

# Predicting distributions, habitat preferences and associated conservation implications for a genus of rare fishes, seahorses (*Hippocampus* spp.)

Xiong Zhang<sup>1,2</sup>  | Amanda C.J. Vincent<sup>1,2</sup>

<sup>1</sup>Project Seahorse, Institute for the Oceans and Fisheries, The University of British Columbia, Vancouver, BC, Canada

<sup>2</sup>Department of Zoology, The University of British Columbia, Vancouver, BC, Canada

## Correspondence

Xiong Zhang, Project Seahorse, Institute for the Oceans and Fisheries, The University of British Columbia, Vancouver, BC, Canada.  
Email: x.zhang@oceans.ubc.ca

## Funding information

China Scholarship Council; Guylian Chocolates (Belgium)

Editor: Maria Beger

## Abstract

**Aim:** To identify useful sources of species data and appropriate habitat variables for species distribution modelling on rare species, with seahorses as an example, deriving ecological knowledge and spatially explicit maps to advance global seahorse conservation.

**Location:** The shallow seas.

**Methods:** We applied a typical species distribution model (SDM), maximum entropy, to examine the utility of (1) two versions of habitat variables (habitat occurrences vs. proximity to habitats) and (2) three sources of species data: quality research-grade (RG) data, quality-unknown citizen science (CS) and museum-collection (MC) data. We used the best combinations of species data and habitat variables to predict distributions and estimate species–habitat relations and threatened status for seahorse species.

**Results:** We demonstrated that using “proximity to habitats” and integrating all species datasets (RG, CS and MC) derived models with the highest accuracies among all dataset variations. Based on this finding, we derived reliable models for 33 species. Our models suggested that only 0.4% of potential seahorse range was suitable to more than three species together; seahorse biogeographic epicentres were mainly in the Philippines; and proximity to sponges was an important habitat variable. We found that 12 “Data Deficient” species might be threatened based on our predictions according to IUCN criteria.

**Main conclusions:** We highlight that using proper habitat variables (e.g., proximity to habitats) is critical to determine distributions and key habitats for low-mobility animals; collating and integrating quality-unknown occurrences (e.g., CS and MC) with quality research data are meaningful for building SDMs for rare species. We encourage the application of SDMs to estimate area of occupancy for rare organisms to facilitate their conservation status assessment.

## KEYWORDS

citizen science, data quality, Maxent, museum collections, rare species, seahorse, species distribution model

## 1 | INTRODUCTION

Understanding global distribution and habitat preference of rare animals is key to ecology and wildlife conservation (Brooks et al., 2002; Hanski, 2011). Over the past few centuries, anthropogenic activities have caused astonishing biodiversity loss, with a detrimental impact on human society (Hooper et al., 2012). Rare species, with low densities or small ranges, are more sensitive to human disturbance and usually have higher extinction risks than common species (Reynolds, Dulvy, Goodwin, & Hutchings, 2005). Estimating threatened status of rare species usually requires a good knowledge about their geographical distributions (Gaston & Fuller, 2009). Their distribution maps are also helpful for identifying global biodiversity hotspots (high species-richness sites that are threatened), where limited conservation resources should be allocated (Mittermeier, Turner, Larsen, Brooks, & Gascon, 2011; Myers, Mittermeier, Mittermeier, Da Fonseca, & Kent, 2000). In addition to distribution maps, habitat-preference knowledge is essential to find species at local scales, where conservation actions are usually taken (Harris, Jenkins, & Pimm, 2005).

Species distribution models (SDMs) are useful for analysing species distributions and habitat preferences for rare species conservation (Franklin, 2010; Marcer, Sáez, Molowny-Horas, Pons, & Pino, 2013). These techniques are built on a variety of algorithms that correlate species occurrences and ecological covariates (i.e., model predictors) based on the “ecological niche” concept (Whittaker, Levin, & Root, 1973). This concept suggests that species choose their habitats based on their fitness to various surrounding factors. The model approach includes quantitative description of the relationship between species occurrences and ecological covariates. By modelling this relationship, biologists can predict species distributions in unsurveyed regions and generate a global view of species distribution patterns (Franklin, 2010). Moreover, integrating habitat variables in the model allows identification of critical habitats based on the parameter estimation function. The spatially explicit maps derived from SDMs can be used to estimate area of occupancy (AOO), which is a critical measure for assessing threatened status (IUCN Standards and Petitions Subcommittee, 2017).

Identifying useful species data and habitat-related predictors is essential for building robust SDMs for rare species (Aubry, Raley, & McKelvey, 2017). Given the difficulty in studying rare organisms, scientists usually need to collate distribution data from quality-unknown sources (e.g., citizen science) to supplement the limited high-accuracy data (e.g., research data). Applying unknown-quality data can be problematic and might, through errors, generate different results than quality data (Aubry et al., 2017; Graham, Ferrier, Huetteman, Moritz, & Peterson, 2004). Studies comparing different data sources are rare in the literature (Jackson, Gergel, & Martin, 2015; Zhang & Vincent, 2017). Furthermore, including appropriate habitat covariates in SDMs can be critical for rare species whose presences are correlated with important resources (e.g., food, shelters) within a particular habitat (Rainho & Palmeirim, 2011). These issues above can degrade model accuracy and distort species-habitat relationships if they are not addressed appropriately (Aubry et al., 2017).

