

Ensemble approaches for estimating congruence between species delimitation and morphological variation: a case study using the Pacific Pond Turtle (*Emys marmorata*)

Peter D Smits^{1*}, Kenneth D Angielczyk², Bryan L Stuart³, James F Parham⁴,

1 Department of Integrative Biology, University of California – Berkeley, Berkeley, California, USA.

2 Integrative Research Center, Field Museum of Natural History, Chicago, Illinois, USA.

3 Section of Research and Collections, North Carolina Museum of Natural Sciences, Raleigh, North Carolina, USA.

4 Department of Geological Sciences, California State University – Fullerton, Fullerton, California, USA.

* psmits@berkeley.edu

Abstract

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*). We used landmark-based morphometric data to quantify plastron shape in 532 adult *E. marmorata*/*E. 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 an ensemble of supervised machine learning approaches to determine which classification hypothesis was best supported by the data. In addition, we applied the same approach to two clear-cut examples, one consisting of seven unambiguously distinct species closely related to *E. marmorata* and an outgroup (*Chrysemys picta*), and the other consisting of two subspecies of *Trachemys scripta*. The analyses of the clear-cut examples produced near perfect classifications, demonstrating that plastron shape typically is a useful marker for differentiating turtle species, and that the methods can recover correct results when an appropriate signal exists. There is no clear “best” grouping of *E. marmorata*/*E. pallida* based on plastron shape. 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 emydid species and sub-species 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 [1–6]. In attempts to streamline the documentation of biodiversity, several methods of species delimitation that rely almost

entirely on genetic data have recently been proposed [7–12]. Although strong caveats on the utility of these methods have been raised [13, 14], they are nevertheless being used to name species [15, 16].

In contrast to those genetically-diagnosed species, 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 [17–19]. It also has serious implications for our records of modern biodiversity: 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. Most studies focus on using morphometrics to discover differences between taxa that were identified by other means [20–27]. Additionally, there has been work on automated taxon identification and classification of taxa into groups [28–32], as well as the development of models that combine genetic, phenotypic, and geographic data to infer evolutionary units of interest [33].

Here, we investigate the morphometric identification of cryptic species using machine learning approaches. We use an ensemble learning approach where multiple methods are used in order to look for consensus among their results. We test our approach on three datasets: plastron (ventral shell) shape of seven species of closely related turtles and one outgroup, plastron shape of two subspecies of a single turtle species, and plastron shape of the *Emys marmorata* species complex. In particular, we ask whether it is possible to determine which among a set of classification hypotheses best aligns with the observed morphology, and examine the implications of our results for the *E. marmorata* complex.

Background and study system

Machine learning is an extension of known statistical methodology that emphasizes predictive accuracy and generality, often at the expense of the interpretability of individual parameters [34]. Basic statistical approaches 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 [34]. Machine learning approaches have found use in medical research, epidemiology, economics, and automated identification of images such as handwritten zip codes [34].

There are two major classes of machine learning methods: 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 [35]. 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 [34, 36]. Our application of the supervised learning approaches used in this study illustrates only a sampling of the various methods available for fitting classification models. The specific methods we used were chosen because they are suited for cases with more two or more response classes.

Geometric morphometric approaches to identifying differences in morphological variation between classes, including cryptic species, have mostly relied on methods like linear discriminate analysis and canonical variates analysis [20–24, 37–41]. 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 Camul et al. [42] and Polly [24] 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. [43] 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 an ensemble of supervised machine learning methods to compare the congruence of morphological data to different classification hypotheses. Each of these methods provide different advantages for understanding how to classify specimens, as well as the accuracy of the resulting classifications. Machine learning methods have been combined with geometric morphometric data to study shape variation in a variety of contexts, including automated taxon identification and classification of groups [28–31, 44, 45]. 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: the area under the Receiver Operating Characteristic curve [34].

We analyzed the problem of whether there are distinct subspecies or cryptic species within the western pond turtle, *Emys marmorata* [46] (formerly *Clemmys marmorata*; see [47]). *Emys marmorata* is distributed from northern Washington State, USA to Baja California, Mexico; populations in western Nevada may have been introduced by recent human activity or they could be a genuine part of the species' range [48]. Traditionally, *E. marmorata* was classified into two named subspecies: the northern *E. marmorata marmorata* and the southern *Emys marmorata pallida* [49], with a central Californian intergrade zone in between. *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 can sometimes be present in *E. marmorata pallida* although they are considerably smaller. Seeliger [49] did not formally include the Baja California populations of *E. marmorata* in either taxon, implying the existence of a third distinct but unnamed subspecies.

Previous work on morphological variation in *E. marmorata* has focused primarily on differentiation between populations over a portion of the species' total range [50–53]; comparatively few studies have included specimens from across the entire range [54]. Most of these studies considered how local biotic and abiotic factors may contribute to differences in carapace length, and they found that size can vary greatly between different populations [50–52]. There also has been interest in size-based sexual dimorphism in *E. marmorata* [50, 52, 54], with males being on average larger than females based on total carapace length and other linear measurements. However, the quality of size as a classifier of sex can vary greatly between populations [54] because of the magnitude of size differences among populations [50, 52]. The effect of sexual dimorphism on shape, *sensu* Kendall [55], has not been assessed [50, 51, 54].

Of particular relevance in the context of cryptic diversity in *E. marmorata* is the morphometric analysis of carapace shape carried out by Holland [54], who compared populations of *E. marmorata* from three areas of the species' range. Holland concluded that geographic distance was a poor indicator of morphological differentiation, and instead hypothesized that geographic features such as breaks between different drainage basis are probably more important barriers to dispersal and interbreeding. Additionally, he suggested that morphological differences were more pronounced as the magnitude of barriers and distance increased, but this variation required many variables to adequately capture, implying only very subtle morphological differentiation between putatively distinct populations. Finally, 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 of [49], except elevated to the species level and without recognition of a distinct Baja species.

More recently, the phylogeography of *E. marmorata* and the possibility of cryptic diversity was investigated using molecular data [16, 56, 57]. Based on mitochondrial DNA, Spinks and Shaffer [56] recognized four subclades within *E. marmorata*, a northern clade, a San Joaquin Valley clade, a Santa Barbara clade, and a southern clade. Analyses with nuclear DNA [57] and single-nucleotide polymorphism (SNP) data suggest a primarily north–south division in *E. marmorata*, although these datasets differed from the mitochondrial-based results of Spinks and Shaffer [56] in the location of the break point [16]. All three studies discussed the potential taxonomic implications of their results, with Spinks et al. [16] 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 [49]. Given that these characters are variable within the proposed species, and that Holland [54] 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 proposed taxonomies of the *E. marmorata* complex.

