WILEY **MOLECULAR ECOLOGY**

Integrating life history traits into predictive phylogeography

Jack Sullivan[1,2]* | Megan L. Smith[3]* | Anahí Espíndola[1,4] | Megan Ruffley[1,2] | Andrew Rankin[1,2] | David Tank[1,2] | Bryan Carstens[3]

[1]Department of Biological Sciences, University of Idaho, Moscow, Idaho

[2]Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, Idaho

[3]Department of Ecology, Evolution and Organismal Biology, The Ohio State University, Columbus, Ohio

[4]Department of Entomology, University of Maryland, College Park, Maryland

Correspondence
Jack Sullivan, Department of Biological Sciences, University of Idaho, Moscow, ID. Email: jacks@uidaho.edu

Abstract

Predictive phylogeography seeks to aggregate genetic, environmental and taxonomic data from multiple species in order to make predictions about unsampled taxa using machine-learning techniques such as Random Forests. To date, organismal trait data have infrequently been incorporated into predictive frameworks due to difficulties inherent to the scoring of trait data across a taxonomically broad set of taxa. We refine predictive frameworks from two North American systems, the inland temperate rainforests of the Pacific Northwest and the Southwestern Arid Lands (SWAL), by incorporating a number of organismal trait variables. Our results indicate that incorporating life history traits as predictor variables improves the performance of the supervised machine-learning approach to predictive phylogeography, especially for the SWAL system, in which predictions made from only taxonomic and climate variables meets only moderate success. In particular, traits related to reproduction (e.g., reproductive mode; clutch size) and trophic level appear to be particularly informative to the predictive framework. Predictive frameworks offer an important mechanism for integration of organismal trait, environmental data, and genetic data in phylogeographic studies.

KEYWORDS
comparative phylogeography, cryptic diversity, machine learning, random forest

1 | INTRODUCTION

The role of geography in the generation of biological diversity has been central to the development of evolutionary theory, including Darwin's focus on the Galapagos Islands and Wallace's studies of the Sahul and Sunda Shelves. Comparative phylogeography (e.g., Bermingham & Avise, 1986) developed as a modern manifestation of this interest and was the focus of a recent colloquium (Avise, Bowen, & Ayala, 2016) that addressed the current limitations, and future prospects of the discipline. Several papers that emerged from that colloquium issued a call for the development of trait-based approaches for phylogeography to move forward (e.g., Papadopoulou

& Knowles, 2016; Riddle, 2016; Zamudio, Bell, & Mason, 2016), echoing earlier studies that integrated traits into phylogeography (e.g., Paz, Ibáñez, Lips, & Crawford, 2015).

Concurrent with this colloquium, Espíndola et al. (2016) introduced a supervised machine-learning approach using random forests (RF) (Breiman, 2001) that integrates comparative phylogeography with distributional and climate data housed in publicly available databases such as GBIF and WorldClim (Fick & Hijmans, 2017). The method of Espíndola et al. (2016) uses these geospatial data as predictor variables to develop a classifier that permits predictions to be made about the nature of genetic variation (e.g., the presence or absence of cryptic diversity, as indicated by deep genetic divergence or phylogeographic breaks) in taxa that have little to no genetic data yet available. Specifically, an existing comparative phylogeographic

*Co-first authors on this paper

wileyonlinelibrary.com/journal/mec

data set is used as the training data set to build the classifier, and, of the various supervised machine-learning options available, RF permit an assessment of the importance of predictor variables in the predictive function.

In its original version, the predictive phylogeographic approach of Espíndola et al. (2016) used a machine learning framework to predict whether or not unsampled species were likely to harbour cryptic diversity, which was defined originally as pre-Pleistocene divergence across a geographic barrier. Since the Espíndola et al. (2016) Pacific Northwest training data set included two woody plants, a mammal, three amphibians, and two terrestrial invertebrates, life history traits that could be coded across all included species were difficult to identify. To address this shortcoming, taxonomic categories served as a proxy for life history traits and were used with occurrence records for species endemic to a particular region or ecosystem as observations to be classified. Predictive accuracy was reasonably good in Espíndola et al. (2016) and the predictive phylogeographic framework remains ideally suited to incorporation of trait data as predictor variables, particularly for comparative phylogeographic investigations that are more taxonomically focused.

Although many current comparative phylogeographic data sets are taxonomically restricted, for example to a regional avifauna (e.g., Kirchman & Franklin, 2007), or to regional anurans (e.g., Freilich et al., 2016), cross kingdom studies (e.g., Carstens, Brunsfeld, Demboski, Good, & Sullivan, 2005; Hewitt, 1996; Sérsic et al., 2011) and syntheses (e.g., Médail & Diadema, 2009; Soltis, Morris, McLachlan, Manos, & Soltis, 2006) are beginning to accumulate. The broad taxonomic sampling of such studies leads to inherent difficulties related to the comparison of genetic variation, but model selection and species-specific simulations represent powerful tools to test ecosystem-wide hypotheses (e.g., Carstens et al., 2005; Carstens & Richards, 2007) and make species-specific predictions (e.g., Ruffley et al., 2018; Smith, Ruffley, Tank, Sullivan, & Carstens, 2017), enabling such cross-kingdom comparisons. Along with simulations and model comparison, the use of the predictive approach provides an invaluable strategy for the integration of trait data into comparative phylogeography because such data are often available even in taxa for which no genetic data have yet been collected.

In the current paper, we expand the approach of Espíndola et al. (2016) and incorporate functional traits as predictor variables in order to integrate life history explicitly into predictive phylogeography. The challenge in doing so for a taxonomically broad comparative data set is to identify traits that may influence the geographic structure of genetic variation and are available and scorable across plants and animals. Traits that we integrate here include aspects of diet, life stage at dispersal, mode of dispersal, presence or absence of selfing, reproductive output, and maximum body size. Including these variables improves the predictive ability of the RF predictor and, more importantly, allows us to make inferences regarding the association of functional traits with geographic patterns of genetic variation, thereby demonstrating the applicability of machine learning in biodiversity studies.

2 | MATERIALS AND METHODS

2.1 | Random forest classifier

Of the many supervised machine-learning approaches available, we focus on the ensemble classifier inherent in RF (Breiman, 2001), which uses collections of decision trees to classify observations based on a series of predictor variables. Details of our use of this method are provided in Table 1. First, we use existing phylogeographic studies (see below) as a training data set (i.e., reference taxa). Here, each observation (georeferenced occurrence records for the reference taxa) was classified based on whether it is from a taxon that exhibits or does not exhibit a particular deep phylogeographic break between disjunct populations (i.e., cryptic diversity; step 1, Table 1). That is, the presence or absence of cryptic diversity is the response variable (i.e., the feature we seek to predict). Inferences as to whether or not taxa harboured cryptic diversity were based either on results from Approximate Bayesian Computation (ABC) and BEAST analyses, or on results from previous studies (details in Espíndola et al., 2016).

In step two (Table 1), we use RF in an attempt to use all the information in the data and we grow multiple classifiers (i.e., multiple decision trees) using a bootstrap aggregating (bagging) approach. Two thirds of observations are randomly assigned to be used as input data to construct decision trees; the remaining third is called the out-of-bag (OoB) sample and is used as an internal validation to assess accuracy of the emergent classifier. At each split (i.e., node) in each decision tree, a random subset of the predictor variables is chosen. From among this set of variables, the one that best partitions the observations between classes is selected. This process is repeated and nodes are added to decision trees until all observations are classified (step 2C, Table 1). Whereas individual decision trees may provide poor predictions (i.e., high OoB error rates), the ensemble classifier obtained from a collection (i.e., a forest) of random trees has been shown to be very accurate in many applications. Furthermore, the proportion of decision trees in the forest that classify an observation into a particular category (i.e., the proportion that classify an occurrence record as being from a taxon that harbours cryptic diversity) provides a measure of strength of inference. For more information, an excellent review of random forest classification is provided by Liaw and Wiener (2002).

