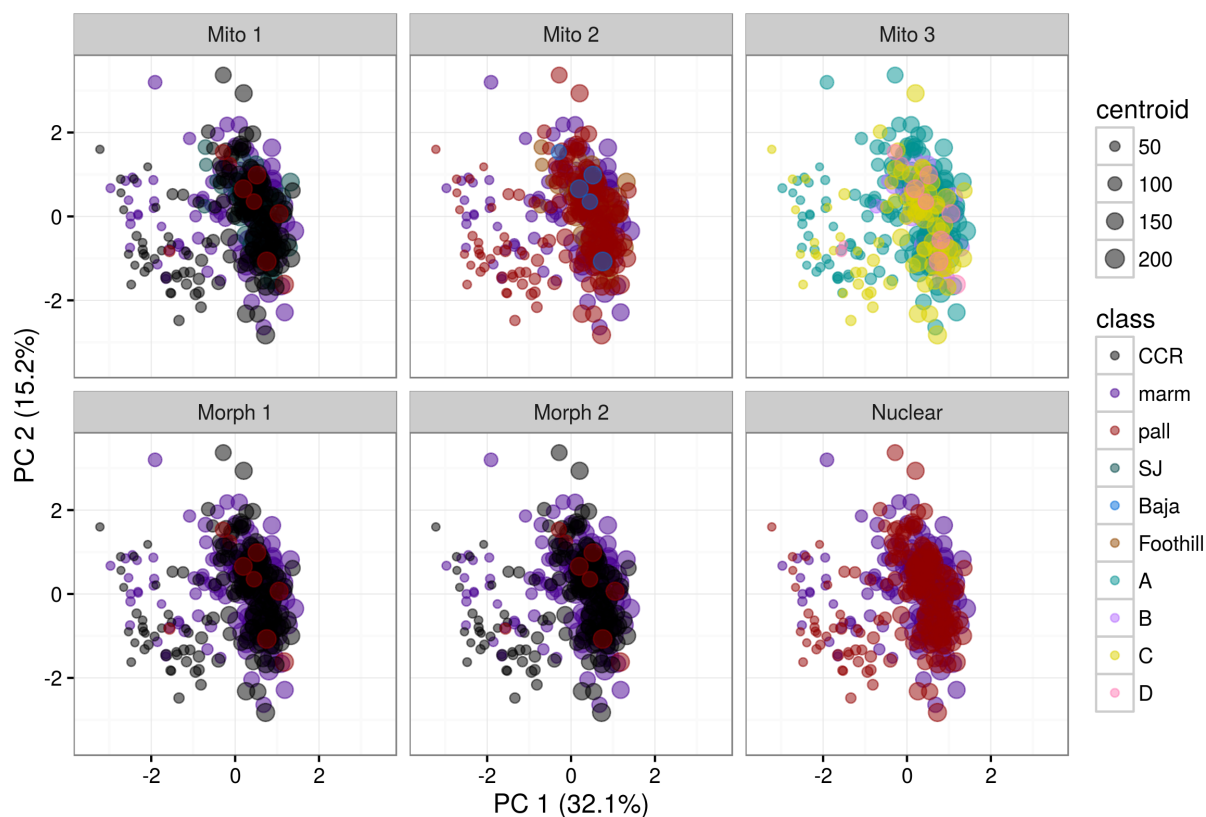

Figure 2: CAPTION

Supervised learning
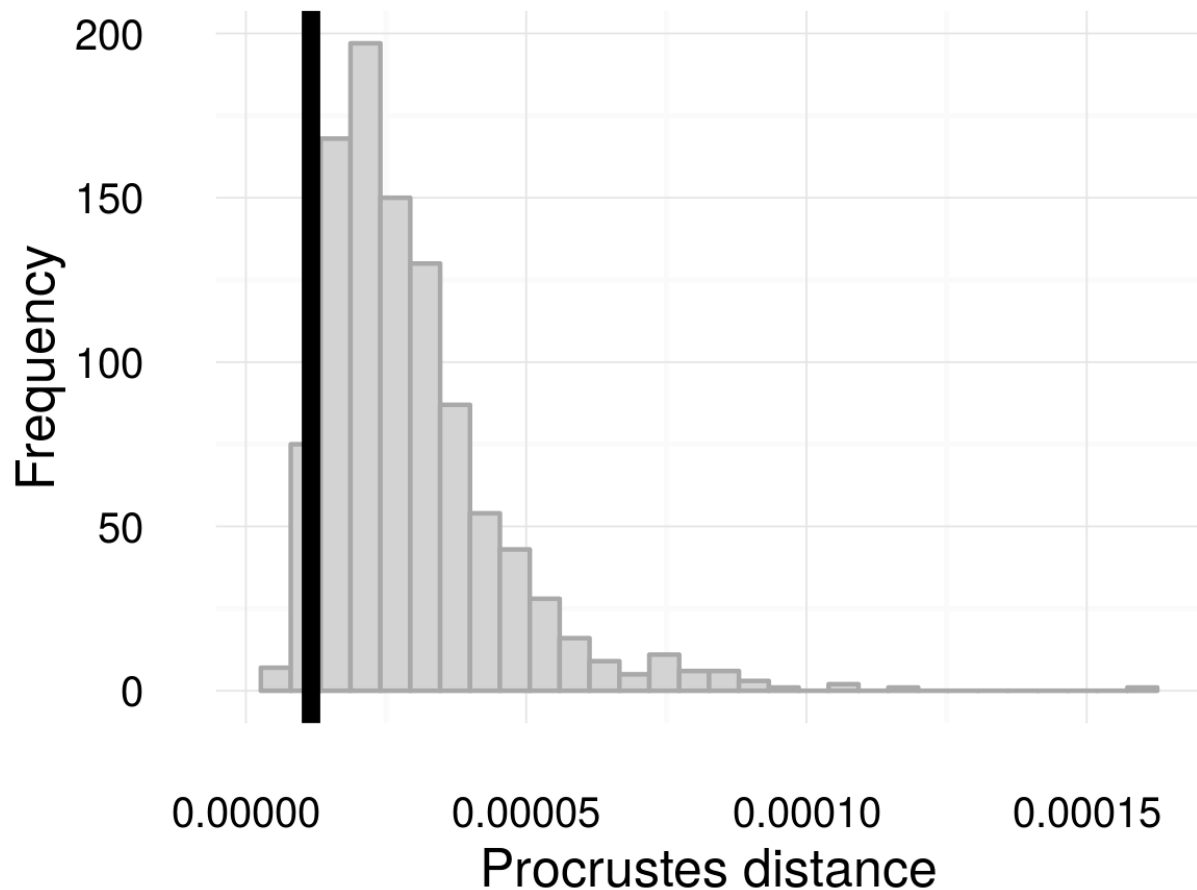
9

Figure 3: CAPTION

Discussion

Bibliography

Figure 4: CAPTION

K. D. Angielczyk and C. R. Feldman. Are diminutive turtles miniaturized? The ontogeny of plastron shape in emydine turtles. *Biological Journal of the Linnean Society*, 108(4): 727–755, apr 2013. ISSN 00244066. doi: 10.1111/bij.12010. URL `http://doi.wiley.com/10.1111/bij.12010`.

K. D. Angielczyk and H. D. Sheets. Investigation of simulated tectonic deformation in fossils using geometric morphometrics. *Paleobiology*, 33(1):125–148, 2007.

K. D. Angielczyk, C. R. Feldman, and G. R. Miller. Adaptive evolution of plastron shape in emydine turtles. *Evolution*, 65(2):377–394, feb 2011. ISSN 1558-5646. doi: 10.1111/j.1558-5646.2010.01118.x.

S. F. Baird and C. Girard. Descriptions of new species of reptiles collected by the U.S. Exploring Expedition under the command of Capt. Charles Wilkes. *Proceedings of the National Academy of Sciences Philadelphia*, 6:174–177, 1852.

A. M. Bauer, J. F. Parham, R. M. Brown, B. L. Stuart, L. Grismer, T. J. Papenfuss,

W. Bohme, J. M. Savage, S. Carranza, J. L. Grismer, P. Wagner, A. Schmitz, N. B. Ananjeva, and R. F. Inger. Availability of new Baysian-delimited gecko names and the importance of character-based species descriptions. *Proceedings of the Royal Society B: Biological Sciences*, 278:490–492, 2000. ISSN 07331347. doi: 10.2307/1467045. URL `http://www.jstor.org/stable/1467045?origin=crossref`.

M. Baylac, C. Villemant, and G. Simbolotti. Combining geometric morphometrics with pattern recognition for the investigation of species complexes. *Biological Journal of the Linnean Society*, 80:89–98, 2003.

D. Bickford, D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. Cryptic species as a window on diversity and conservation. *Trends in ecology & evolution*, 22(3):148–55, mar 2007. ISSN 0169-5347. doi: 10.1016/j.tree.2006.11.004. URL `http://www.ncbi.nlm.nih.gov/pubmed/17129636`.

L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and regression trees*. Wadsworth International Group, Belmont, 1984.