We choose to analyze plastron shape for multiple reasons. First, it is very easy to collect geometric morphometric data on plastron shape from two-dimensional photographs as the structure is virtually flat. This approach allows both museum specimens and individuals photographed in the field to be analyzed together. Second, previous work has suggested that there are strong differences in plastron shape among traditionally-recognized emydine species [58–60]. Finally, due to these previous studies a large dataset was readily available.

In the case of the *E. marmorata* species complex, we hypothesize that if one or more of the proposed classification schemes are consistent with the morphological data then our ensemble approach fit to those hypotheses will have higher out-of-sample predictive performance than the more inconsistent hypotheses. However, if all of the classification schemes lead to equal out-of-sample predictive performance then we would conclude that the proposed hypotheses are inconsistent with whatever information is present in the morphological data. Because of unclear geographic boundaries between subgroups of *E. marmorata*, we compare multiple permutations of the Spinks et al. [57] and Spinks et al. [16] hypotheses.

Materials and methods

Specimens, sampling, morphometrics

Three different geometric morphometric datasets describing turtle plastron variation were assembled for this analysis: 1) specimens from seven distinct emydine species and a deriochelyine outgroup (*C. picta*); 2) *T. scripta* specimens from the two main subspecies (*T. scripta elegans* and *T. scripta scripta*); and 3) *E. marmorata* specimens from across the species' geographic range. The first two datasets are intended to serve as a test of whether machine learning techniques can differentiate species-level groupings of emydid turtles using plastron shape. We expect that the first case represents a low complexity dataset because of the high level of plastron shape disparity

Table 1. Table of species delimitation hypotheses for *E. marmorata*

Abbreviation	Number of classes	citation
SP10.1	3	[57]
SP10.2	3	[57]
SP10.3	4	[57]
SP14.1	2	[16]
SP14.2	4	[16]
Morph	2	[57]

that exists among these species [59,61,62], whereas the second dataset should be relatively higher in complexity and more analogous to the *E. marmorata* example. We predict that the *E. marmorata* dataset should be of the highest complexity and our greatest challenge given the previous finding that only very subtle differences existed between geographically-distinct populations [54].

The first dataset we analyzed includes 578 total specimens from the following species: *Chrysemys picta*, *Clemmys guttata*, *Emys blandingii*, *Emys orbicularis*, *Glyptemys insculpta*, *Glyptemys mühlenbergii*, *Terrapene coahuila*, and *Terrapene ornata*. These specimens are a subset of those used in previous studies [59,60].

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

The final dataset is of 532 adult *E. marmorata* museum specimens, though not all specimens were able to be assigned a class for all schemes (Fig. 1). These specimens represent a subset of those included in previous studies [58–60]. Because previous genetic studies did not use vouchered specimens [16,56,57] we were not able to directly sample the individuals in their studies. Instead, our specimen classifications were based solely on the geographic information and not explicit assignment using molecular data. For each taxonomic hypothesis, specimens were assigned to one of the possible classes based on geographic occurrence data recorded in museum collections. In cases where precise latitude and longitude information were not available we estimated them from other locality information. Because the exact barriers between different biogeographic regions are unknown and unclear, we represented each hypothesis with multiple possible realizations representing the classification uncertainty for specimens present at the geographic boundaries. The taxonomic hypotheses and sub-hypotheses for *E. marmorata* used here are presented in Table 1 and Fig. 1.

Fig 1. Geographic distribution of specimens sampled for comparing the hypothesized subdivisions of *Emys marmorata*. Each hypothesized scheme has two or more possible classes. Sample size differs between schemes because of our ability to confidently assign museum specimens to the various schemes. The number of localities shown on each map is less than the number of specimens sampled because some localities produced multiple specimens. The different classification abbreviations are as follows: *E. marmorata* = “marm”, *E. pallida* = “pall”, Central Coast Ranges = “CCR”, San Joaquin Valley = “SJ,” Baya Peninsula = “Baja,” and Sierra Foothills = “Foothill.”

We used three main binning schemes. All three schemes include a class for *E. marmorata* specimens from northern populations (marm) as well as a class for those assigned to *E. pallida* (pall) and an intergrade zone in the Central Coast Ranges (CCR). The schemes differ in the assignment of samples from the San Joaquin Valley (Fig. 1). Scheme SP10.1 and SP10.2 differ in the assignment of specimens from the western San Joaquin Valley to either CCR or marm reflecting uncertainty regarding their genetic affinity as explained above. In scheme SP10.3 these specimens are assigned to a San

Joaquin class reflecting the mitochondrial distinctiveness shown by Spinks and Shaffer [56]. We used two versions of the binning scheme based on the results of Spinks et al. [16]: SP14.1 reflects their phylogenetic network analysis and SP14.2 corresponds to their Bayesian species delimitation analysis. SP14.2 requires the addition of two new classes, “Baja” and “Foothill,” to accommodate the genetic groupings recovered by the SNP Structure analysis that was used to create the guide tree for the BPP species delimitation analysis [16]. Finally, we proposed a conservative morphological scheme (“Morph”) in order to compare the molecular hypotheses with something approximating the original taxonomic hypothesis for the group. This scheme is made up solely of the marm and pall classes from the SP10.3 scheme.

Sex was known only for a subset of the total dataset and was not included as a predictor of classification. Instead, we estimated the degree to which specimens cluster morphologically by sex in order to determine how much of a potential biasing factor sexual dimorphism could be for our analysis of the *E. marmorata* species complex (see below).

Following previous work on plastron shape [58–60], we used TpsDig 2.04 [63] to digitize 19 two-dimensional landmarks (Fig. 2). Seventeen of the landmarks are at the endpoints or intersection of the keratinous 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 symmetric 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 [64], after which we calculated the principal components (PC) of shape using the *shapes* package for R [65,66]. All specimens were used for superimposition, after which the subset labeled for each of the schemes were used in model training and testing (see below).

Fig 2. Depiction of general plastral shape of *E. marmorata* and position of the 19 landmarks used in this study. Anterior is towards the top of the figure.

Biasing effects

We estimated the possible effect of digitization error [67–69] on our results by comparing within-specimen (replicated) Procrustes distances to the distances between classification scheme centroids. Ten randomly-selected *E. marmorata* specimens were each digitized four times, with the original set of digitized coordinates serving as a fifth replicate. These 50 landmark configurations were Procrustes superimposed. A range of four Procrustes distances was then calculated as the average of the pairwise distances between each of the replicate configurations of a given specimen.

For each specimen, the difference in shape caused by digitization was calculated as the mean of all pairwise Procrustes distances between the five replicates of that specimen. The average distance between any two digitizations was calculated as the mean of all pairwise Procrustes distances between all replicates for all specimens. The ratio between these two values was used to assess the magnitude of variation caused by digitization. The goal of this ratio is to determine if the within group distances are smaller on average 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.