2.2 | Study systems and predictor variables

2.2.1 | Pacific Northwest of North America

Following Espíndola et al. (2016), we use two model systems. First is the disjunct rainforest of the Pacific Northwest of North America (PNW) (e.g., Brunsfeld, Sullivan, Soltis, & Soltis, 2001; Carstens et al., 2005). The PNW supports the world's greatest extent of temperate coniferous rainforests (Franklin & Dyrness, 1973), and this ecosystem occurs in two distinct bands: one along the Coast and

TABLE 1 The application of random forest to predictive phylogeography

1. Compile a reference data set where the response variable has been assessed (e.g., presence or absence of a deep phylogeographic split)

    (a) Investigate phylogeographic patterns for several codistributed taxa using molecular data and classify each taxon according to a response variable. In this case, whether or not reference taxa are determined to harbour cryptic diversity is the response variable, as evaluated based on the presence/absence of a pre-Pleistocene phylogeographic break. Sequence/genomic data may be generated de novo or mined from existing databases (i.e., Genbank). We use phylogeographic model selection to assess the response variable, in a comparative phylogeographic approach

    (b) Compile occurrence records for all taxa in reference data from public databases (e.g., GBIF, Arctos, etc.), field work, and/or publications. These represent the observations that we wish to classify with respect to the response variable (i.e., as belonging to a taxon that harbours or lacks cryptic diversity)

    (c) For each observation (occurrence record), we obtain and tabulate climate data (e.g., from WorldClim), taxonomic data, and trait data. These are the predictor variables

2. Apply random forest (following Breiman, 2001; Liaw & Wiener, 2002) to build a predictive model (i.e., classifier)

    (a) Randomly select 2/3 of the observations to train the model. This is the in-bag sample

    (b) The remaining random 1/3 of the observations are the OoB sample to be used for internal validation (i.e., used to calculate OoB error rates)

    (c) Build a random decision tree to classify the training set

        (i) From a random subset of predictor variables, select the best predictor variable (i.e., the predictor variable that most effectively splits observations into classes). Classify the training data based on the selected variable. This represents the first node in the decision tree

        (ii) Grow the decision tree by adding predictor variables at random (i.e., new nodes in decision tree) until all observations in the training set are classified correctly

    (d) Repeat steps 2a–2c many times to create a forest of random decision trees

    (e) Classify each OoB sample using the forest of decision trees. This classifies an OoB observation, whose state is known (i.e., from a taxon that has or lacks cryptic diversity) allowing an assessment of OoB error. In addition, the frequency of inference (across the forest of decision trees) for each OoB observation provides an assessment of strength of that inference

    (f) Assess variable importance

3. Cross validate model

    (a) Leave all observations from a particular taxon out of the reference data

    (b) Apply steps 2a–2e to build a classifier without using data for the omitted taxon

    (c) Classify each occurrence record for the omitted taxon to assess accuracy of the classifier

    (d) Repeat steps 3a–3c until each taxon has been used in cross validation

4. Make predictions about the value of the response variable in taxa that have yet to be assessed

Cascades Ranges and a disjunct inland segment in the northern Rocky Mountains of Idaho, Montana, and British Columbia (e.g., DellaSala et al., 2011; see figure 1a in Espíndola et al., 2016). The presence of disjunct populations of several conspecifics or putative sister-species pairs was the focus of the predictive approach developed by Espíndola et al. (2016). The RF classifier they developed had very high accuracy (98.78% overall accuracy) in predicting the presence or absence of cryptic diversity across this disjunction when environmental variables, along with taxonomy as a proxy for life history traits, were used as predictor variables. We include here the same seven taxa and add five other mesic forest disjuncts that have been recently examined: robust lancetooth snails (*Haplotrema vancouverense*) (Smith et al., 2017), three species/complexes of taildropper slugs (*Prophysaon andersoni*, *P. dubium*, and *P. vanattae/P. humile*; Wilke & Duncan, 2004; Smith et al., 2018) and red alders (*Alnus rubra*) (Ruffley et al., 2018). This resulted in a total of three species/species complexes classified as cryptic and nine classified as noncryptic.

Here, we used occurrence data from Espíndola et al. (2016), along with occurrence data from Smith et al. (2018), Smith et al. (2017) and Ruffley et al. (2018). These data are composed of occurrences obtained from the literature, GBIF, and the field. Because GBIF also includes observations that correspond to living collections (e.g., botanical gardens, private gardens, parks), GBIF occurrences were excluded if they fell outside the focal range of the focal taxa (35° to 65° latitude, −160° to −100° longitude). This filter applied mostly to plant localities and was similar to that applied by Espíndola et al. (2016). We downloaded the bioclimatic variables from WORLDCLIM version 2 on 8 June 2018 (Fick & Hijmans, 2017) and we collected trait data from different sources, such as public US Forest Service reports, NatureWeb, and the literature (Supporting Information Table S1). We included the following traits: trophic level (detritivore, herbivore, predator, primary), reproductive strategy (selfing, outcrossing, or both), dispersal stage (adult, juvenile, or embryo), dispersion means (wind, wind/water, and self), and maximum size of the organism.

2.2.2 | Southwest Arid Lands

Espíndola et al. (2016) also examined a set of taxa from comparative phylogeography of the Southwest Arid Lands (SWAL) for the presence or absence of a deep phylogeographic break across the Colorado River, which separates Baja Californian from the Sonoran Desert (see figure 1b in Espíndola et al., 2016). This system has played a central role in the maturation of comparative phylogeography (e.g., Riddle, Hafner, Alexander, & Jaeger, 2000; Zink, 2002), and we follow Espíndola et al. (2016) in using 14 amphibian, bird, and mammal species/complexes as a reference set of taxa for predicting the presence or absence of this deep phylogeographic break. Their initial classifier achieved a moderate prediction accuracy (~69%) with environmental variables and taxonomy used as predictor variables.

As above, occurrence data were from Espíndola et al. (2016), and we downloaded the bioclimatic variables from WORLDCLIM version 2 on 8 June 2018 (Flick & Hijmans, 2017). We did not include

the species *Melozone fuscus* (formerly *Pipilo fuscus*) because, as currently defined, it does not include populations in Baja California. We collected trait data from PANTHERIA (Jones et al., 2009), ELTONTRAITS (Wilman et al., 2014), AMPHIBIAWEB (AmphibiaWeb, 2002), birdsna.org (Rodewald, 2015), the Encyclopedia of Life (EOL), animaldiversity.org (Myers et al., 2018), and allaboutbirds.org (The Cornell Lab of Ornithology, 2018). We included the following traits: diet (five categories, from ELTONTRAITS), nocturnality, body mass, clutch/litter size, and reproductive mode (oviparous or viviparous). Specific information on sources is given in Supporting Information Table S2. Where more than one value was available for a species or complex, the average value was used.

2.2.3 | Pacific Northwest of North America training

We trained eight different RF classifiers, using bioclimatic and trait data as predictor variables and the presence or absence of cryptic diversity as the response variable in the R package RANDOMFOREST (Liaw & Wiener, 2002). The eight classifiers differed in three ways: whether or not the taxon set was curated, whether or not the trait data were curated with respect to their effects on RF accuracy, and whether or not taxonomy was included as a predictor variable. For the first RF classifier, we considered all available data, including taxonomy (All Taxa, All Traits & Taxonomy). For the second RF classifier, we omitted taxonomy (All Taxa, All Traits & No Taxonomy). Because including incorrectly classified taxa is likely to have substantial impacts on the overall accuracy of a classifier, the RF classifiers built from a curated set of taxa may be more robust. Therefore, the third and fourth RF classifiers excluded two species/species complexes: *Chonaphe armata* and the *Prophysaon vanattae/humile* species complex (Curated Taxa & All Traits). These taxa have been difficult to classify as having or lacking cryptic diversity because of either incongruence between morphological and genetic data (Smith et al., 2018), or incongruence between validation approaches using the same genetic data (Espíndola et al., 2016), and therefore may mislead the RF classifier (Espíndola et al., 2016; Smith et al., 2018). The third and fourth RF classifiers were constructed with and without taxonomy, respectively (Curated Taxa, All Traits & Taxonomy; Curated Taxa, All Traits & No Taxonomy). Based on results from the comparisons between the full and curated taxon sets (see below), we only included the subset of the taxa for which we had confidence in the classifications (the curated taxon set) in the remaining classifiers. The final RF classifiers were constructed based on the results of assessments of variable importance. Based on these measures (see below), we chose to remove maximum size from the model, as it tended to mislead the classifier. We constructed these final classifiers both with and without taxonomy as a predictor variable (Curated Taxa, Curated Traits & Taxonomy and Curated Taxa, Curated Traits & No Taxonomy). Finally, we constructed a classifier using only bioclimatic variables and taxonomy (Curated Taxa, No Traits & Taxonomy) to compare to the Final Classifiers.