R. B. Bury, D. J. Germano, and G. W. Bury. Population Structure and Growth of the Turtle Actinemys marmorata from the KlamathSiskiyou Ecoregion: Age, Not Size, Matters. *Copeia*, 2010(3):443–451, sep 2010. ISSN 0045-8511. doi: 10.1643/CH-08-096. URL `http://www.bioone.org/doi/abs/10.1643/CH-08-096`.

A. Cardini, D. Nagorsen, P. O'Higgins, P. D. Polly, R. W. Thorington Jr, and P. Tongiorgi. Detecting biological distinctiveness using geometric morphometrics: an example case from the Vancouver Island marmot. *Ethology Ecology & Evolution*, 21:209–223, 2009.

B. C. Carstens and T. A. Dewey. Species Delimitation Using a Combined Coalescent and Information-Theoretic Approach: An Example from North American Myotis Bats. *Systematic Biology*, 59(4):400–414, 2010. URL `papers2://publication/doi/10.1093/sysbio/syq024`.

B. C. Carstens, T. a. Pelletier, N. M. Reid, and J. D. Satler. How to fail at species delimitation. *Molecular ecology*, 22(17):4369–83, sep 2013. ISSN 1365-294X. doi: 10.1111/mec.12413. URL `http://www.ncbi.nlm.nih.gov/pubmed/23855767`.

R. Caumul and P. D. Polly. Phylogenetic and environmental components of morphological variation: skull, mandible, and molar shape in marmots (Marmota, Rodentia). *Evolution; international journal of organic evolution*, 59(11):2460–72, nov 2005. ISSN 0014-3820. URL `http://www.ncbi.nlm.nih.gov/pubmed/16396186`.

E. L. Clare. Cryptic species? Patterns of maternal and paternal gene flow in eight neotropical bats. *PloS one*, 6(7):e21460, jan 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0021460.

M. H. Demandt and S. Bergek. Identification of cyprinid hybrids by using geometric morphometrics and microsatellites. *Journal of Applied Ichthyology*, 25(6):695–701, dec

2009. ISSN 01758659. doi: 10.1111/j.1439-0426.2009.01329.x. URL `http://doi.wiley.`
`com/10.1111/j.1439-0426.2009.01329.x`.

G. Dobigny, L. Granjon, V. Aniskin, K. Ba, and V. Voloboulev. A new sigling species of
Taterillus (Muridae, Gerbillinae) from West Agrica. *Mammalian Biology*, 68:299–316, 2003.

I. L. Dryden. *shapes: Statistical shape analysis*, 2013. URL `http://CRAN.R-project.org/`
`package=shapes`. R package version 1.1-8.

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

N. Eldredge and S. J. Gould. Punctuated equilibria: an alternative to phyletic gradualism.
In T. J. M. Schopf, editor, *Models in Paleobiology*, pages 82–115. Freeman Cooper, San
Francisco, 1972.

C. R. Feldman and J. F. Parham. Molecular phylogenetics of emydine turtles: taxonomic
revision and the evolution of shell kinesis. *Molecular Phylogenetics and Evolution*, 22(3):
388–98, mar 2002. ISSN 1055-7903. doi: 10.1006/mpev.2001.1070. URL `http://www.ncbi.`
`nlm.nih.gov/pubmed/11884163`.

T. M. Francoy, R. A. O. Silva, P. Nunes-Silva, C. Menezes, and V. L. Imperatriz-Fonseca.
Gender identification of five genera of stingless bees (Apidae, Meliponini) based on wing
morphology. *Genetics and molecular research*, 8(1):207–214, 2009.

W. C. Funk, M. Caminer, and S. R. Ron. High levels of cryptic species diversity uncovered
in Amazonian frogs. *Proceedings of the Royal Society B: Biological Sciences*, 279(1734):
1806–14, may 2012. ISSN 1471-2954. doi: 10.1098/rspb.2011.1653.

P. Gaubert, P. J. Taylor, C. a. Fernandes, M. W. Bruford, and G. Veron. Patterns of cryptic
hybridization revealed using an integrative approach: a case study on genets (Carnivora,
Viverridae, Genetta spp.) from the southern African subregion. *Biological Journal of the
Linnean Society*, 86(1):11–33, aug 2005. ISSN 00244066. doi: 10.1111/j.1095-8312.2005.
00518.x. URL `http://doi.wiley.com/10.1111/j.1095-8312.2005.00518.x`.

D. J. Germano and R. B. Bury. Variation in body size, growth, and population structure
of Actinemys marmorata from lentic and lotic habitats in Southern Oregon. *Journal of
Herpetology*, 43(3):510–520, 2009.