Table 2. Table of within-class sex ratios from those specimens with sex information for each classification schemes used in this analysis.

Scheme	class ratios
SP10.1	CCR: 70 F/60 M; Marm: 112/113; Pall: 20/26
SP10.2	CCR: 63/50; Marm: 119/123; Pall: 20/26
SP10.3	CCR: 63/50; Marm: 92/95; Pall: 20/26; SJ: 27/28
SP14.1	Marm: 112/117; Pall: 65/61
SP14.2	Baja: 0/5; Foothill: 20/19; Marm: 70/63; Pall: 45/38
Morph	Marm: 92/94; Pall: 20/26

Table 3. Table of the supervised learning methods used in this analysis.

Method name	abbreviation	R package	citation
multinomial logistic regression	MLR	nnet	[70]
linear discriminate analysis	LDA	MASS	[70]
penalized discriminate analysis	PDA	mda	[71]
single-hidden-layer neural network	NN	nnet	[70]
random forests	RF	randomForest	[72]

Emys marmorata is known to display sexual dimorphism in plastral shape, particularly the presence of a plastral concavity in males [49]. To test for biases resulting from sexual dimorphism in our *E. marmorata* dataset, 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 (257 of 532). This analysis was also done for each of the classification schemes in order to determine if sex is a biasing factor within these groups. As with the total dataset, only a limited number of specimens in each classification scheme have sex information, and the observed sex ratios for each of the classes within the classification schemes are summarized in Table 2. Importantly, these within-class sex ratios are based only on the subset of specimens with sex information.

Supervised learning approaches

Instead of relying on a single supervised learning method, we chose to use an ensemble approach where multiple model types are used in concert so that any congruence between them increases our support for that conclusion over another [34]. The supervised learning methods used here are named in Table 3. Each of these methods makes different assumptions, treats data differently, and can produce different classification results depending on the nature of the data [34]. For example, multinomial logistic regression is a type of generalized linear model, whereas random forest is itself an ensemble approach where multiple decision trees are fit to subsets of the full dataset and then averaged.

The maximum set of possible predictors or features used for any model of our dataset is comprised of the first 25 principal components (PCs), scaled centroid size, the interaction between scaled centroid size and PC 1, and the interaction between scaled centroid size and PC 2. Additional interaction terms were not considered because of model complexity/sample size concerns. Size and the interaction between size and PCs 1 and 2 were included as predictors to account for known ontogenetic variation in plastron shape [60] as well as potential size differences between classes, even if this is unlikely [49, 54]. These data constitute a “maximum set” because the best or selected models based on five-fold cross-validation need not, and likely will not, include all predictors possible (see below). Because our supervised learning models use PCs as

predictors, this approach is in many ways analogous to PCA regression. PCA regression takes advantage of the reduction and orthogonality PCs to improve regression fit [34]. Because the PCs of shape are by definition orthogonal, they can easily serve as independent predictors or features of class membership without fear of collinearity.

We adopted a training and testing paradigm for selecting parsimonious models and estimating their overall error rates [34, 73]. 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 are split into a training dataset and a testing dataset. The former is used for fitting the model whereas the later is used for measuring model performance, a process called model generalization. For each scheme, we limited the model training and testing to only those individuals with class labels for that scheme. In this analysis, we randomly divided 80% of samples into the training set and the remaining 20% into the testing set.

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 [34]. The area under the ROC curve (AUC) is the derived estimate of model performance; AUC ranges from 0.5 to 1 which correspond to performance similar to random guesses and perfect classification rates, respectively [34]. Both ROC and AUC are preferable to simple classification accuracy when class membership is unbalanced, as it is in these analyses [34]. The standard ROC and AUC calculations are defined only for binary classifications, which is not the case for our eight species and *Emys* complex datasets. To generalize this approach for situations with multiple 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 [74].

For a given supervised learning method, we compared the fit of 27 models as the average AUC from 10 rounds of five-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 dataset itself [34]. In a single round of k -fold cross-validation, the training data are divided into k blocks where the model is fit to $k - 1$ blocks and the values of the k th block are predicted. This is repeated for all combinations of blocks. Within each round, the predictive performance metrics are 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` [75]. For a given supervised learning method, the “best” trained model is that with the highest mean AUC as estimated from five-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 [34]. The purpose of this rule is to ameliorate the chances of selecting an overly complex model that will perform poorly when predicting the classes of out-of-sample data.

Results

Geometric morphometrics

The results of the PCA of plastron shape in both the eight species and *Trachemys* datasets demonstrate strong association between shape and the recognized classification schemes (Fig. 3).

The results of the PCA of plastron shape in the *Emys marmorata* dataset show no clear connection between plastron shape and any of the proposed classification schemes

Fig 3. Two scatterplots of morphological differences from two of the three datasets analyzed in this study. (a) Scatterplot of the first two PCA axes from the landmarks from the eight different species dataset, and (b) the first two axes of variation from two subspecies of *Trachemys* dataset. Point colors correspond to the categories within each dataset while point size is proportional to individual centroid size. In parentheses next to the axis labels are the percent of total variation accounted for by that axis. For both datasets there are clear distinctions between the different categories.

(Fig. 4). The first PC axis of shape variation appears to be primarily structured by differences in individual centroid size (Fig. 4); a linear regression of PC 1 vs log centroid size shows that size explains approximately 81 percent of the variation along this axis. This was the motivation for including centroid size and its interaction with PC1 as predictors in all of the supervised learning models. Size typically explains less than three percent of the variance of the remaining PC axes (e.g. 2.5% on PC 2).

Fig 4. Scatterplot of the first two axes of morphological variation in the *Emys marmorata* dataset. Each panel corresponds to one of the six different classification schemes analyzed as part of this study (Tab. 1). Point color corresponds to the categories within each scheme, and the class names correspond to geographic regions. Point size is proportional to centroid size of that specimen and the numbers in parentheses next to the axis labels are the percent of total variation accounted for along that axis.

Analysis of the differences between sexes of *E. marmorata* indicates that sex does not appear to strongly structure differences in shape (Fig. 5). The difference in mean shape between the sexes is very small; the sexes overlap about as much as expected given a null distribution based on permuting the sex-labels. The results from sex comparisons within each classification scheme have similar results with none of the observed differences being significantly different from the permuted distribution, indicating that sex is most likely not a biasing factor in our analyses (Fig. 6).

Fig 5. Comparison of observed Procrustes distance between the centroids of each sex (vertical line) to a null distribution generated from 1000 permutations of the sex-labels. This result indicates that the difference between the centroids is as small/smaller than expected by random.

Comparison of the within to between Procrustes distances of the digitization replicates gives an approximate estimate of the error between distinct groupings (Table 4). The ratio of the average within-individual distance to the average distance between individuals for the replicated datasets is 1.11; this indicates that the grouping is slightly counter-intuitive to the data and is consistent with all shapes being very similar regardless of individual identity. This value also provides a baseline by which to understand how distinct the groupings are, where other ratios are compared to the correction ratio 1.11/1.