We assessed the accuracy of these RF classifiers using two approaches following Espíndola et al. (2016): OoB error rates and cross-validation. OoB error rates were calculated using the R package RANDOMFOREST (Liaw & Wiener, 2002). To calculate cross-validation accuracies, we followed Espíndola et al. (2016); we ran separate RF analyses by dropping a single taxon from the reference set and building the classifier without that taxon. Occurrence records for the deleted taxon were then assessed using the classifier built without it and accuracy was calculated as follows: (a) $accuracy_{overall}$ = 100 * ($n_{\text{true cryptic localities}}$ + $n_{\text{true noncryptic localities}}$)/$n_{\text{total predicted localities}}$), (b) $accuracy_{cryptic}$ = 100 * ($n_{\text{true cryptic localities}}$/$n_{\text{total cryptic localities}}$), (c) $accuracy_{noncryptic}$ = 100 * ($n_{\text{true noncryptic localities}}$/$n_{\text{total noncryptic localities}}$).

2.2.4 | Southwest Arid Lands training

We trained four RF classifiers for the SWAL using bioclimatic and trait data as predictor variables and, again, with the presence or absence of cryptic diversity as the response variable in the R package RANDOMFOREST (Liaw & Wiener, 2002). These classifiers differed in whether or not taxonomy was included as a predictor variable and whether or not trait data were curated based on their effects on prediction accuracies. For the first RF classifier, we considered all available data, including taxonomy (All Traits & Taxonomy). For the second RF classifier, we omitted taxonomy (All Traits & No Taxonomy). Two classifiers were then constructed from a reduced set of variables, with two variables with negative impacts on predictive accuracies (see below) removed. We removed body mass and nocturnality, because, in combination, these two variables tended to mislead the classifier (see below). These final two classifiers were constructed both with and without taxonomy as a predictor variable (Curated Traits & Taxonomy; Curated Traits & No Taxonomy). Finally, we trained a classifier with only taxonomy and bioclimatic variables as predictor variables (No Traits & Taxonomy) for comparison. We assessed accuracy of the RF classifiers as above, using both OoB error rates and cross-validation.

2.3 | Assessment of variable importance

We used three metrics to assess variable importance. First, we calculated the Mean Decrease in Accuracy (MDA) and Mean Decrease in Gini (GINI) metrics using the R package RANDOMFOREST (Liaw & Wiener, 2002). MDA is calculated by permuting variable *i*, calculating the OoB error rate, and comparing this error rate to the error rate when variable *i* is included in the model unpermuted. This reveals whether a particular variable increases the accuracy of the classifier and by how much, allowing for variables to be ranked based on their importance in the accuracy of the classifier. GINI measures the gain of node purity in splits for a given variable. Variables that are informative about classification should tend to split observations between, rather than within, classes, and node purity is a measure of whether a node splits between or within classes. The GINI coefficient measures how much each variable contributes to node purity, with lower values associated with higher node purity. Thus, variables with a higher Mean Decrease in Gini contribute more to node purity. Mean decrease in Gini and MDA are both contingent on the training

TABLE 2 Accuracy of Pacific Northwest of North America random forest classifiers

Model	Overall OoB[a] error rate (%)	Overall prediction accuracy[b] (%)	Cryptic prediction accuracy (%)	Noncryptic prediction accuracy (%)
All Taxa, All Traits & Taxonomy	0.000	100.00	100.00	100.00
All Taxa, All Traits & No Taxonomy	0.000	97.0	97.0	97.0
Curated Taxa, All Traits & Taxonomy	0.000	100.00	100.00	100.00
Curated Taxa, All Traits & No taxonomy	0.000	97.7	96.9	98.4
Curated Taxa, Curated Traits & Taxonomy	0.000	100.0	99.9	100.00
Curated Taxa, Curated Traits & No Taxonomy	0.000	98.3	97.0	99.7
Curated Taxa, No Traits & Taxonomy	0.000	100.00	100.00	100.00

[a]OoB error rate. [b]Predition accuracies are calculated as in Espíndola et al. (2016), as the ratio of correctly predicted localities to total predicted localities *100.

data set. For the problem addressed here, multiple observations per species are included in the training data set. When trait data are used, the values are consistent within a species and different among species. By virtue of relating all observations from the same species, a variable may seem important based on MDA or GINI, even if it offers very little information about similarities among species. To address this concern, we also calculated the decrease in prediction accuracy for each variable by repeating the cross-validation step in which individual taxa were omitted from the classifier one at a time. In addition to permuting each taxon, we also permuted each of the trait variables. We calculated the overall accuracies and the cryptic and noncryptic accuracies and then compared these to the base measure of overall accuracy. For each variable, we calculated a mean decrease in prediction accuracies (averaged across taxa) when the variable was removed. These results were used to determine which predictor variables were included in our final RF classifier for each region. For the SWAL, these results generated uncertainty in how to build the most accurate classifier, so we also used a leave-two-out permutation approach, where we removed two variables at a time and measured the effect on prediction accuracy.

3 | RESULTS

3.1 | Classifiers

#### 3.1.1 | Pacific Northwest of North America

Our final PNW data set had 2,734 observations that passed quality filters. The OoB error rates overall and for each class were zero; all OoB observations were correctly classified as cryptic or noncryptic, regardless of which variables and taxa were used to construct the RF classifier (Table 1). When taxonomy and all traits were used, prediction accuracies were 100%, regardless of which taxa were included. Without taxonomy and with all traits, overall accuracy was higher for the curated taxon set than for the full taxon set (97.7% vs. 97.0%; Table 2). Without taxonomy, and with the curated taxon set, cryptic prediction accuracy was 96.9% and noncryptic prediction accuracy was 98.4% using the full set of trait variables (Table 2). Based on this, the remaining analyses were conducted with the curated set of

taxa. For the classifier built with a curated set of traits and a curated set of taxa, but without taxonomy, the overall prediction accuracy was 98.3%, the cryptic prediction accuracy was 97.0%, and the noncryptic prediction accuracy was 99.7%. For the classifier built with a curated set of traits, a curated set of taxa, and taxonomy, prediction accuracies were nearly 100%, and for the classifier built with a curated set of taxa, no traits, and taxonomy, prediction accuracies were 100%.

In the cross-validation analyses based on the curated set of taxa, curated trait data, and taxonomy, all taxa were classified correctly (Figure 1, red densities). To explore how replacing or supplementing taxonomy with trait data affected prediction accuracy, we compared the classifications from this classifier with those from the classifier that included only taxonomy and bioclimatic variables (Figure 1, blue densities) and with those from the classifier that included curated taxa, curated trait data, but no taxonomy (Figure 1, yellow densities) to explore how replacing or supplementing taxonomy with trait data affected prediction accuracy. For all species except *Haplotrema vancouverense*, accuracy did not change substantially. For *H. vancouverense*, prediction accuracy decreased when trait data were used, and decreased further when trait data were used without taxonomy. For this species, taxonomy contains information useful for predictions that is not captured by climatic variables or the traits that we used, and the trait data are misleading when analyzed in the absence of taxonomy. Likewise, for *Microtus richardsoni*, the inclusion of trait data decreases prediction accuracy, but not to the same extent as was observed for *H. vanouverense*.