D. J. Germano and G. B. Rathbun. Growth, population structure, and reproduction of
western pond turtles (Actinemys marmorata) on the Central Coast of California. *Chelonian
Conservation and Biology*, 7(2):188–194, 2008.

S. J. Gould and N. Eldredge. Punctuated equilibria: the tempo and mode of evolution
reconsidered. *Paleobiology*, 3(2):115–151, 1977.

I. Gündüz, M. Jaarola, C. Tez, C. Yeniyurt, P. D. Polly, and J. B. Searle. Multigenic and
morphometric differentiation of ground squirrels (Spermophilus, Sciuridae, Rodentia) in
Turkey, with a description of a new species. *Molecular phylogenetics and evolution*, 43

(3):916–35, jun 2007. ISSN 1055-7903. doi: 10.1016/j.ympev.2007.02.021. URL http://www.ncbi.nlm.nih.gov/pubmed/17500011.

D. J. Hand and R. J. Till. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning*, 45:171–186, 2001.

T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction.* Springer, New York, 2nd edition, 2009.

B. Hausdorf and C. Hennig. Species delimitation using dominant and codominant multilocus markers. *Systematic biology*, 59(5):491–503, oct 2010. ISSN 1076-836X. doi: 10.1093/sysbio/syq039. URL http://www.ncbi.nlm.nih.gov/pubmed/20693311.

D. C. Holland. *Level and pattern in morphological variation: a phylogeographic study of the western pond turtle (Clemmys marmorata).* PhD thesis, University of Southwestern Louisiana, 1992.

J. P. Huelsenbeck, P. Andolfatto, and E. T. Huelsenbeck. Structurama: bayesian inference of population structure. *Evolutionary bioinformatics online*, 7:55–9, jan 2011. ISSN 1176-9343. doi: 10.4137/EBO.S6761. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3118697{&}tool=pmcentrez{&}rendertype=abstract.

G. Hunt. Gradual or pulsed evolution: when should punctuational explanations be preferred? *Paleobiology*, 34(3):360–377, 2008. doi: 10.1666/07073.1.

L. Kaufman and P. J. Rousseeuw. *Finding groups in data : an introduction to cluster analysis.* Wiley, New York, 1990.

D. G. Kendall. The diffusion of shape. *Advances in Applied Probability*, 9(3):428–430, 1977.

M. Kuhn and K. Johnson. *Applied predictive modeling.* Springer, New York, NY, 2013.

A. D. Leaché and M. K. Fujita. Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). *Proceedings. Biological sciences / The Royal Society*, 277(1697):3071–7, oct 2010. ISSN 1471-2954. doi: 10.1098/rspb.2010.0662. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2982061{&}tool=pmcentrez{&}rendertype=abstract.

G. M. Lubcke and D. S. Wilson. Variation in shell morphology of the Western Pond Turtle (Actinemys marmorata Baird and Giarard) from three aquativ habitats in Northern California. *Journal of Herpetology*, 41(1):107–114, 2007.

N. MacLeod. *Automated taxon identification in systematics: theory, approaches and applications.* CRC Press, Boca Raton, 2007.

A. Mitrovski-Bogdanovic, A. Petrovic, M. Mitrovic, A. Ivanovic, V. Žikic, P. Starý, C. Vorburger, and Ž. Tomanovic. Identification of two cryptic species within the Praon abjectum group (Hymenoptera: Braconidae: Aphidiinae) using molecular markers and geometric morphometrics. *Annals of the entomological society of America*, 106(2):170–180, 2013.

B. C. O'Meara. New heuristic methods for joint species delimitation and species tree inference. *Systematic biology*, 59(1):59–73, jan 2010. ISSN 1076-836X. doi: 10.1093/sysbio/syp077. URL http://www.ncbi.nlm.nih.gov/pubmed/20525620.

M. Pfenninger and K. Schwenk. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC evolutionary biology*, 7:121, jan 2007. ISSN 1471-2148. doi: 10.1186/1471-2148-7-121. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1939701{&}tool=pmcentrez{&}rendertype=abstract.

P. D. Polly. Paleophylogeography of Sorex araneus: molar shape as a morphological marker for fossil shrews. *Mammalia*, 68(2):233–243, 2003.

P. D. Polly. Phylogeographic differentiation in Sorex araneus: morphology in relation to geography and karyotype. *Russian Journal of Theriology*, 6(1):73–84, 2007.