The results from the eight species and *Trachemys* datasets indicate that both of these classification schemes are more recognizable than not given our estimate of digitization error (Table 4). In contrast, the different *E. marmorata* classification schemes appear to be barely distinct, with their within:between ratios approximating 1. This indicates that the magnitude of the differences between groupings is approximately the same as the difference between any two random individuals (Table 4).

Fig 6. Comparison of observed Procrustes distance between the centroids of each sex (vertical line) to a null distribution generated from 1000 permutations of the sex-labels for each of the six classification schemes. This result indicates that the difference between the centroids is as small/smaller than expected by random. For explanation of classification scheme names see Table 1.

Table 4. Results from the within-individual to between-individual Procrustes distances for the replicated plastron shape data. Results are presented for all three datasets analyzed here: the *Trachemys* dataset, the eight species dataset, and each of the *Emys marmorata* classification schemes.

Dataset	Scheme	Ratio	Corrected ratio
Replicates		1.06	
Seven species		0.43	0.46
<i>Trachemys</i>		0.76	0.81
<i>E. marmorata</i>	SP10.1	0.99	1.05
	SP10.2	1.00	1.06
	SP10.3	0.95	1.01
	SP14.1	1.01	1.07
	SP14.2	0.94	1.00
	Morph	0.99	1.05

Supervised learning

Analysis of the eight morphologically- and genetically-distinct species and the *T. scripta scripta*–*T. scripta elegans* datasets indicate that these taxa are sufficiently morphologically distinct to be differentiated on the basis of plastron shape. Both in-sample and out-of-sample classifications have AUC values of approximately 1 for all methods, implying near-perfect classification rates (Fig. 7, 8). For both datasets, the ROC scores from testing datasets are tightly clustered near AUC = 1 (Fig. 8). These results demonstrate that when there are distinctions between the states of the classification schemes (i.e., differences in plastron shape that correlate with the different taxonomic groups), the methods used here can recover them.

Fig 7. Comparisons of model fit to the training dataset for each of the supervised learning methods applied to the first two datasets; the results from the eight species dataset are presented in the left column, while those from the *Trachemys* dataset are presented in the right column. Models were fit to datasets of varying complexity, with the number of parameters listed along the x-axis. Model fit is measured as the area under the receiver operating characteristic (AUC), which ranges from 0.5 to 1. Error bars correspond to one standard error estimated from 10 rounds of 5-fold cross-validation. The red dot corresponds to the model fit with the highest mean AUC while the blue dot corresponds to the model selected for further analysis. In some cases, there is no difference in complexity between the best and selected models.

AUC-based model selection revealed some important patterns of variation and congruence between the classification schemes and the actual data. Generally, the best performing models tended to include about half the total number of possible PCs (Fig. 9).

Observed AUC values for all of the optimal models are lower for the *E. marmorata* dataset than for the other two datasets (Fig. 7, 9). In most cases the different proposed classification schemes are generally poor descriptors of the observed variation. It appears that the dataset is overwhelmed by noise (likely biological and analytical),

Fig 8. The results of out-of-sample predictive performance of the selected models for both the eight species (left) and *Trachemys* datasets. Predictive performance is measured as the area under the receiver operating characteristic (AUC), which ranges from 0.5 to 1. Points correspond to the individual out-of-sample predictive performance of the specific model, indicated along the x-axis. The horizontal bars correspond the average out-of-sample predictive performance of all the models.

Fig 9. AUC values for models of varying complexity fit to the *Emys marmorata* training datasets for each classification scheme. The x-axis corresponds to the total number of predictors included in each model, while the y-axis corresponds to the AUC value which is a measure of goodness of fit for classification datasets. A model with a high AUC value corresponds to better classification performance than a model with a lower AUC value. Standard errors on AUC estimates are calculated from 10 rounds of 5-fold cross-validation. Indicated are the best performing and the selected models, in red and blue respectively. In some cases, there is no difference in complexity between the best and selected models.

making any accurate classifications difficult at best. This observation is cemented with the generalizations of the models to the testing dataset (Fig. 10).

Fig 10. Comparison of out-of-sample AUC estimates from the predictions of selected models (Fig. 9), grouped by classification scheme. The horizontal line in each panel corresponds to the average AUC value across all models of that classification scheme.

Mean AUC values for the model generalizations, in most cases, are approximately equal to the observed AUC values from the training dataset (Fig. 9, 10). The cases in which the AUC from the generalizations is less than the observed indicate poor model fit and a poor classification scheme. Comparison of AUC values from the model generalizations do not indicate a clear “best” classification scheme (Fig. 9, 10). Only in the case of the conservative morphological hypothesis (“Morph”) is the mean AUC value potentially distinct from that of other schemes; in this case mean AUC is lower than the average of the other five schemes which indicates that the morphologically-based scheme performs more poorly than the molecularly-based ones. It is important to note, however, that the training and testing dataset for the “Morph” scheme is the smallest of the six schemes which may lead to poorer performance in in-sample and out-of-sample comparisons.

Discussion

As expected, our ensemble approach yields high out-of-sample classification performance for the first two datasets. These results indicate that in cases of clear class separation (Fig. 3) our approach is able to detect this and make good out-of-sample predictions.

In the case of the *E. marmorata* dataset, our results show that none of the proposed taxonomic hypotheses for the *E. marmorata* species complex are more consistent with morphological differentiation than any other proposal (Fig. 10). Both the low out-of-sample AUC values and the significant difference between the correctly and incorrectly classified observations support the conclusion that none of the hypothesized classification schemes are good descriptions of the observed plastral variation within *E. marmorata*. An analytical explanation of this result is that the level of digitization error in the *E. marmorata* dataset is so great as to swamp out any biological signal. We think this is unlikely because all of the specimens considered in our three analyses were

digitized by one of us (K.D.A.), and digitization error was not a problem in the eight species or *Trachemys* examples. There are also no features of the plastron of *E. marmorata* that would make it significantly more difficult to accurately digitize than the plastra of the other species.

Biological explanations include the possibility that genetic differentiation is not associated with plastron shape variation and/or that local selective pressures (e.g. from hydrological regime) overwhelm morphological differentiation. Both of these options seem plausible given that shell shape is influenced by selection for both protection and streamlining, but not necessary mate choice [76–80], and that shell shape in *E. marmorata* is known to vary among populations inhabiting water bodies with different flow regimes [50, 52, 54]. Plastron shape does not seem to preserve a strong phylogenetic signal at the interspecific level in emydine turtles, at least compared to the effect of the presence or absence of a plastral hinge [59], and our current results suggest that this may be the case for phylogeographic signal within emydine species as well. A final possibility (explored below) is that the proposed classification schemes themselves do not represent significant evolutionary lineages.