Seahorses are rare animals, whose conservation is of global concern (Vincent, Foster, & Koldewey, 2011). Seahorses are a genus (*Hippocampus* Rafinesque, 1810) of small, cryptic and sedentary marine fishes in the family Syngnathidae, well known for their male pregnancy and charismatic appearance. They are usually found at low population densities (0–0.51 individuals/m<sup>2</sup>; Foster & Vincent, 2004). Because of this and other biological traits (e.g., low fecundity, extensive paternal care and often, mate fidelity), seahorses are vulnerable to various human activities, especially poorly managed fisheries and habitat degradation (Foster & Vincent, 2004). Wild seahorses are often caught in fisheries and traded worldwide, mainly for traditional medicines. In 2002, seahorses became the first marine fishes to be listed on Appendix II of the Convention on International Trade in Endangered Species (CITES). This listing mandates 183 Parties to CITES to ensure that their exports do not threaten wild seahorse populations. The recent estimate of annual seahorse catches in 22 countries, totalling at least 37 million individuals (Lawson, Foster, & Vincent, 2017), emphasizes the importance of global actions to conserve these rare species.

Global distribution, habitat preference and conservation status are poorly known for seahorses. Studies of seahorse ecology have only centred on a small fraction of species (Cohen, Valenti, Planas, & Calado, 2017). Available species range maps are not spatially explicit enough to inform global and local conservation actions (Lourie, Pollom, & Foster, 2016). Comprehensive habitat knowledge is lacking for most species, although site-level or regional-scale habitat studies are available for some populations (Aylesworth et al., 2015; Caldwell & Vincent, 2013; Harasti, Martin-Smith, & Gladstone, 2014; Zhang & Vincent, 2017). Currently, 14 species are considered as threatened (Endangered or Vulnerable), but 17 species are still “Data Deficient” according to the latest IUCN Red List assessment (IUCN, 2017; Pollom et al., in preparation).

Our study aims to apply SDMs to advance global conservation for all seahorse species. We are interested in examining the utility of quality-unknown species data and identifying proper habitat variables to build robust SDMs for rare species. In undertaking this work, we collated species-level seahorse occurrences and ecological covariates including habitat variables. We then identified the best species datasets and habitat variables deriving the most accurate models. We used these models to predict seahorse distributions, to understand species-habitat preferences and to inform their conservation status assessment.

## 2 | METHODS

### 2.1 | Species distribution model

We used maximum entropy (Maxent; Phillips et al., 2004) to build species distribution models. Maxent is one of the most powerful and popular SDMs (Hernandez, Graham, Master, & Albert, 2006; Phillips & Dudík, 2008). It estimates presence probability by finding a distribution with maximum entropy (i.e., closest to uniform), subject to constraints defined by conditions at known occurrence locations

(Phillips et al., 2004). Earlier research indicated that Maxent is robust to positional uncertainty/errors in species occurrences (Fernandez et al., 2009; Graham et al., 2008) and particularly suitable for rare species with limited occurrences (Pearson, Raxworthy, Nakamura, & Townsend Peterson, 2007). We developed our models with the latest Maxent software (Version 3.4.1, Phillips et al., 2017).

## 2.2 | Study species

Our study focused on 42 valid species of the genus *Hippocampus* (Lourie et al., 2016; Zhang, Qin, Wang, & Lin, 2016). Seahorses were typically found in shallow waters (depth < 200 m) from tropic to temperate zones. They are cryptic, sedentary, small fishes (usually body height < 35 cm) that ambush zooplankton and benthic organisms (e.g., Crustacea and Amphipoda; Manning, 2017). Their predators are believed rare, although they were occasionally found in diets of various marine species such as larger fishes and birds (Kleiber, Blight, Caldwell, & Vincent, 2011). Identification of seahorses at the species level is sometimes challenging because of morphological similarity across species and individual phenotypic

plasticity within species (Curtis, 2006; Roos, Van Wassenbergh, Aerts, Herrel, & Adriaens, 2011). We used work by Lourie et al. (2016), which clarified seahorse nomenclature and revealed many synonyms, as the basis for modelling species-level distributions in our study.