J. Pons, T. Barraclough, J. Gomez-Zurita, A. Cardoso, D. Duran, S. Hazell, S. Kamoun, W. Sumlin, and A. Vogler. Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. *Systematic Biology*, 55(4):595–609, aug 2006. ISSN 1063-5157. doi: 10.1080/10635150600852011. URL http://sysbio.oxfordjournals.org/cgi/doi/10.1080/10635150600852011.

F. J. Rohlf. TpsDig 2.04, 2005.

B. C. Schilck-Steiner, B. Seifert, C. Stauffer, E. Christian, R. H. Crozier, and F. M. Steiner. Without morphology, cryptic species stay in taxonomic crypsis following discovery. *Trends in ecology & evolution*, 22(8):391–392, aug 2007. ISSN 0169-5347. doi: 10.1016/j.tree.2007.05.003. URL http://www.ncbi.nlm.nih.gov/pubmed/17573147.

L. M. Seeliger. Variation in the Pacific Mud Turtle. *Copeia*, 1945(3):150–159, 1945.

P. Q. Spinks and H. B. Shaffer. Range-wide molecular analysis of the western pond turtle (Emys marmorata): cryptic variation, isolation by distance, and their conservation implications. *Molecular ecology*, 14(7):2047–64, jun 2005. ISSN 0962-1083. doi: 10.1111/j.1365-294X.2005.02564.x. URL http://www.ncbi.nlm.nih.gov/pubmed/15910326.

P. Q. Spinks, R. C. Thomson, and H. B. Shaffer. Nuclear gene phylogeography reveals the historical legacy of an ancient inland sea on lineages of the western pond turtle, Emys marmorata in California. *Molecular ecology*, 19(3):542–56, feb 2010. ISSN 1365-294X. doi: 10.1111/j.1365-294X.2009.04451.x. URL http://www.ncbi.nlm.nih.gov/pubmed/20051011.

P. Q. Spinks, R. C. Thomson, and H. Bradley Shaffer. The advantages of going large: genome wide SNPs clarify the complex population history and systematics of the threatened western pond turtle. *Molecular Ecology*, pages n/a–n/a, mar 2014. ISSN 09621083. doi: 10.1111/mec.12736. URL http://doi.wiley.com/10.1111/mec.12736.

B. L. Stuart, R. F. Inger, and H. K. Voris. High level of cryptic species diversity revealed by sympatric lineages of Southeast Asian forest frogs. *Biology letters*, 2(3):470–4, sep 2006.

448 ISSN 1744-9561. doi: 10.1098/rsbl.2006.0505. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1686201{&}tool=pmcentrez{&}rendertype=abstract.

450 A. Sztencel-Jabonka, G. Jones, and W. BogdanowicZ. Skull Morphology of Two Cryptic Bat Species: Pipistrellus pipistrellus and P. pygmaeus A 3D Geometric Morphometrics
452 Approach with Landmark Reconstruction. *Acta Chiropterologica*, 11(1):113–126, jun 2009. ISSN 1508-1109. doi: 10.3161/150811009X465730. URL http://www.bioone.org/doi/
454 abs/10.3161/150811009X465730.

B. Van Bocxlaer and G. Hunt. Morphological stasis in an ongoing gastropod radiation from
456 Lake Malawi. *Proceedings of the National Academy of Sciences*, aug 2013. ISSN 0027-8424. doi: 10.1073/pnas.1308588110. URL http://www.pnas.org/cgi/doi/10.1073/
458 pnas.1308588110.

V. van den Brink and F. Bokma. Morphometric shape analysis using learning vector
460 quantization neural networks an example distinguishing two microtine vole species. *Annales Zoologici Fennici*, 48:359–364, 2011.

462 Z. Yang and B. Rannala. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, 107(20):9264–9, may 2010. ISSN
464 1091-6490. doi: 10.1073/pnas.0913022107. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2889046{&}tool=pmcentrez{&}rendertype=abstract.