3.1.2 | Southwest Arid Lands

For the SWAL, 71,832 observations passed quality filters. When all variables were used, and only taxonomy was omitted, OoB error rates were zero. However, overall prediction accuracies were low (40.6%; Table 3). By contrast, omission of nocturnality and body mass (with taxonomy included) resulted in OoB error rates of zero. This improved overall prediction accuracy increased to 95.1%, with the cryptic prediction accuracy increasing to 97.8%, and the noncryptic prediction accuracy increasing to 86.8% (Table 3).
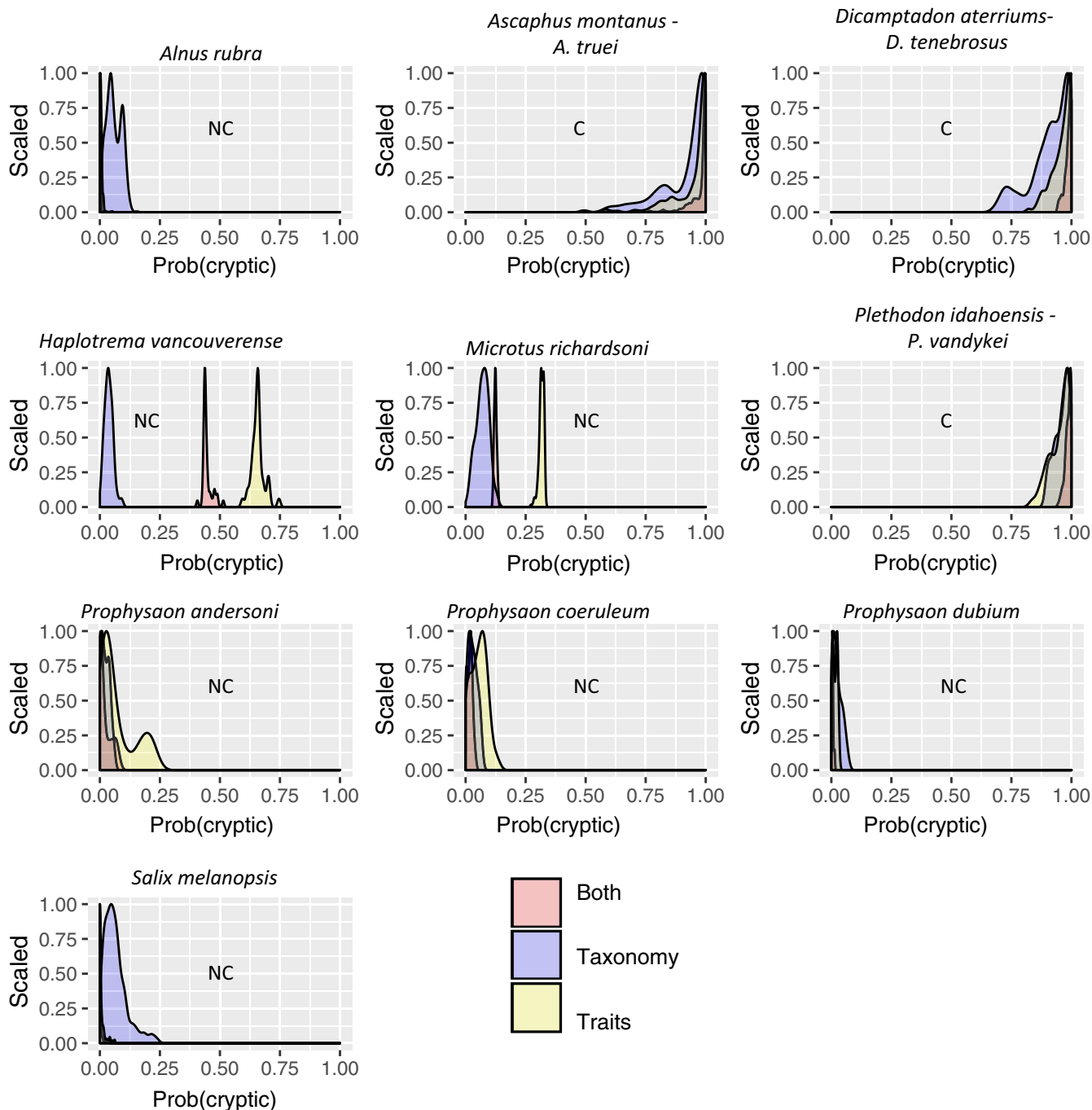
FIGURE 1 Comparisons of accuracies as assessed by cross validations for PNW taxa between the classifiers built using climatic data plus taxonomy (blue), plus curated traits (yellow), and plus both curated traits and taxonomy (red). Classifiers are for the PNW taxa. The *x*-axis is the probability of an observation being cryptic. The *y*-axis is the density of occurrences with a given posterior probability. A "C" on the graph indicates that the taxon harbours cryptic diversity based on genetic data, and a "NC" on the graph indicates that the taxon lacks cryptic diversity based on genetic data

When only the curated traits and bioclimatic variables were used as predictor variables, OoB error rates were zero, overall prediction accuracy was 89.9%, cryptic accuracy was 97.6%, and noncryptic accuracy was 72.0%. When only taxonomy and bioclimatic variables were used (No Traits), OoB error rates were 21%, and prediction accuracies were 62.5% (75.9% for cryptic and 21.4% for noncryptic).

Some differences were apparent in comparisons of classifications made from curated trait data with those from the classifier that included only taxonomy and bioclimatic variables. For two taxa (*Melozone aberti/M. crissalis* and *Oreothlypis ruficapilla/O. luciae*), inclusion of trait data has a nearly neutral effect on prediction accuracy (Figure 2). For five taxa (*Anaxyrus punctatus*, *Callipepla gambelii/C. californica*, *Chaetodipus baileyi*, *Dipodomys merriami*, and *Peromyscus eremicus/P. fraterulus*)

Model	Overall OoB[a] error rate (%)	Overall prediction accuracy[b] (%)	Cryptic prediction accuracy (%)	Noncryptic prediction accuracy (%)
All traits and taxonomy	0.00	40.6	85.7	26.3
All traits and no taxonomy	0.00	40.6	87.1	26.6
Curated traits and taxonomy	0.00	95.1	97.8	86.8
Curated traits and no taxonomy	0.00	89.9	97.6	72.0
No traits	20.8	62.5	75.9	21.4

[a]OoB error rate. [b]Predition accuracies are calculated as in Espíndola et al. (2016), as the ratio of correctly predicted localities to total predicted localities *100.

inclusion of trait data increases certainty in the correct classification. For four other complexes, addition of trait data salvages the classifier from being inaccurate and permits correct prediction. *Campylorhynchus brunnicapilus* and *Toxostoma leconti* are predicted as harbouring cryptic diversity in classifications with only taxonomy (plus climate data), whereas the classifiers with trait data included correctly predict a lack of cryptic diversity (Figure 2). For *Toxosotma cinereum/T. bindirei,* the classifier with taxonomy (plus climatic data) incorrectly predicts a lack of cryptic diversity, whereas inclusion of trait data in the classifiers results in correct prediction of cryptic diversity (Figure 2). For *Auriparus flaviceps,* the correct prediction of absence of cryptic diversity is only attainable with all three data types (climatic, taxonomic, and trait data) included. For *Polioptila melanura/P. californica,* the classifier with trait data led to an incorrect classification, but with taxonomy included, the presence of cryptic diversity is predicted accurately. Finally, for one complex (*Ammospermophilus leucurus*), predictions tended to be inaccurate regardless of whether or not trait data were used (Figure 2).