## 2.3 | Data collection

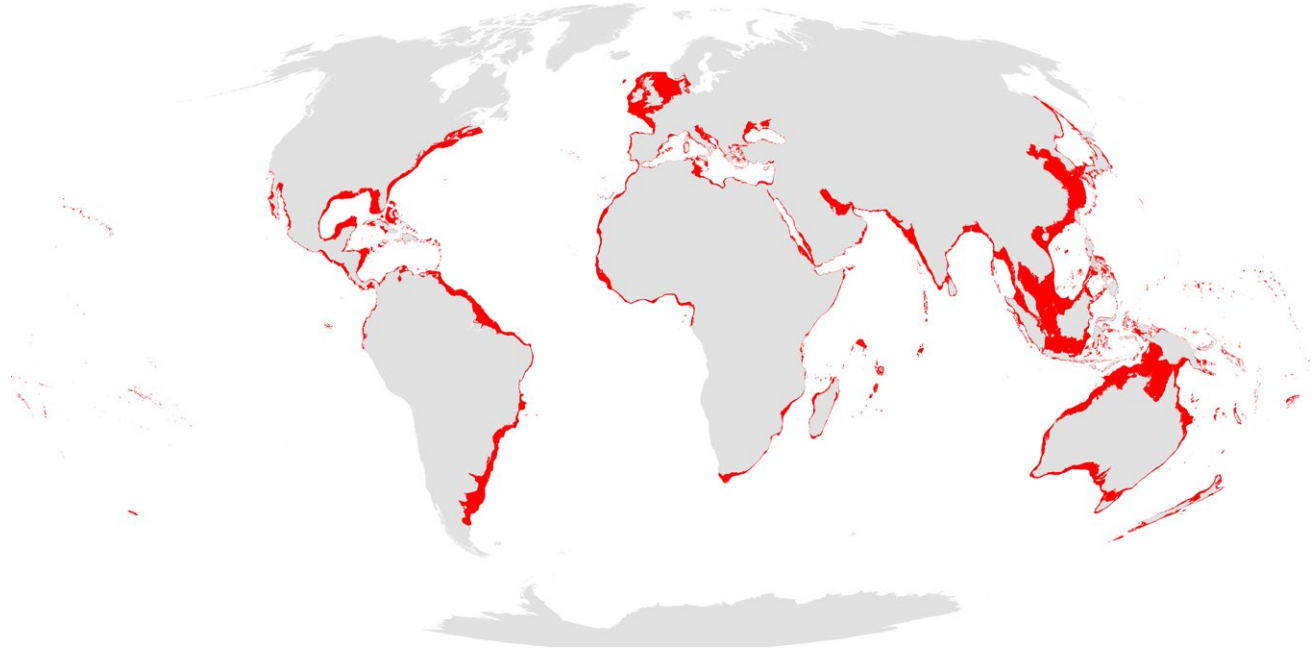
We gathered data for model predictors that have ecological relevance and available information. We first derived data for 12 variables related to seahorse physiological suitability and primary productivity (Foster & Vincent, 2004; Table 1). We selected seven variables from these twelve covariates to minimize collinearity based on Pearson correlation tests ( $|r| > 0.7$ , Dormann et al., 2013). The resulting seven predictors were depth, pH, salinity, sea surface temperature mean and range, and chlorophyll a mean and range.

We then derived datasets of nine categories of potentially key habitats (e.g., coral reefs) from online biogeographic databases (Table 1, Table S1.1, Appendix S1) and generated a binary and a continuous variable for each habitat category. The binary variable was

**TABLE 1** Original data resolutions and sources for seahorse ecological variables. “\*” indicates the seven selected predictors from the original twelve covariates of physiological suitability and primary productivity

Categories	Parameters	Resolutions (arc-degree)	Data sources
Physiological suitability	depth*	0.00833	Weatherall et al. (2015)
	pH*	0.0833	Tyberghein et al. (2012)
	salinity*	0.0833	Tyberghein et al. (2012)
	dissolved oxygen	0.0833	Tyberghein et al. (2012)
	SST mean*	0.0833	Tyberghein et al. (2012)
	SST range*	0.0833	Tyberghein et al. (2012)
	SST maximum	0.0833	Tyberghein et al. (2012)
Primary productivity	SST minimum	0.0833	Tyberghein et al. (2012)
	Chlorophyll a (mean)*	0.0833	Tyberghein et al. (2012)
	Chlorophyll a (range)*	0.0833	Tyberghein et al. (2012)
	Chlorophyll a (maximum)	0.0833	Tyberghein et al. (2012)
Habitat availability	Chlorophyll a (minimum)	0.0833	Tyberghein et al. (2012)
	coral reefs	0.0003–0.01	UNEP-WCMC et al. (2010)
	seagrass beds	0.0026	UNEP-WCMC and Short FT (2017)
	mangroves	0.0003	Giri et al. (2011)
	estuary	–	Alder 2003
	soft corals	–	GBIF & OBIS
	seaweed	–	GBIF & OBIS
	sponge	–	GBIF & OBIS
	sea pen	–	GBIF & OBIS
hydrozoa	–	GBIF & OBIS	

GBIF, Global Biodiversity Information Facility; OBIS, Ocean Biogeographic Information System; SST, sea surface temperature.



**FIGURE 1** A map of the potential distribution ranges of all seahorse species combined that we used in the model to constrain model prediction

habitat presence/absence (1/0), and the continuous variable was the distance to the nearest location of each habitat. We generate all these model predictors as global maps with a resolution of  $1 \text{ km}^2$  (cylindrical equal-area projection), constrained within the 200-m depth range in ArcMap (version 10.2.2). We predicted that continuous habitat variables could be more useful than binary ones, as the former would be resilient to some extent of mismatch between species occurrences and habitat locations.

We collected species-level (presence-only) locations of seahorses from multiple databases including online biogeographic databases, published peer-reviewed literature, unpublished research data (from Project Seahorse) and iSeahorse ([www.iseahorse.org](http://www.iseahorse.org))—a global citizen-science platform for gathering seahorse occurrences. We collated the species location data by checking nomenclature and spatial errors. Based on the collated georeferenced species occurrences, we determined the potential geographical range (i.e., modelling envelope, Figure 1) for each species to constrain model prediction. We predicted that the quality of multiple sources of species data might differ. Therefore, we divided all occurrences of each species into three commonly used categories: research grade, citizen science and museum collection.

A more detailed description of this section is in Appendix S1.