466 M. L. Zelditch, D. L. Swiderski, and H. D. Sheets. *Geometric morphometrics for biologists: a primer.* Elsevier Academic Press, Amsterdam, 2004.
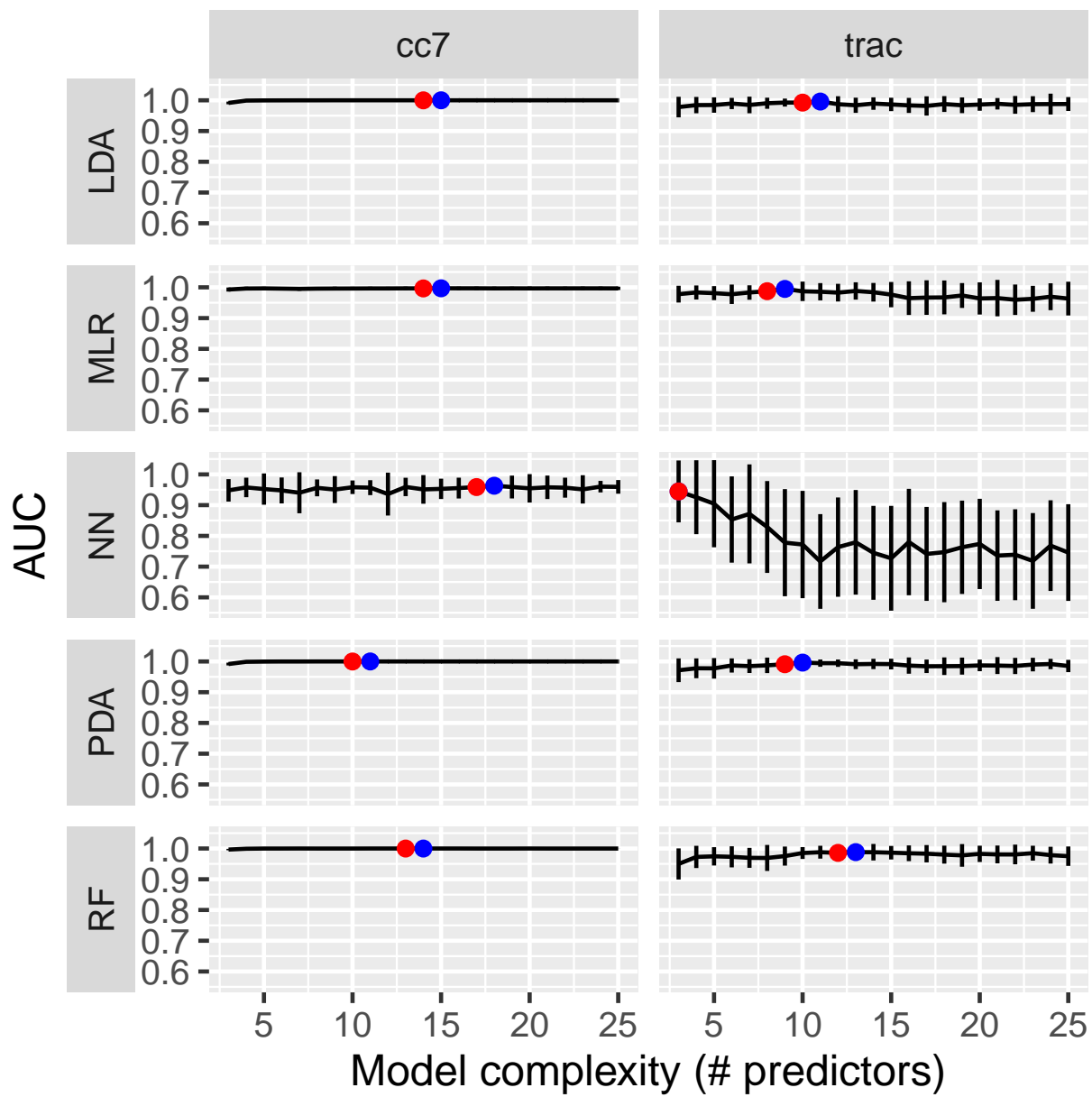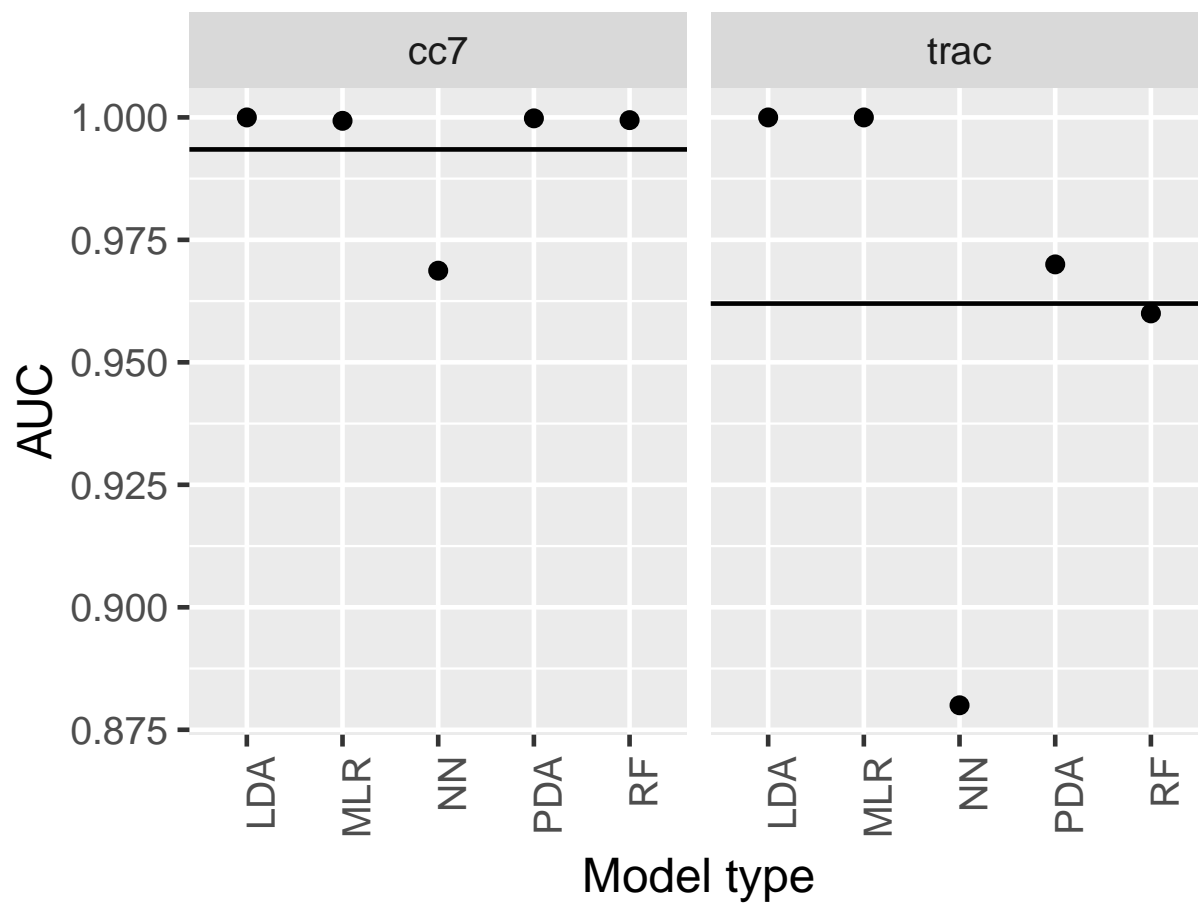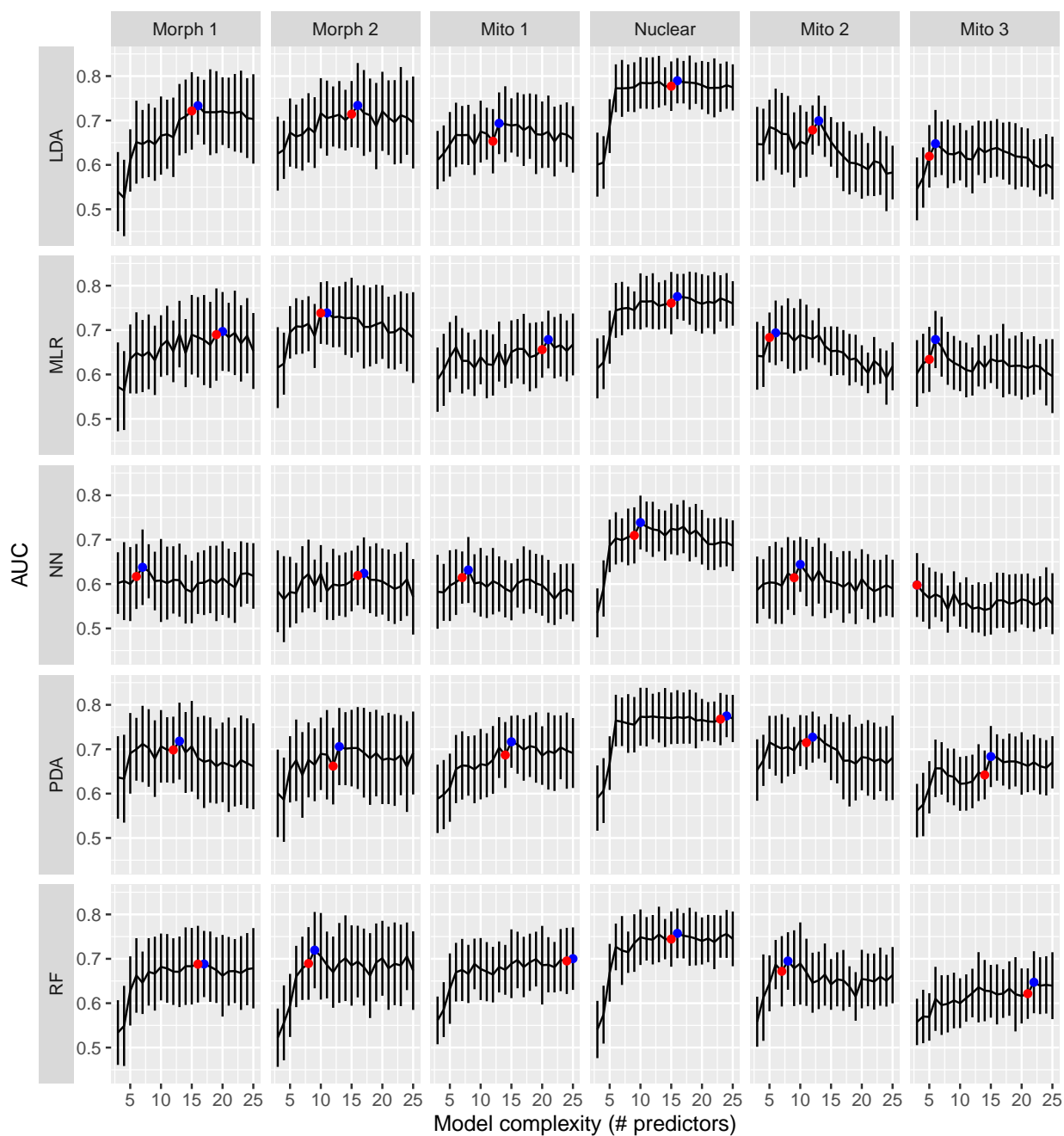