3.2 | Variable importance

3.2.1 | Pacific Northwest of North America

When taxonomy was included, this variable had the highest measures of MDA and Mean Decrease in Gini, followed by the trait variables (Figure 3). With taxonomy excluded, the trait variables had the highest measures of MDA and GINI (Supporting Information Figure S1). Based on our variable permutation measures of importance, all trait variables except maximum size had nearly neutral or positive effects on the accuracy of the RF classifier (Figure 3a). In the classifier with curated traits, curated taxa, and no taxonomy, reproductive mode (selfing vs. outcrossing) was the most important variable, followed by trophic level and life stage at dispersal as measured by MDA (Figure 3b). However, decrease in GINI indicated that trophic level had the largest effect on node purity, followed by life stage at dispersal, dispersion mode, and reproductive mode (Figure 3c).

3.2.2 | Southwest arid lands

Regardless of whether or not taxonomy was included in the classifier, clutch/litter size was the most important predictor variable based

on MDA and GINI (Figure 4c,d). All other trait variables, except reproductive mode, ranked higher in importance than bioclimatic variables based on MDA and GINI (Supporting Information Figure S2). Based on our variable permutation measures of importance, body mass and nocturnality in combination were strongly misleading the classifier (Figure 4a). Our final leave-out-one permutation measure of variable importance (when body mass and nocturnality were excluded from the model) indicated that reproductive mode and clutch size were valuable predictors, and that diet may have a slightly negative effect on predictive accuracy (Figure 4b).

4 | DISCUSSION
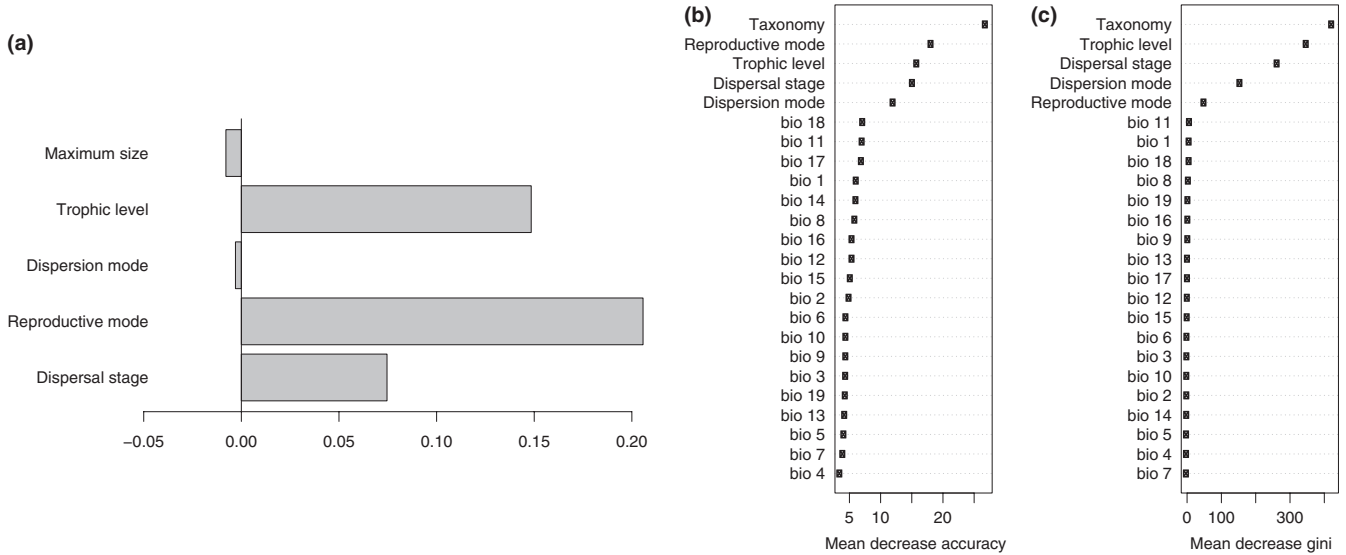
4.1 | Predictions

Incorporating life history traits as predictor variables improves the performance of the supervised machine-learning approach to predictive phylogeography introduced by Espíndola et al. (2016). Our approach, along with the regression model approach introduced by Paz et al., (2015), provides phylogeographers with a framework to integrate ecology and life history into the maturing discipline of comparative phylogeography (e.g., Papadopoulou & Knowles, 2016; Riddle, 2016; Zamudio et al., 2016). Our predictive framework has the potential to expand the importance and impact of comparative phylogeography and its relevance to other disciplines because it enables researchers to use existing databases to make predictions about the structure of genetic variation that may be present in taxa before genetic data are collected. In the two systems examined here, the use of cross validation, where *n* − 1 taxa are used to build the random forest classifier (where *n* is the number of taxa in the reference data set), demonstrates the power of the approach to make predictions about a taxon that was not used in building the classifier (Figures 1 and 2). Furthermore, in using a collection of random decision trees to classify a test observation (in this case an occurrence record), the ensemble approach provides some measure of strength of the prediction in the congruence of the prediction across the forest of decision trees.

For the PNW system, prediction accuracy is high across classifiers and combinations of predictor variables. This is the case both for internal accuracy, as measured by low OoB error rates (Table 1),
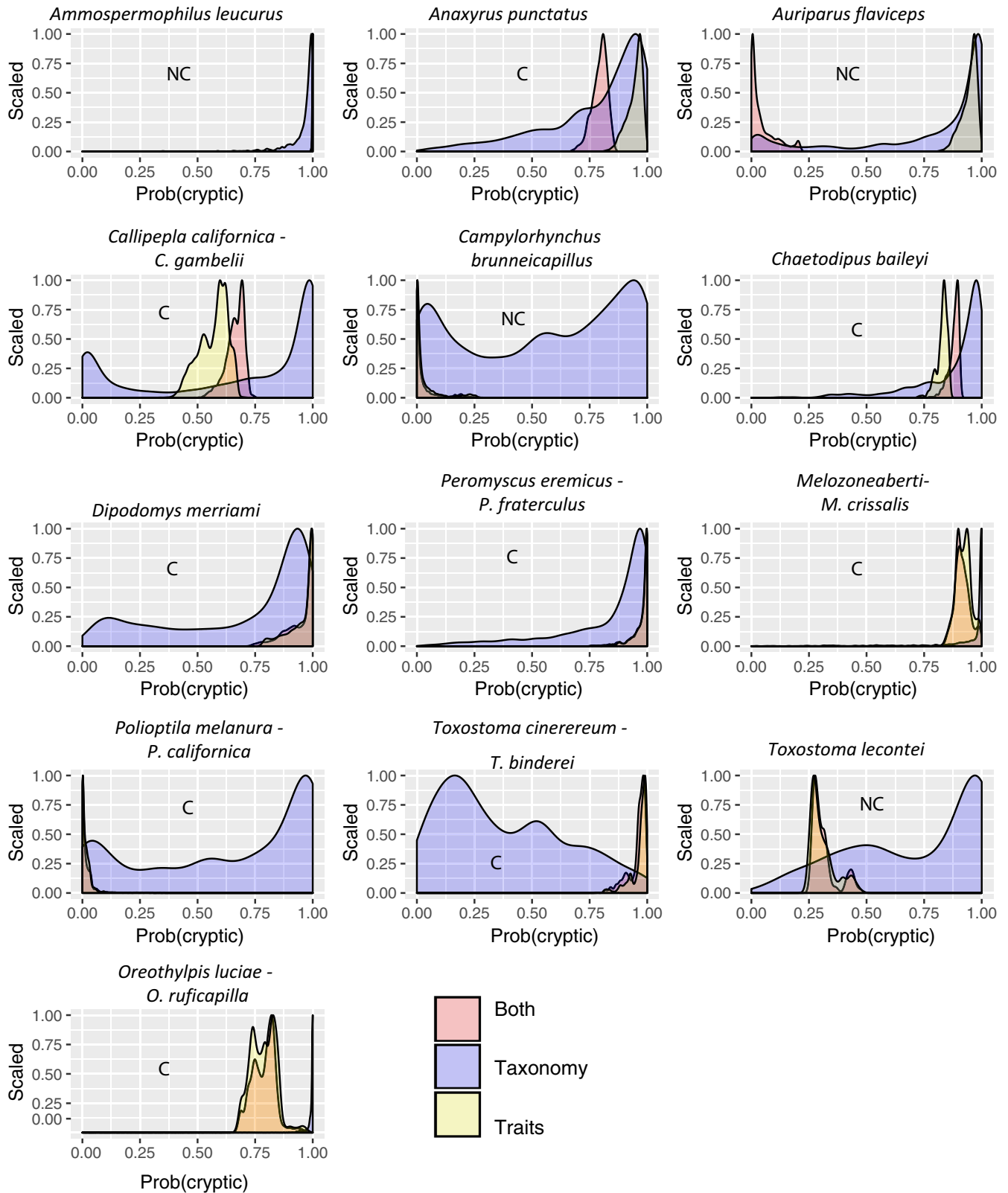
FIGURE 2 Comparisons of accuracies as assessed by cross validations between the classifiers for SWAL taxa built using climatic data plus taxonomy (blue), plus curated traits (yellow), and plus both curated traits and taxonomy (red). Classifiers are for the SAL taxa. The x-axis is the probability of an observation being cryptic. The y-axis is the density of occurrences with a given posterior probability. A "C" on the graph indicates that the taxon harbours cryptic diversity based on genetic data, and a "NC" on the graph indicates that the taxon lacks cryptic diversity based on genetic data