## 2.4 | Model description and evaluation

We executed two groups of model variations (Table 2) to examine our datasets. To do so, we created different predictor datasets and species datasets. First, we developed three different predictor datasets which represented the three methods of habitat data usage in the model (none, binary and continuous; Table 2): Dataset 1—seven

selected predictors reflecting species' physiological suitability and primary productivity (Table 1); Dataset 2—combination of Dataset 1 and nine *binary* habitat variables (see Section 2.3); and Dataset 3—combination of Dataset 1 and nine *continuous* habitat variables (see Section 2.3). Secondly, we generated seven different datasets of species occurrences (Table 2): (1) research grade (RG), (2) citizen science (CS), (3) museum collections (MC), (4) CS plus MC, (5) RG plus CS, (6) RG plus MC and (7) ALL (i.e., RG plus CS plus MC).

*Model Group 1:* To examine the utility of habitat variables, we built models for each species based on the species dataset ALL and each of the three predictor datasets (Datasets 1–3). We then compared the three models across species using three statistics: area under the curve of the receiver operating characteristic (AUC; Hanley & McNeil, 1982), habitat variable importance (HVI) and predicted area ratio (PAR). The AUC is a standard technique for measuring models' omission and commission errors, producing a score (0–1) for general predictive accuracy. Score “1” means no errors of commission or omission, and “ $\text{AUC} \leq 0.5$ ” means the model prediction is no better than random selection. We defined HVI as the cumulative permutation importance of all habitat variables and used it as an index of species-habitat dependency. The PAR is the proportion of the predicted area to the modelling envelope of each species (see Section 2.3). Both HVI and PAR can be easily derived from Maxent's summary file (i.e., “maxentResults.csv”).

*Model Group 2:* To examine the utility of different sources of species data, we conducted models for each species based on the “best performing” predictor dataset (derived from Model Group 1) and each of the seven datasets of seahorse occurrences: RG, CS, MC, CS plus MC, RG plus CS, RG plus MC and ALL. We compared the accuracy in predicting presences (i.e., sensitivity; Altman & Bland, 1994) among the models and estimated prediction agreement between models of

**TABLE 2** Description of model variations used to test the utility of different types of macrohabitat variables (Group 1) and different sources of seahorse data (Group 2). Note that one type of model variation (based on all occurrences and predictor Dataset 3) was used in both groups

Model group	Species dataset	Predictor dataset	Model statistics
Group 1: Testing the utility of different habitat variables (3 models per species).	All occurrences	Each model uses one of the three types:  Dataset 1: seven selected variables; Dataset 2: Dataset 1 + nine binary macrohabitat variables; Dataset 3: Dataset 1 + nine distance-based macrohabitat variables	Area under the curve (AUC), Habitat variable importance (HVI), Predicted area ratio (PAR)
Group 2: Testing the utility of different sources of species data (seven models per species)	Each model uses one of the six types: Research grade (RG), citizen science (CS), museum collection (MC), CS + MC, RG + CS, RG + MC and all occurrences (ALL)	Dataset 3 as described above	Sensitivity, I Similarity Statistic, Presence Agreement

individual species-data sources (RG, CS and MC) based on I Similarity Statistic (Warren, Glor, & Turelli, 2008) and Presence Agreement (Zhang & Vincent, 2017). The former was used to estimate similarity between two presence-probability maps, while the latter evaluated overlapping rate between two predictive-presence maps.

To ensure robust statistical analyses, we only built models on species with a total sample size  $\geq 50$  in Group 1 and executed models on species with a sample size  $\geq 30$  from each source (i.e., RG, CS and MC) in Group 2. Detailed methods of Section 2.4 can be found in Appendix S2.

## 2.5 | Model prediction and interpretation

We applied Maxent to generate the presence-probability map for every species that had at least five occurrences (Pearson et al., 2007), based on the best predictor and species datasets identified above. To estimate species richness, we stacked the presence-probability maps of all species that had fair to excellent model performance (i.e.,  $AUC \geq 0.7$ ; Calabrese, Certain, Kraan, & Dormann, 2014; Zhang & Vincent, 2017). For species that we were unable to derive acceptable predictive maps or that had few occurrences ( $n < 5$ ), we added their collated occurrences directly to corresponding pixels of the stacked map in ArcMap.

We applied Maxent's variable permutation importance and marginal response curve to interpret species-habitat relationships (Searcy & Shaffer, 2016; Stirling, Boulcott, Scott, & Wright, 2016). To this end, we first converted permutation importance values into ranks, with rank 1 assigned to the largest value (Appendix S2). We then used the mean rank (of each variable across species) as an index of the general importance of that variable in determining seahorse distributions. We also calculated habitat variable importance (HVI) for each seahorse species. We identified habitats that were positively correlated with species presence probability through examining the marginal response curves and ranked them by permutation

importance. This allowed us to determine the rank of importance of each habitat for each species.

## 2.6 | Identifying potentially threatened species

With the above distribution information, we calculated and compared geographical metrics (extent of occurrence and area of occupancy) against the IUCN threatened thresholds for geographical range (Criteria B and D), and number of locations (Criterion D) to identify potentially threatened species, following the latest IUCN Red List Guidelines (IUCN Standards and Petitions Subcommittee, 2017; Appendix S2).

## 3 | RESULTS

We obtained a total of 6,128 unique occurrences (collated from the original 6,316 occurrences) for 42 species, with a wide range of sample size (1–1,990 occurrences per species). The temporal range spanned from 1828 to 2016 (96% from 1950 to 2016). Most occurrences had no location-precision information, except 448 records (precision =  $727 \pm 609$  m). Most seahorse occurrences were sourced from MC (75%), followed by RG (18%) and CS (7%). In general, seahorses were found in a very wide geographical range (46.73°S to 54.75°N, 160.62°W to 179.12°E) and diverse environmental spaces (Table S3.1 in Appendix S3). According to the sample-size requirement to build robust models (see Section 2.4), a total of 16 species were used in Model Group 1, and two species (*Hippocampus erectus* and *Hippocampus kuda*) were examined in Model Group 2.