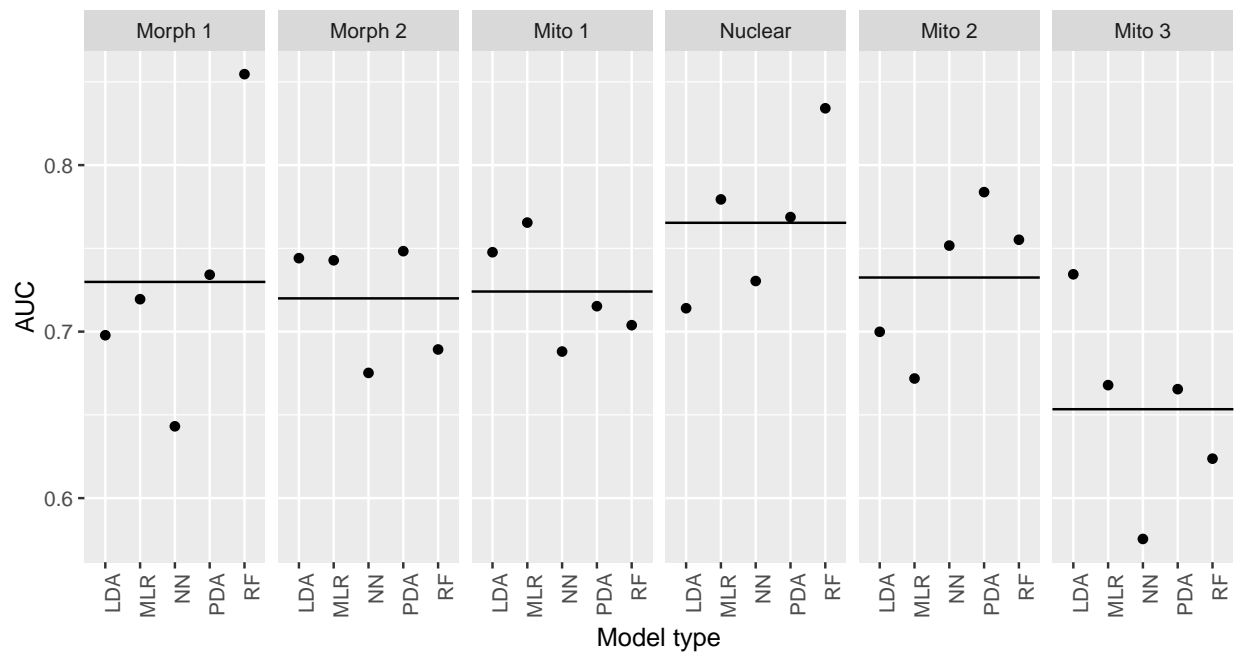
Figure 5: CAPTION

Figure 6: CAPTION

Figure 7: CAPTION

Figure 8: CAPTION