Despite the negative result for *E. marmorata*, it is important to note that plastron shape is an extremely effective method for differentiating classes in the additional datasets we investigated. The magnitude of shape differences between the eight species (measured as Procrustes distance between the eight species' mean shapes) is approximately an order of magnitude greater than the differences between the *E. marmorata* subgroups, and not surprisingly the machine learning methods had no trouble classifying the specimens correctly. However, the magnitude of the shape differences between the *T. scripta* subspecies is comparable to those separating the different *E. marmorata* subgroups, yet even in this case the machine learning methods returned an almost perfect classification. These results demonstrate that plastron shape is normally a good marker for differentiating real subgroups in close relatives of *E. marmorata*, and that our lack of results for *E. marmorata* is not simply a shortcoming of the methods we applied. Indeed, it begs the question of what factors have suppressed morphological differentiation of plastron shape in *E. marmorata* and *E. pallida* if they are distinct species. Invoking issues such as the role of the plastron in protection or the need for streamlining are insufficient because the other species are expected to be subject to similar constraints [78, 80]. Although it may seem counterintuitive that plastron shape is both useful for species delimitation but has weak or absent phylogenetic signal, it is important to remember that these are different goals. While phylogenetically similar species may not be morphologically similar (e.g. compare the box turtles of the genus *Terrapene* to the closely related spotted turtle *Clemmys guttata*), the variation within a species typically is much less than the variation between species. Therefore, the consistent plastron shapes that characterize different emydid species leads to plastron shape being a useful tool for species delimitation, even when other selective factors have overprinted similarities stemming from patterns of descent from common ancestors.

Conclusions

The lack of morphological support for the distinctiveness of *E. pallida* does not, on its own, preclude the recognition of this taxon. However, the apparent lack of congruence between phenotypic and genetic data does prompt a reexamination of the methods and concepts that led to that taxonomic revision, especially considering that plastron shape is demonstrably capable of differentiating species and subspecies among other emyids. In other words, before we can assess the significance of the morphological non-diagnosability, it is essential to evaluate the methods and concepts that led to the

initial taxonomic revision.

Spinks et al. [16] elevated *E. pallida* based on a species delimitation analysis of SNP data using BPP [11]. However, Spinks et al. [16] did not heed the caveats about such species delimitation methods raised by Carstens et al. [14]. In addition to specifically addressing the shortcomings of validation methods such as BPP that rely on guide trees and “should be interpreted with caution,” Carstens et al. [14] also strongly emphasized that “Inferences regarding species boundaries based on genetic data alone are likely inadequate, and species delimitation should be conducted with consideration of the life history, geographical distribution, morphology and behaviour (where applicable) of the focal system. . .” These caveats evoke the development of the Unified Species Concept [81, 82], Integrative Taxonomy [83], and other pluralist approaches to species delimitation. None of these considerations were brought to bear on the *E. marmorata* system until now, and in doing so we find the proposal that *E. pallida* is a distinct species to be lacking.

The only other range-wide analysis of morphological variation in the *E. marmorata* complex is that of Seeliger [49], which was cited by Spinks et al. [16] as being “consistent” with their results. A closer examination of Seeliger [49] raises questions about that claim. For example, Spinks et al. [16] claim that Seeliger [49] suggested that “an area of extensive intergradation was restricted to the San Joaquin Valley.” But Seeliger actually hypothesized that her *E. m. pallida* and *E. m. marmorata* also intergrade in the vicinity of the San Francisco Bay (Coast Ranges), and lists localities in Alameda, Contra Costa, and Santa Clara counties where this occurs. Spinks et al. lack samples from the Bay Area, but the closest specimens to the east of the Bay Area show signs of intergradation. Thus, the putative genetic distinctiveness of *E. marmorata* and *E. pallida* may be partially due to a sampling artifact.

A second issue is that the morphological characters that Seeliger [49] used to distinguish the two putative taxa show considerable variation. The condition of the inguinal scales supposedly can be used to diagnose *E. marmorata* (present and larger) and *E. pallida* (smaller or absent). However, she noted that 11% of of the *E. marmorata marmorata* specimens in her dataset lack scales, 40% of her *E. marmorata pallida* specimens possess them, and specimens from inland central California have “all types of inguinals” (p. 156). Although she qualitatively described differences in the size of the inguinals in the subspecies, she did not quantify these, and modern methods for quantifying shape differences in the scales she observed lay decades in the future. Seeliger [49] also mentions distinct tendencies in coloration of the neck between the subspecies, but this character shows a similar pattern of variation to the inguinal scales. Nineteen percent and 33% of *E. marmorata marmorata* and *E. marmorata pallida* specimens, respectively, possess the coloration pattern expected of the other subspecies, and specimens from the areas of intergradation “have both types of coloration and neither is predominant” (p.157). Further studies of variation in these characters is clearly warranted, especially in the broad intergrade zones.

The natural history and geographical distribution of *E. marmorata* and *E. pallida* also make the recognition of these two taxa implausible. The mitochondrial data from Spinks et al. [16] show extensive introgression and admixture in Central California, which is likely not just a mitochondrial sweep because there are no significant barriers to nuclear gene flow in this region. Spinks et al. also lack sampling from the populations between the two putative species in the San Francisco Bay Area, which we predict would likely show even more genetic mixing. Combined with the well-demonstrated ability for testudinoid turtles, including emydids and even *Emys*, to hybridize (e.g. [84–86]) it is hard to imagine how *E. marmorata* and *E. pallida* could maintain their integrity in the face of such admixture. Any argument for the validity of *E. pallida* as a distinct species needs to address these points. Other emydines show

fairly high levels of genetic diversity that is often geographically structured (e.g. *G. insculpta* [87], *C. guttata* [88], and *E. blandingii* [89]) but are still regarded as single species. Because the geography, natural history, limited sampling from key areas, demonstrated genetic admixture of *E. marmorata*, and comparisons with other morphologically diagnosable species and subspecies conflict with the recognition of *E. pallida*, we hypothesize that *E. pallida* is not a distinct species.

We fully agree with Spinks et al. [16] that *E. marmorata* (*sensu lato*) is a species deserving of strong conservation efforts, and we do not wish to trivialize this need. Moreover, the genetic diversity uncovered by the analysis of Spinks et al. [16] should be accounted for explicitly in any conservation plan. Given the apparent lack of morphological distinction combined with the broad range of intergradation and other problems with the species hypothesis outlined above, we recommend that the populations elevated to *E. pallida* by [16] are best considered Evolutionary Significant Units or Distinct Population Segments instead of distinct species.