### 3.1 | Group 1: Utility of habitat predictors

Our results indicated that including habitat variables had statistically significant improvements on model predictive accuracy, with

models using continuous habitat predictors (Dataset 3) having the best performance (Figure 2a). The AUC values (model performance) of Dataset 3 and Dataset 2 were significantly higher than those of Dataset 1 ( $n = 16$ , paired Wilcoxon tests; Dataset 3 vs. 1,  $p < .005$ ; Dataset 2 vs. 1,  $p < .01$ ), and Dataset 3 derived higher performance than Dataset 2 ( $n = 16$ , paired Wilcoxon test,  $p < .005$ ).

We did not find statistically significant difference between predicted area ratios estimated from different predictor datasets ( $n = 16$ , paired Wilcoxon tests,  $p = .12, .12, .46$ ). But models based on Dataset 3 consistently derived the smallest predicted areas (Figure 2b).

We found that Dataset 3 was more informative given it revealed the importance of habitats to seahorse distributions (Figure 2c). On average, models using Dataset 3 derived high values of habitat variable importance ( $46.2 \pm 20.3$ ), which were significantly larger than those using binary counterparts in Dataset 2 ( $9.2 \pm 8.9$ ;  $n = 16$ , paired Wilcoxon test,  $p < .001$ ).

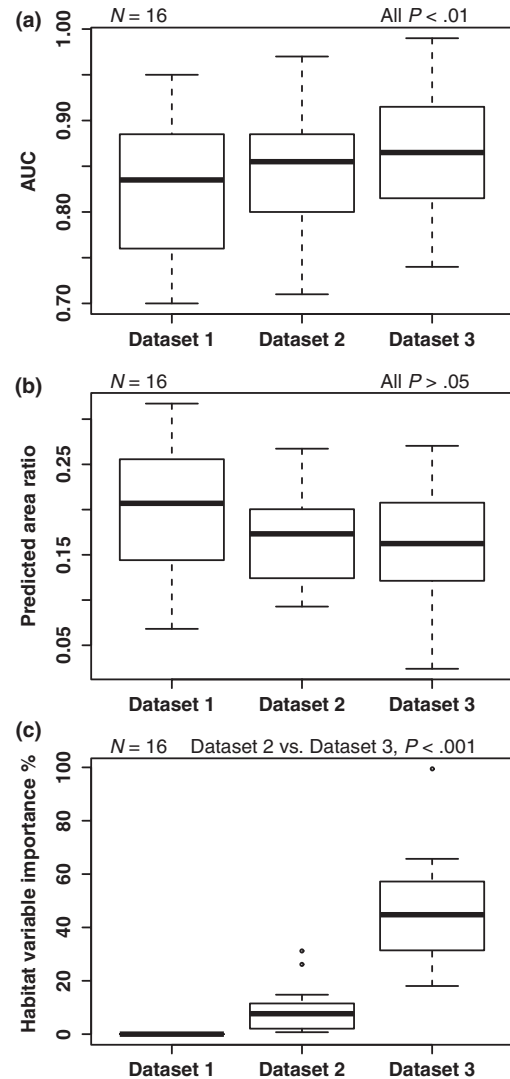
### 3.2 | Group 2: Utility of different sources of species data

Our comparisons indicated that CS and MC data derived less accurate models than RG data did and integrating all three derive the best models (Figure 3a). We found that RG consistently derived the most accurate models (0.55 on *H. erectus* and 0.83 on *H. kuda*) among the individual datasets (Figure 3a). When individual datasets were combined, adding MC or CS to RG consistently improved model performance (Figure 3a). Moreover, using all occurrences derived the most accurate models (0.69 and 0.94, respectively; Figure 3a). Interestingly, while “CS plus MC” derived a model with the 2nd highest accuracy on *H. erectus* (0.66), it resulted in the lowest-accurate model on *H. kuda* (0.27; Figure 3a), suggesting that the reliability of this type of combination varies between species.

There were moderate to high consistencies (I Similarity = 0.68–0.95) in predicting presence probability, but low agreements in predictive presences (PA = 2%–34%) between different species-data sources (Figure 3b,c). The I Similarity Statistics (i.e., presence-probability similarity) suggested that the agreement between RG models and CS models (0.71 and 0.93, respectively) was slightly higher than those between RG models and MC models (0.68 and 0.91, respectively; Figure 3b). This difference was more prominent in the measure of Presence Agreement (i.e., overlapping rate, Figure 3c).