(a)

FIGURE 3 Variable Importance for the PNW. (a) Mean decrease in accuracy based on our permutation approach. Negative values indicate that inclusion of a variable decreases prediction accuracies. (b) Mean decrease in MDA. (c) Mean decrease in Gini coefficient
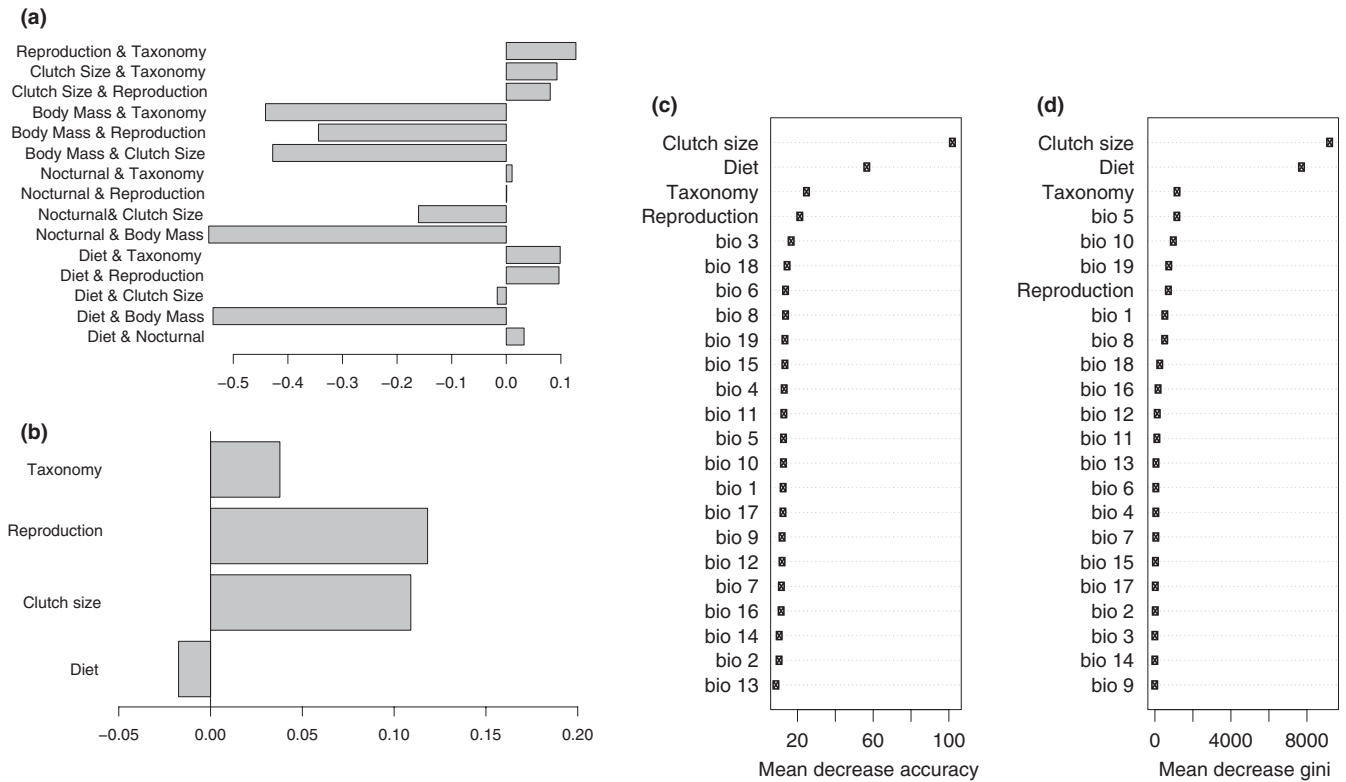
FIGURE 4 Variable Importance for the SWAL. (a) Mean decrease in accuracy based on our leave-two-out permutation approach. Negative values indicate that inclusion of a variable decreases prediction accuracies. (b) Mean decrease in accuracy based on our leave-one-out permutation approach after body mass and nocturnality were excluded from the classifier. Negative values indicate that inclusion of a variable decreases prediction accuracies. (c) Mean decrease in accuracy. (d) Mean decrease in GINI coefficient

and in most of the cross-validation analyses (Figure 1) in which classifiers are built with each taxon successively deleted, and then used to assess accuracy of the classifier. In all taxa, addition of trait data improved the accuracy of the classifier (or was neutral) relative to the original implementation (Espíndola et al., 2016) that used only

taxonomy and climate data. However, while the predatory snail *Haplotrema vancouverense* is still correctly predicted (to lack cryptic diversity) when taxonomy and traits are used in the classifier, it is not correctly predicted when taxonomy is omitted as a predictor variable (Figure 1). This suggests that for this taxon, there is important

information captured by the taxonomy variable that is not captured by the trait data that we used.

The situation was somewhat different for the SWAL system. The original implementation (using only taxonomy plus climate data; Espíndola et al., 2016) had moderate prediction accuracy. By incorporating informative trait data (clutch/litter size, reproductive mode, diet; Figure 4 and Supporting Information Figure S2), we improved overall accuracy considerably (Table 2, Curated Traits classifiers). In the cross-validation analyses (Figure 2), inclusion of all three data types (traits, taxonomy and climatic data) had a positive or neutral impact on predictions for 12 of 13 taxa. Accurate predictions were not attained in any classifier for *Ammospermophilus leucurus* (Figure 2), whereas inclusion of trait data hampered predictions for *Polioptila melanura/P. californica* that were accurate with climate data plus taxonomy (Figure 2). For the *Polioptila* complex, it could be that data for additional traits would improve the classifiers, suggesting that it may be difficult to select trait data a priori.

4.2 | Variable selection and importance

The taxonomic breadth differed considerably in the two systems examined here; the PNW system includes plants, invertebrates, and vertebrates, whereas the SWAL data set includes only tetrapods. As a result, the suite of traits that we examined differed. In the PNW system, the most important traits to include were reproductive mode (selfing vs. outcrossing) and trophic level, with less importance identified for dispersal stage, maximum size, and dispersion mode (Figure 3). The ability to self has long been thought to relate to dispersal ability (Baker's Law; Baker, 1955) and may play an important role in the recolonization of previously glaciated habitat, which is thought to have occurred in PNW rainforest endemics following the Pleistocene glacial cycles (e.g., Carstens et al., 2014). In the taxa studied here, all taxa capable of selfing lacked cryptic diversity and were inferred to have been recent dispersers to the inland rainforests. However, in general, predictions made by Espíndola et al. (2016) using only climate and taxonomic data were quite accurate; there is little room for increasing prediction accuracy by adding trait data. Because of the taxonomic breadth, life history features that are difficult to elucidate across all the taxa, but that may nevertheless be important in structuring genetic diversity, may correlate with taxonomy.