Finally, it is important to note that the data and analyses we present do not let us definitively say whether the apparent lack of morphological divergence within *E. marmorata* truly reflects the presence of a single species, or if it is an artifact of plastron shape being a poor morphological marker for phylogenetic and phylogeographic divergences in the case of *E. marmorata*. Further tests of both our preferred conclusion (*E. marmorata* as a single species) and that of Spinks et al. [16] should include morphological and molecular analyses of specimens from within the Bay Area intergrade zone, ideally the same set of voucher specimens. We also recommend additional tests of species delimitation using alternative methods and corroborating evidence as suggested by Carstens et al. [14]. From a morphological standpoint, support for the validity of “*E. pallida*” may come from other aspects of morphology, such as carapace shape or other features. Likewise, further investigation of the phylogeographic utility of plastron shape in other turtle species will help to clarify whether the lack of differentiation seen in *E. marmorata*, and the strong differentiation among the other emydids, is typical or an unusual case.

Acknowledgments

Data collection for this project was supported in part by NSF DBI-0306158 (to KDA). G. Miller assisted with data collection and her participation in this research was supported by NSF REU DBI-0353797 (to R. Mooi of CAS). For access to emydine specimens, we thank: J. Vindum and R. Drewes (CAS); A. Resetar (FMNH); R. Feeney (LACM); C. Austin (LSUMNS); S. Sweet (MSE); J. McGuire and C. Conroy (MVZ); A. Wynn (NMNH); P. Collins (SBMNH); B. Hollingsworth (SDMNH); P. Holroyd (UCMP). We are grateful for S. Sweet for field assistance and the California Department of Fish and Game for permits. We would also like to thank M. Lambruschi (FMNH) for help with figure 1.

References

1. Stuart BL, Inger RF, Voris HK. High level of cryptic species diversity revealed by sympatric lineages of Southeast Asian forest frogs. *Biology letters*. 2006;2(3):470–4. doi:10.1098/rsbl.2006.0505.
2. Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, et al. Cryptic species as a window on diversity and conservation. *Trends in ecology & evolution*. 2007;22(3):148–55. doi:10.1016/j.tree.2006.11.004.

3. Schilck-Steiner BC, Seifert B, Stauffer C, Christian E, Crozier RH, Steiner FM. Without morphology, cryptic species stay in taxonomic crypsis following discovery. *Trends in ecology & evolution*. 2007;22(8):391–392. doi:10.1016/j.tree.2007.05.003.
4. Pfenninger M, Schwenk K. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC evolutionary biology*. 2007;7:121. doi:10.1186/1471-2148-7-121.
5. Clare EL. Cryptic species? Patterns of maternal and paternal gene flow in eight neotropical bats. *PloS one*. 2011;6(7):e21460. doi:10.1371/journal.pone.0021460.
6. Funk WC, Caminer M, Ron SR. High levels of cryptic species diversity uncovered in Amazonian frogs. *Proceedings of the Royal Society B: Biological Sciences*. 2012;279(1734):1806–14. doi:10.1098/rspb.2011.1653.
7. Pons J, Barraclough T, Gomez-Zurita J, Cardoso A, Duran D, Hazell S, et al. Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. *Systematic Biology*. 2006;55(4):595–609. doi:10.1080/10635150600852011.
8. Carstens BC, Dewey TA. Species Delimitation Using a Combined Coalescent and Information-Theoretic Approach: An Example from North American Myotis Bats. *Systematic Biology*. 2010;59(4):400–414.
9. Hausdorf B, Hennig C. Species delimitation using dominant and codominant multilocus markers. *Systematic biology*. 2010;59(5):491–503. doi:10.1093/sysbio/syq039.
10. O'Meara BC. New heuristic methods for joint species delimitation and species tree inference. *Systematic biology*. 2010;59(1):59–73. doi:10.1093/sysbio/syp077.
11. Yang Z, Rannala B. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*. 2010;107(20):9264–9. doi:10.1073/pnas.0913022107.
12. Huelsenbeck JP, Andolfatto P, Huelsenbeck ET. Structurama: bayesian inference of population structure. *Evolutionary bioinformatics online*. 2011;7:55–9. doi:10.4137/EBO.S6761.
13. Bauer AM, Parham JF, Brown RM, Stuart BL, Grismer L, Papenfuss TJ, et al. Availability of new Bayesian-delimited gecko names and the importance of character-based species descriptions. *Proceedings of the Royal Society B: Biological Sciences*. 2000;278:490–492. doi:10.2307/1467045.
14. Carstens BC, Pelletier Ta, Reid NM, Satler JD. How to fail at species delimitation. *Molecular ecology*. 2013;22(17):4369–83. doi:10.1111/mec.12413.
15. Leaché AD, Fujita MK. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings Biological sciences / The Royal Society*. 2010;277(1697):3071–7. doi:10.1098/rspb.2010.0662.
16. Spinks PQ, Thomson RC, Bradley Shaffer H. The advantages of going large: genome wide SNPs clarify the complex population history and systematics of the threatened western pond turtle. *Molecular Ecology*. 2014; p. n/a–n/a. doi:10.1111/mec.12736.
17. Eldredge N, Gould SJ. Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf TJM, editor. *Models in Paleobiology*. San Francisco: Freeman Cooper; 1972. p. 82–115.