### 3.3 | Seahorse global distributions and biodiversity epicentres

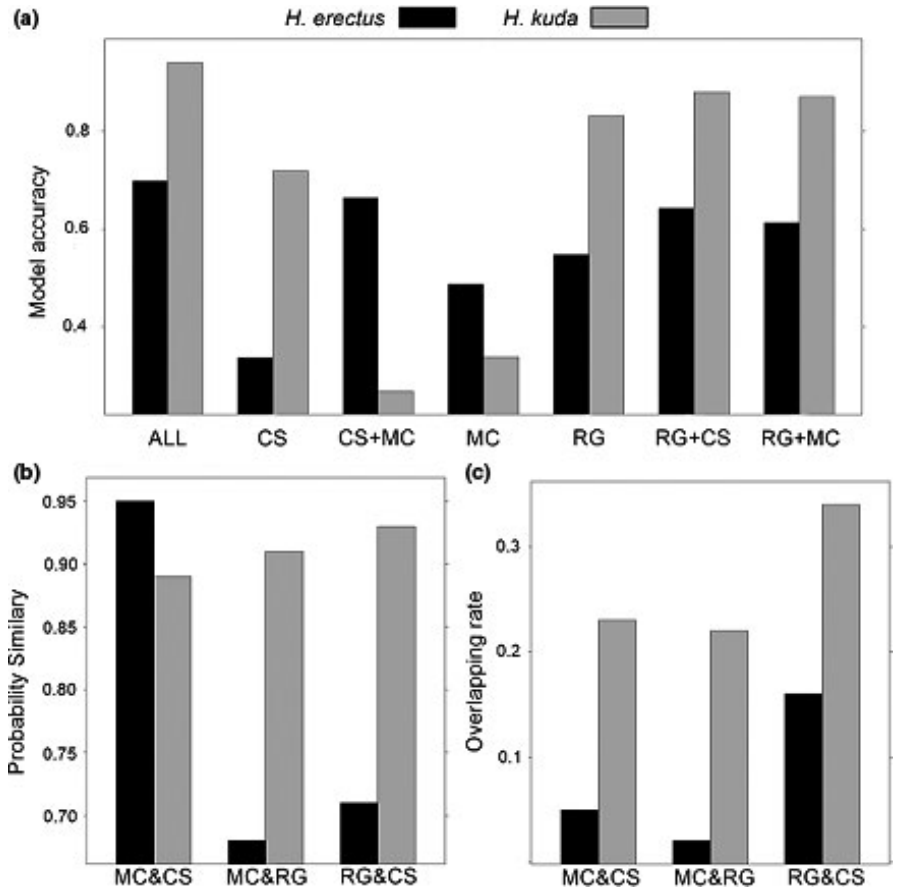
Based on the above results, we chose Dataset 3 (the best predictor dataset) and dataset ALL (the best species dataset) to construct models for the 34 species with a least five occurrences (Tables S3.2 and S3.3). The remaining eight species with few occurrences were



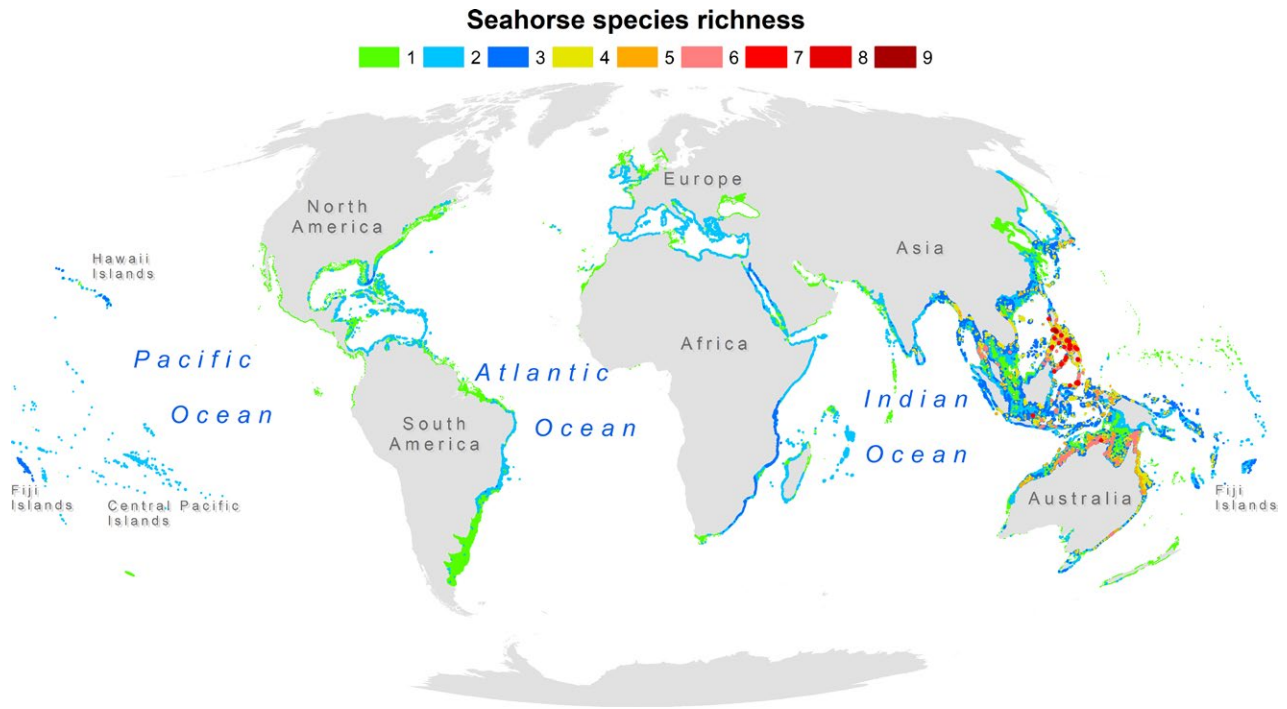
**FIGURE 2** Maxent models compared among three predictor datasets (Dataset 1, 2 and 3) across 16 species using AUC (model predictive accuracy), predicted area ratio (proportion of predictive presences to the potential range) and habitat variable importance (cumulative importance of all habitat variables used in the model)

not modelled (Table S3.4). Finally, we derived acceptable models for 33 species, with excellent performance (AUC = 0.90–1) for ten species, good performance (AUC = 0.80–0.89) for twelve species and fair performance (AUC = 0.70–0.78) for eleven species (Table S3.3). The model was poor (AUC = 0.56) and rejected for one species (*H. casscsio*, endemic to China and only recently described by Zhang et al., 2016).