By contrast, in the SWAL system, we identified a single trait, clutch size (Figure 4), as the most important trait (as measured by MDA). Propagule size is one of the variables affecting population size, which in turns affects the probability of survival of populations and species (e.g., Reed, O'Grady, Brook, Ballou, & Frankham, 2003), as well as influencing nucleotide diversity (Romiguier et al., 2014). Although body size is known to relate to dispersal ability in many groups (e.g., beetles, Dingle & Arora, 1973; mammals and birds, Sutherland, Harestad, Price, & Lertzman, 2000), it was misleading in both systems studied here. This may be due to the broad taxonomic scale at which our study was conducted, because the relationship between body size and dispersal ability differs widely across taxonomic groups. For example, although Sutherland et al. (2000)

found a positive relationship between natal dispersal and body size in mammals and carnivorous birds, the same was not true for herbivorous and omnivorous birds.

4.3 | Traits and comparative phylogeography

The identification of factors that influence patterns of biodiversity has been a long-standing goal of biogeography. In addition to distribution and diversity gradients, attention has focused on traits that influence genetic diversity within species, including life history traits in plants (e.g., Loveless & Hamrick, 1984) and propagule size animals (e.g., Romiguier et al., 2014). It has become increasingly clear that trait data, whether from existing databases or collected from natural populations, allow phylogeographers to integrate their work with other disciplines in the biological sciences (Zamudio et al., 2016), and a recent call has been made for phylogeographers to integrate traits into phylogeography in a more meaningful way. However, some authors have been using traits as a priori predictors of pattern for several years. For example, Alvarez et al. (2009) found that the ecological specialization of alpine plants on specific substrate types explained their phylogeographic patterns in the Alps. More recently, Papadopoulou and Knowles (2015) identified an organismal trait, edaphic specialization, and used it to predict which taxa were likely to have responded concordantly to fluctuations in sea level. Others have evaluated a set of organismal trait variables in a post-hoc manner. For example, Paz et al. (2015) collected nine traits from 31 anurans and devised models consisting of combinations of traits that could be compared with a linear regression. In these examples, the authors asked if traits were associated with a particular phylogeographic pattern, either a single trait (Papadopoulou & Knowles, 2015) or several (Alvarez et al., 2009; Paz et al., 2015). However, both of these approaches are limited in the number of traits they can examine. In the approach of Papadopoulou and Knowles (2015), the number of traits that could be considered was limited due to the use of the hierarchical ABC framework to test the prediction of codivergence across a geographic barrier. In Alvarez et al. (2009), the number of traits analyzed was so small that they could apply standard statistical approaches. Paz et al. (2015) investigated a larger set of traits by selecting a set of 28 generalized linear models to compare via multi-model inference techniques. However, these models represent only a small subset of the possible models, which may be a limitation in comparisons across broad taxonomic groups where researchers' intuitions may be misleading and cause important trait variables to be missed. By allowing for the incorporation of a potentially large number of trait variables into phylogeographic studies, machine-learning approaches such as the one used here offer a solution to this problem. They also dramatically expand the scope of comparative phylogeography because it is no longer limited to codistributed taxa.

The use of supervised machine learning to detect variables that predict patterns of genetic variation enables comparative phylogeography to be expanded to continental or global scales. For example, Pelletier and Carstens (2018) examined 8,000 fungi, plants, and animals on a global scale to identify aspects of the species range

(i.e., total size, latitude, elevation) as important predictors of which species contain structured genetic variation. They did this by testing for isolation by distance and environment within each species and building a set of predictor variables that include organismal trait data related to metabolism along with other variables measuring aspects of the species range, habitat, and taxonomy. More detailed organismal trait data can be incorporated by limiting the taxon sampling. For example, Carstens, Morales, Field, and Pelletier (2018) investigated >300 species of bats to identify organismal traits associated with demographic responses to end Pleistocene climate change. Notably, since each of these supervised machine-learning approaches utilize data from public databases such as GBIF, WORLDCLIM, and GENBANK, the continued growth and refinement of these databases will expand the scope of predictive phylogeography.

5 | CONCLUSION: PREDICTIVE APPROACHES IN BIODIVERSITY STUDIES

The inclusion of trait data into supervised machine-learning approaches can improve the accuracy of a classifier that can predict the presence or absence of deep phylogeographic splits (i.e., cryptic divergence) in uninvestigated taxa. We achieved good (in the SWAL system) to excellent (in the PNW) performance with respect to prediction accuracies. In particular, incorporation of trait data improved the accuracies produced by Espíndola et al., (2016), where only taxonomy and climatic data were used, especially for the taxonomically restricted SWAL test data set. Alternatively, one could use such approaches to test the effect of traits that have hypothesized a priori to impinge on the geographic structure of genetic variation (e.g., Alvarez et al., 2009; Paz et al., 2015).

AUTHOR CONTRIBUTION

M.S. and A.E. conducted most of the analyses. J.S. and M.S. wrote the first draft. J.S., B.C.C. and D.C.T. secured funding and conceived of the research program. All authors contributed to the intellectual development of the paper, made multiple revisions, and approved of the final draft.

DATA ACCESSIBILITY

Input files for all random forest analyses are available on Dryad (https://doi.org/10.5061/dryad.s6v210k). Scripts developed as a part of the work presented here are available on github (https://github.com/meganlsmith).

ORCID

Jack Sullivan (iD) https://orcid.org/0000-0003-0216-6867

Bryan Carstens (iD) https://orcid.org/0000-0002-1552-227X

REFERENCES

Alvarez, N., Thiel-Egenter, C., Tribsch, A., Holderegger, R., Manel, S., Schönswetter, P., …IntraBioDiv Consortium. (2009). History or ecology? Substrate type as a major driver of patial genetic structure in alpine plants. *Ecology Letters*, 12(7), 632–640. https://doi.org/10.1111/j.1461-0248.2009.01312.x

AmphibiaWeb. (2002). *AmphibiaWeb: Information on amphibian biology and conservation*. AmphibiaWeb.

Avise, J. C., Bowen, B. W., & Ayala, F. J. (2016). In the light of evolution X: Comparative phylogeography. *Proceedings of the National Academy of Sciences, USA*, 113(29), 7957–7961.

Baker, H. G. (1955). Self-compatibility and establishment after '"long-distance" dispersal. *Evolution*, 9(3), 347–349.

Bermingham, E., & Avise, J. C. (1986). Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics*, 113, 939–965.

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brunsfeld, S. J., Sullivan, J., Soltis, D. E., & Soltis, P. S. (2001). A comparative phylogeography of northwestern North America: A synthesis. In J. Silvertown, & J. Antonovics (Eds.), *Integrating ecology and evolution in a spatial context* (pp. 319–340). Oxford, UK: Blackwell Science.

Carstens, B. C., Brennan, R. S., Chua, V., Duffie, C. V., Harvey, M. G., Koch, R. A., … Sullivan, J. (2014). Model selection as a tool for phylogeographic inference: An example from the willow Salix melanopsis. *Molecular Ecology*, 22, 4014–4028.

Carstens, B. C., Brunsfeld, S. J., Demboski, J. R., Good, J. M., & Sullivan, J. (2005). Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: Hypothesis testing within a comparative phylogeographic framework. *Evolution*, 59, 1639–1652. https://doi.org/10.1554/04-661.1

Carstens, B. C., Morales, A. E., Field, K. A., & Pelletier, T. A. (2018). A global analysis of bats using automated comparative phylogeography uncovers a surprising impact of Pleistocene glaciation. *Journal of Biogeography*, 45, 1795–1805. https://doi.org/10.1111/jbi.13382

Carstens, B. C., & Richards, C. L. (2007). Integrating coalescent and ecological niche modeling in comparative, phylogeography. *Evolution*, 61, 1439–1454.