18. Gould SJ, Eldredge N. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology*. 1977;3(2):115–151.
19. Van Bocxlaer B, Hunt G. Morphological stasis in an ongoing gastropod radiation from Lake Malawi. *Proceedings of the National Academy of Sciences*. 2013;doi:10.1073/pnas.1308588110.
20. Polly PD. Paleophylogeography of *Sorex araneus*: molar shape as a morphological marker for fossil shrews. *Mammalia*. 2003;68(2):233–243.
21. Zelditch ML, Swiderski DL, Sheets HD. Geometric morphometrics for biologists: a primer. Amsterdam: Elsevier Academic Press; 2004.
22. Gaubert P, Taylor PJ, Fernandes Ca, Bruford MW, Veron G. Patterns of cryptic hybridization revealed using an integrative approach: a case study on genetids (Carnivora, Viverridae, *Genetta* spp.) from the southern African subregion. *Biological Journal of the Linnean Society*. 2005;86(1):11–33. doi:10.1111/j.1095-8312.2005.00518.x.
23. Gündüz I, Jaarola M, Tez C, Yenyurt C, Polly PD, Searle JB. Multigenic and morphometric differentiation of ground squirrels (*Spermophilus*, *Sciuridae*, *Rodentia*) in Turkey, with a description of a new species. *Molecular phylogenetics and evolution*. 2007;43(3):916–35. doi:10.1016/j.ympev.2007.02.021.
24. Polly PD. Phylogeographic differentiation in *Sorex araneus*: morphology in relation to geography and karyotype. *Russian Journal of Theriology*. 2007;6(1):73–84.
25. Demandt MH, Bergek S. Identification of cyprinid hybrids by using geometric morphometrics and microsatellites. *Journal of Applied Ichthyology*. 2009;25(6):695–701. doi:10.1111/j.1439-0426.2009.01329.x.
26. Markolf M, Rakotonirina H, Fichtel C, von Grumbkow P, Brameier M, Kappeler PM. True lemurs... true species - species delimitation using multiple data sources in the brown lemur complex. *BMC Evolutionary Biology*. 2013;13(1):233. doi:10.1186/1471-2148-13-233.
27. Fruciano C, Franchini P, Raffini F, Fan S, Meyer A. Are sympatrically speciating Midas cichlid fish special? Patterns of morphological and genetic variation in the closely related species *Archocentrus centrarchus*. *Ecology and Evolution*. 2016;6(12):4102–4114. doi:10.1002/ece3.2184.
28. Baylac M, Villemant C, Simbolotti G. Combining geometric morphometrics with pattern recognition for the investigation of species complexes. *Biological Journal of the Linnean Society*. 2003;80:89–98.
29. Dobigny G, Granjon L, Aniskin V, Ba K, Voloboulev V. A new sigling species of *Taterillus* (*Muridae*, *Gerbillinae*) from West Africa. *Mammalian Biology*. 2003;68:299–316.
30. MacLeod N. Automated taxon identification in systematics: theory, approaches and applications. Boca Raton: CRC Press; 2007.
31. van den Brink V, Bokma F. Morphometric shape analysis using learning vector quantization neural networks — an example distinguishing two microtine vole species. *Annales Zoologici Fennici*. 2011;48:359–364.

32. Vitek NS, Manz CL, Gao T, Bloch JI, Strait SG, Boyer DM. Semi-supervised determination of pseudocryptic morphotypes using observer-free characterizations of anatomical alignment and shape. *Ecology and Evolution*. 2017;7(14):5041–5055. doi:10.1002/ece3.3058.
33. Guillot G, Renaud S, Ledevin R, Michaux J, Claude J. A unifying model for the analysis of phenotypic, genetic, and geographic data. *Systematic Biology*. 2012;61(6):897–911. doi:10.1093/sysbio/sys038.
34. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer; 2009.
35. Kaufman L, Rousseeuw PJ. Finding groups in data : an introduction to cluster analysis. New York: Wiley; 1990.
36. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. Belmont: Wadsworth International Group; 1984.
37. Franco TM, Silva RAO, Nunes-Silva P, Menezes C, Imperatriz-Fonseca VL. Gender identification of five genera of stingless bees (Apidae, Meliponini) based on wing morphology. *Genetics and molecular research*. 2009;8(1):207–214.
38. Sztencel-Jablonka A, Jones G, Bogdanowicz W. Skull Morphology of Two Cryptic Bat Species: *Pipistrellus pipistrellus* and *P. pygmaeus* — A 3D Geometric Morphometrics Approach with Landmark Reconstruction. *Acta Chiropterologica*. 2009;11(1):113–126. doi:10.3161/150811009X465730.
39. Edwards S, Claude J, Van Vuuren BJ, Matthee CA. Evolutionary history of the Karoo bush rat, *Myotomys unisulcatus* (Rodentia: Muridae): Discordance between morphology and genetics. *Biological Journal of the Linnean Society*. 2011;102(3):510–526. doi:10.1111/j.1095-8312.2010.01583.x.
40. Mitrovski-Bogdanovic A, Petrovic A, Mitrovic M, Ivanovic A, Žikic V, Starý P, et al. 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*. 2013;106(2):170–180.
41. Dillard KC. A comparative analysis of geometric morphometrics across two *Pseudemys* turtle species in east central Virginia [Masters]. Virginia Commonwealth University; 2017.
42. Caumul R, Polly PD. Phylogenetic and environmental components of morphological variation: skull, mandible, and molar shape in marmots (*Marmota*, Rodentia). *Evolution; international journal of organic evolution*. 2005;59(11):2460–72.
43. Cardini A, Nagorsen D, O'Higgins P, Polly PD, Thorington Jr RW, Tongiorgi P. Detecting biological distinctiveness using geometric morphometrics: an example case from the Vancouver Island marmot. *Ethology Ecology & Evolution*. 2009;21:209–223.
44. Van Bocxlaer B, Schultheiß R. Comparison of morphometric techniques for shapes with few homologous landmarks based on machine-learning approaches to biological discrimination. *Paleobiology*. 2010;36(3):497–515.

45. Navega D, Vicente R, Vieira DN, Ross AH, Cunha E. Sex estimation from the tarsal bones in a Portuguese sample: a machine learning approach. *International Journal of Legal Medicine*. 2015;129(3):651–659. doi:10.1007/s00414-014-1070-5.
46. Baird SF, Girard C. 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*. 1852;6:174–177.
47. Feldman CR, Parham JF. Molecular phylogenetics of emydine turtles: taxonomic revision and the evolution of shell kinesis. *Molecular Phylogenetics and Evolution*. 2002;22(3):388–98. doi:10.1006/mpev.2001.1070.
48. Bury RB. Biogeography of Western Pond Turtles in the western Great Basin: Dispersal Across a Northwest Passage ? *Western Wildlife*. 2017;4:72–80.
49. Seeliger LM. Variation in the Pacific Mud Turtle. *Copeia*. 1945;1945(3):150–159.
50. Lubcke GM, Wilson DS. Variation in shell morphology of the Western Pond Turtle (*Actinemys marmorata* Baird and Giarard) from three aquativ habitats in Northern California. *Journal of Herpetology*. 2007;41(1):107–114.
51. Germano DJ, Rathbun GB. Growth, population structure, and reproduction of western pond turtles (*Actinemys marmorata*) on the Central Coast of California. *Chelonian Conservation and Biology*. 2008;7(2):188–194.
52. Germano DJ, Bury RB. Variation in body size, growth, and population structure of *Actinemys marmorata* from lentic and lotic habitats in Southern Oregon. *Journal of Herpetology*. 2009;43(3):510–520.
53. Bury RB, Germano DJ, Bury GW. Population Structure and Growth of the Turtle *Actinemys marmorata* from the Klamath–Siskiyou Ecoregion: Age, Not Size, Matters. *Copeia*. 2010;2010(3):443–451. doi:10.1643/CH-08-096.
54. Holland DC. Level and pattern in morphological variation: a phylogeographic study of the western pond turtle (*Clemmys marmorata*). University of Southwestern Louisiana; 1992.
55. Kendall DG. The diffusion of shape. *Advances in Applied Probability*. 1977;9(3):428–430.
56. Spinks PQ, Shaffer HB. Range-wide molecular analysis of the western pond turtle (*Emys marmorata*): cryptic variation, isolation by distance, and their conservation implications. *Molecular ecology*. 2005;14(7):2047–64. doi:10.1111/j.1365-294X.2005.02564.x.
57. Spinks PQ, Thomson RC, Shaffer HB. 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*. 2010;19(3):542–56. doi:10.1111/j.1365-294X.2009.04451.x.
58. Angielczyk KD, Sheets HD. Investigation of simulated tectonic deformation in fossils using geometric morphometrics. *Paleobiology*. 2007;33(1):125–148.
59. Angielczyk KD, Feldman CR, Miller GR. Adaptive evolution of plastron shape in emydine turtles. *Evolution*. 2011;65(2):377–394. doi:10.1111/j.1558-5646.2010.01118.x.