The predicted biodiversity map demonstrated that locations with high species richness (value = 4–9) of seahorses were largely concentrated in tropical shallow waters of the central Indo-Pacific, with the epicentres mainly in the Philippines (Figures 4 and 5, Figures S3.1–S3.13 in Appendix S3). Other biodiversity epicentres near the central Indo-Pacific were in southern India and Sri Lanka (Figure S3.3), Taiwan (China, Figure S3.4), subtropical Japan (Figure S3.4), and Sydney and Melbourne in Australia (Figure S3.6). We also



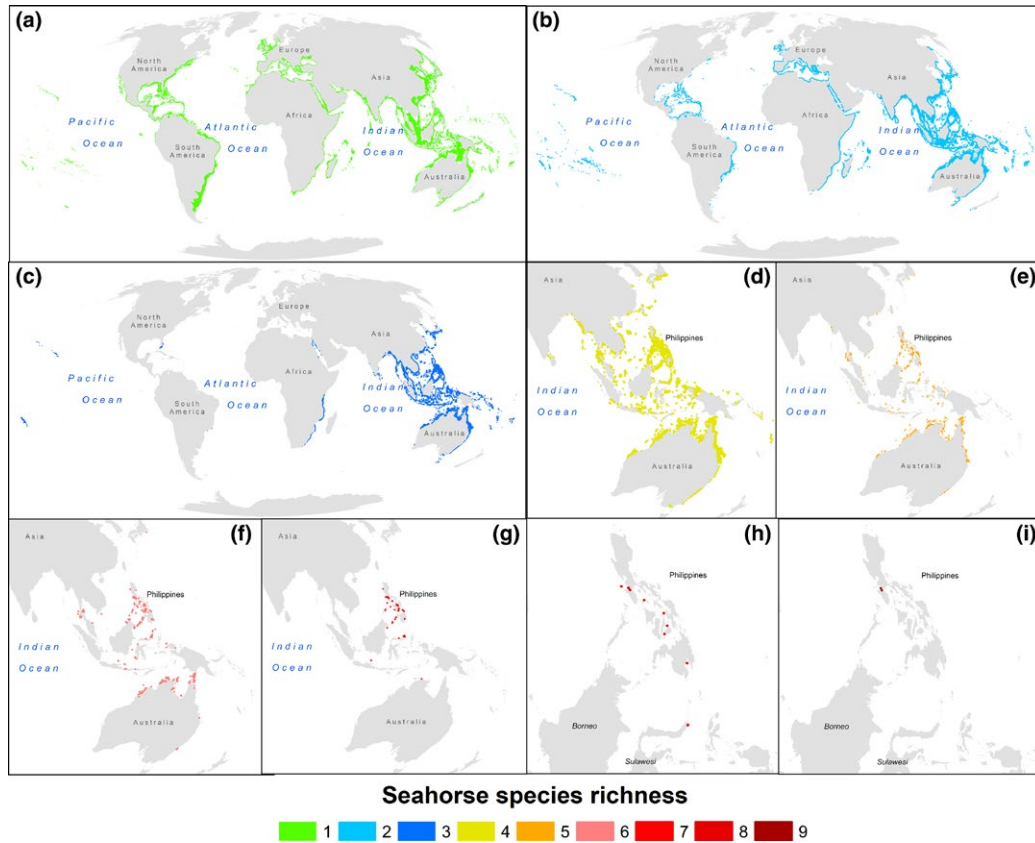
**FIGURE 3** Comparisons among Maxent models based on three different sources of seahorse occurrences, that is research grade, citizen science, and museum collections, and their combinations. They were compared on two species (*Hippocampus erectus*, *Hippocampus kuda*) with sufficient sightings from each source



**FIGURE 4** Global map of biodiversity distributions of seahorse species

predicted moderate species richness (value = 3) in southern Florida (Figure S3.1), northern and central Red Sea (Figure S3.2), south-east Africa (Figure S3.2), Hawaii and Fiji (Figure S3.7).

In total, the predicted suitable area for seahorses was 9 million km<sup>2</sup> (2.5% of the ocean's surface), with large extents of geographical separation among species; 84% of the "potential range" (Figure 1)



**FIGURE 5** Separated maps of different levels of seahorse species richness

was either unsuitable to seahorses (40.8%) or suitable to one species alone (43.2%); 15.5% of the potential range was suitable to two (13.3%) or three (2.3%) species together, and 0.4% was fit for more than three species together.