DellaSala, D. A., Alaback, P., Craighead, L., Goward, T., Paquet, P. C., & Spribille, T. (2011). Temperate and boreal rainforests of inland northwestern North America. In D. A. DellaSala (Ed.), *Temperate and boreal rainforests of the world* (pp. 82–110). Washington, DC: Island Press.

Dingle, H., & Arora, G. (1973). Experimental studies of migration in bugs of the genus Dysdercus. *Oecologia*, 12(2), 119–140. https://doi.org/10.1007/BF00345512

Espíndola, A., Ruffley, M., Smith, M. L., Carstens, B. C., Tank, D. C., & Sullivan, J. (2016). Identifying cryptic diversity with predictive phylogeography. *Proceedings of the Royal Society B: Biological Sciences*, 283(1841), 20161529. https://doi.org/10.1098/rspb.2016.1529

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37, 4302–4315.

Franklin, J. F., & Dyrness, C. T. (1973). Natural Vegetation of Oregon and Washington. General Technical Report PNW-8, USDA Forest Service, Portland, Oregon.

Freilich, X., Anadón, J., Bukala, J., Calderon, O., Charaborty, R., Biossinot, S., & Evolutionary Genetics Class of 2013. (2016). Comparative phylogeography of Ethopian anurans: Impact of the Great Rift Valley and Pleistocene climat change. *BMC Evolutionary Biology*, 16, 206.

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58, 247–276. https://doi.org/10.1111/j.1095-8312.1996.tb01434.x

Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., ... Purvis, A. (2009). PanTHERIA: A species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, 90(9), 2648–3000. https://doi.org/10.1890/08-1494.1

Kirchman, J. J., & Franklin, J. D. (2007). Comparative phylogeography and genetic structure of Vanuatu birds: Contro region variation in rail, a dove, and a passerine. *Molecular Phylogenetics and Evolution*, 43, 14–23.

Liaw, A., & Wiener, M. (2002). Classification and regression by random-Forest. *R News*, 2(3), 18–22.

Loveless, M. D., & Hamrick, J. L. (1984). Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*, 15, 65–95. https://doi.org/10.1146/annurev.es.15.110184.000433

Médail, F., & Diadema, K. (2009). Glacial refugia influence plant diversity in the Mediterranean Basin. *Journal of Biogeography*, 36(7), 1333–1345.

Myers, P., Espinosa, R., Parr, C. S., Jones, T., Hammond, G. S., & Dewey, T. A. (2018). The Animal Diversity Web (online). Accessed at https://animaldiversity.org. Accessed 10 June 2018.

Papadopoulou, A., & Knowles, L. L. (2015). Species-specific responses to island connectivity cycles: Refined models for testing phylogeographic concordance across a Mediterranean Pleistocene aggregate island complex. *Molecular Ecology*, 24(14), 4252–4268. https://doi.org/10.1111/mec.13305

Papadopoulou, A., & Knowles, L. L. (2016). Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proceeding of the National Academy of Sciences, USA*, 113(29), 8018–8024. https://doi.org/10.1073/pnas.1601069113

Paz, A., Ibáñez, R., Lips, K. R., & Crawford, A. J. (2015). Testing the role of ecology and life history in structuring genetic variation across a landscape: A trait-based phylogeographic approach. *Molecular Ecology*, 24(14), 3723–3737. https://doi.org/10.1111/mec.13275

Pelletier, T. A., & Carstens, B. C. (2018). Geographical range size and latitude predict population genetic structure in a global survey. *Biology Letters*, 14, 20170566. https://doi.org/10.1098/rsbl.2017.0566

Pelletier, T. A., Carstens, B. C., Tank, D. C., Sullivan, J., & Espíndola, A. (2018). redicting plant conservation priorities on a global scale. Proceedings of the National Academy of Sciences, USA. 115, 13027–13032.

Reed, D. H., O'Grady, J. J., Brook, B. W., Ballou, J. D., & Frankham, R. (2003). Estimates of minimum viable population sizes for vertebrates and factors influencing those estimates. *Biological Conservation*, 113, 23–34. https://doi.org/10.1016/S0006-3207(02)00346-4

Riddle, B. R. (2016). Comparative phylogeraphy clarifies the complexity and problems of continental distribution that drove A. R. Wallace to favor islands. *Proceedings of the National Academy of Sciences, USA*, 113(29), 7970–7977.

Riddle, B. R., Hafner, D. J., Alexander, L. F., & Jaeger, J. R. (2000). Cryptic vicariance in the historical assembly of a Baja California penninsular desert biota. *Proceeding of the National Academy of Sciences, USA*, 97(26), 14438–14443.

Rodewald, P. (2015). The Birds of North America. https://birdsna.org. Cornell Laboratory of Ornithology, Ithaca. New York.

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., ... Galtier, N. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, 515(7526), 261–263. https://doi.org/10.1038/nature13685

Ruffley, M., Smith, M. L., Espíndola, A., Carstens, B. C., Sullivan, J., & Tank, D. C. (2018). Combining allele frequency and tree-based approaches improves phylogeographic inference from natural history collections. *Molecular Ecology*, 27(4), 1012–1024. https://doi.org/10.1111/mec.14491

Sérsic, A. N., Cosacov, A., Cocucci, A. A., Johnson, L. A., Pozner, R., Avila, L. J., ... Morando, M. (2011). Emerging phylogeographical patterns of plants and terrestrial vertebrates from Patagonia. *Biological Journal of the Linnean Society*, 103(2), 475–494. https://doi.org/10.1111/j.1095-8312.2011.01656.x

Smith, M. L., Ruffley, M., Rankin, A. M., Espíndola, A., Tank, D. C., Sullivan, J., & Carstens, B. C. (2018). Testing for the presence of cryptic diversity in tail-dropper slugs (Prophysaon) using molecular data. *Biological Journal of the Linnean Society*, 124(3), 518–532. https://doi.org/10.1093/biolinnean/bly067

Smith, M. L., Ruffley, M., Tank, D. C., Sullivan, J., & Carstens, B. C. (2017). Demographic model selection using random forests and the site frequency spectrum. *Molecular Ecology*, 26(17), 4562–4573. https://doi.org/10.1111/mec.1422

Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., & Soltis, P. S. (2006). Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, 15(14), 4261–4293. https://doi.org/10.1111/j.1365-294X.2006.03061.x

Sutherland, G. D., Harestad, A. S., Price, K., & Lertzman, K. P. (2000). Scaling of natal dispersal distances in terrestrial birds and mammals. *Conservation Ecology*, 4(1), 16–000. https://doi.org/10.5751/ES-00184-040116

The Cornell Lab of Ornithology. (2018). Available from http://www.birds.cornell.edu/netcommunity/page.aspx?pxml:id=1636. Accessed 10 June 2018.

Wilke, T., & Duncan, N. (2004). Phylogeographical patterns in the American Pacific Northwest: Lessons from the arionid slug *Prophysaon coeruleum*. *Molecular Ecology*, 13, 2303–2315. https://doi.org/10.1111/j.1365-294X.2004.02234.x

Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M. M., & Jetz, W. (2014). EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals: Ecological Archives E095–178. *Ecology*, 95(7), 2027.

Zamudio, K. R., Bell, R. C., & Mason, N. A. (2016). Phenotypes in phylogeography: Species' traits, environmental variation and vertebrate diversification. *Proceeding of the National Academy of Sciences, USA*, 113(29), 8041–8088. https://doi.org/10.1073/pnas.1602237113

Zink, R. M. (2002). Methods in comparative phylogeography, and their application to studying evolution in the North American aridlands. *Integrative and Comparative Biology*, 42, 953–959. https://doi.org/10.1093/icb/42.5.953

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.