60. Angielczyk KD, Feldman CR. Are diminutive turtles miniaturized? The ontogeny of plastron shape in emydine turtles. *Biological Journal of the Linnean Society*. 2013;108(4):727–755. doi:10.1111/bij.12010.
61. Claude J, Paradis E, Tong H, Auffray JC. A geometric morphometric assessment of the effects of environment and cladogenesis on the evolution of the turtle shell. *Biological Journal of the Linnean Society*. 2003;79(December):485–501. doi:10.1046/j.1095-8312.2003.00198.x.
62. Claude J. Convergence induced by plastral kinesis and geometric morphometric assessment: a geometric morphometric assessment. *Fossil Turtle Research*. 2006;1:34–45.
63. Rohlf FJ. TpsDig 2.04; 2005.
64. Dryden IL, Mardia KY. Statistical shape analysis. New York: Wiley; 1998.
65. R Core Team. R: A Language and Environment for Statistical Computing; 2016. Available from: <http://www.R-project.org/>.
66. Dryden IL. shapes: Statistical shape analysis; 2013. Available from: <http://CRAN.R-project.org/package=shapes>.
67. Arnqvist G, Mårtensson T. Measurement error in geometric morphometrics: Empirical strategies to assess and reduce its impact on measures of shape; 1998.
68. von Cramon-Taubadel N, Frazier BC, Lahr MM. The problem of assessing landmark error in geometric morphometrics: theory, methods, and modifications. *American journal of physical anthropology*. 2007;132(4):535–544. doi:10.1002/ajpa.
69. Munoz-Munoz F, Perpignan D. Measurement error in morphometric studies: comparison between manual and computerized methods. *Ann Zool*. 2010;47(1):46–56.
70. Venables WN, Ripley BD. Modern Applied Statistics with S. 4th ed. New York: Springer; 2002. Available from: <http://www.stats.ox.ac.uk/pub/MASS4>.
71. Hastie T, Tibshirani R, Leisch F, Hornik K, Ripley BD. mda: Mixture and Flexible Discriminant Analysis; 2015. Available from: <https://CRAN.R-project.org/package=mda>.
72. Liaw A, Wiener M. Classification and Regression by randomForest. *R News*. 2002;2(3):18–22.
73. Kuhn M, Johnson K. Applied predictive modeling. New York, NY: Springer; 2013.
74. Hand DJ, Till RJ. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning*. 2001;45:171–186.
75. Kuhn M. caret: Classification and Regression Training; 2013. Available from: <http://CRAN.R-project.org/package=caret>.
76. Rivera G. Ecomorphological variation in shell shape of the freshwater turtle *Pseudemys concinna* inhabiting different aquatic flow regimes. *Integrative and comparative biology*. 2008;48(6):769–87. doi:10.1093/icb/icn088.

77. Rivera G, Stayton CT. Finite element modeling of shell shape in the freshwater turtle *Pseudemys concinna* reveals a trade-off between mechanical strength and hydrodynamic efficiency. *Journal of morphology*. 2011;272(10):1192–203. doi:10.1002/jmor.10974.
78. Stayton CT. Biomechanics on the half shell: functional performance influences patterns of morphological variation in the emydid turtle carapace. *Zoology (Jena, Germany)*. 2011;114(4):213–23. doi:10.1016/j.zool.2011.03.002.
79. Rivera G, Davis JN, Godwin JC, Adams DC. Repeatability of Habitat-Associated Divergence in Shell Shape of Turtles. *Evolutionary Biology*. 2014; p. 29–37. doi:10.1007/s11692-013-9243-6.
80. Polly PD, Stayton CT, Dumont ER, Pierce SE, Rayfield EJ, Angielczyk KD. Combining geometric morphometrics and finite element analysis with evolutionary modeling: towards a synthesis. *Journal of Vertebrate Paleontology*. 2016;4634(March). doi:10.1080/02724634.2016.1111225.
81. Dayrat B. Towards integrative taxonomy. *Biological Journal of the Linnean Society*. 2005;85:407–415.
82. De Queiroz K. Species concepts and species delimitation. *Systematic Biology*. 2007;56(6):879–86. doi:10.1080/10635150701701083.
83. Padial JM, Miralles A, De la Riva I, Vences M. The integrative future of taxonomy. *Frontiers in Zoology*. 2010;7(16):1–14.
84. Buskirk SW, Parham JF, Feldman CR. On the hybridisation between two distantly related Asian turtles (Testudines: *Scalia* x *Mauremys*). *Salamandra*. 2005;41:21–26.
85. Spinks PQ, Shaffer HB. Conflicting mitochondrial and nuclear phylogenies for the widely disjunct *Emys* (Testudines: Emydidae) species complex, and what they tell us about biogeography and hybridization. *Systematic biology*. 2009;58(1):1–20. doi:10.1093/sysbio/syp005.
86. Parham JF, Papenfuss TJ, Dijk PPV, Wilson BS, Marte C, Schettino LR, et al. Genetic introgression and hybridization in Antillean freshwater turtles (*Trachemys*) revealed by coalescent analyses of mitochondrial and cloned nuclear markers. *Molecular phylogenetics and evolution*. 2013;67(1):176–87. doi:10.1016/j.ympev.2013.01.004.
87. Amato ML, Brooks RJ, Fu J. A phylogeographic analysis of populations of the wood turtle (*Glyptemys insculpta*) throughout its range. *Molecular Ecology*. 2008;17(2):570–581. doi:10.1111/j.1365-294X.2007.03580.x.
88. Davy CM, Murphy RW. Conservation genetics of the endangered Spotted Turtle (*Clemmys guttata*) illustrate the risks of “bottleneck tests”. *Canadian Journal of Zoology*. 2014;92(2):149–162. doi:10.1139/cjz-2013-0188.
89. Sethuraman A, McGaugh SE, Becker ML, Chandler CH, Christiansen JL, Hayden S, et al. Population genetics of Blanding’s turtle (*Emys blandingii*) in the midwestern United States. *Conservation Genetics*. 2014;15(1):61–73. doi:10.1007/s10592-013-0521-8.