### 3.4 | Key predictors and habitats

The ranks of predictor importance suggested that all 16 predictors provided valuable information in our models, although the most important factors varied among species (Figure 6). Generally, depth, distance to the nearest sponge, distance to the nearest macroalgae, pH and ocean temperature (range and mean) were the most influential predictors (Figure 6, Table S3.5). The marginal response curves revealed that generally seahorses were more likely to live in shallower waters (<50 m; Figure 7a); many species tended to live close to sponges but have divergent adaptations to other factors such as ocean temperature (Figure 7b,c). Habitat variable importance (HVI = 3.9%–99%) and key habitat types differed largely among seahorse species (Table S3.2). Two pygmy seahorses specializing in gorgonian corals, *H. denise* and *H. bargibanti*, had the highest HVI values.

### 3.5 | Potentially threatened species based on IUCN Criteria

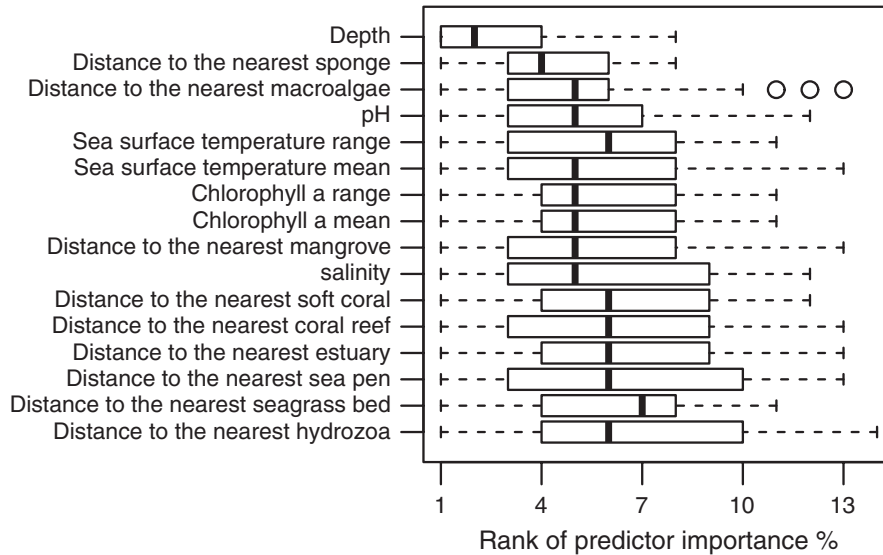
Our predicted results suggested that four species met the “Endangered” thresholds, and two met the “Vulnerable” thresholds

based on geographical range alone (IUCN Criterion B; Table S3.3). Among the six species, four met the IUCN thresholds based on area of occupancy (AOO) but not extent of occurrence (EOO), and the other two met thresholds for both AOO and EOO. However, information about population fragmentation, fluctuation or decline is required to fully justify these categories under Criterion B of IUCN. We also identified seven species with fewer than five known locations (Table S3.4) that might be considered as “Vulnerable” based on IUCN Criterion D2.

## 4 | DISCUSSION

Our study demonstrates that appropriate integration of multiple sources of species occurrences and habitat datasets is vital to derive robust SDMs to inform rare species conservation. We provide global-scale, spatially explicit maps and conservation knowledge that are urgently needed for a group of rare and data-poor marine fishes (i.e., seahorses). Our analyses highlight that proximity to habitats is more informative than habitat presence/absence for improving model accuracy and detecting key habitats. Meanwhile, we indicate that it is better to combine CS and/or MC with RG, whenever available, to improve model accuracy. Our study also demonstrates that SDM-based predictive maps can help to identify potentially threatened species with small area of occupancy.



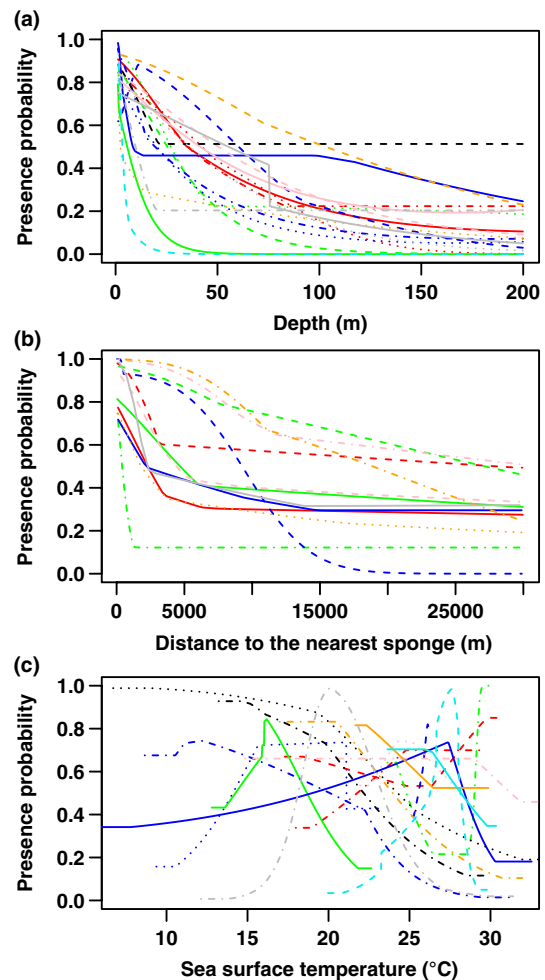


**FIGURE 6** Ranks of predictor importance across 33 seahorse species with acceptable models in our study. The box plots are shown with the rank on the x-axis and predictors on the y-axis ordered by the mean rank (not the median) from the highest to the lowest. Note that lower numbers indicate higher ranks

**4.1 | Global biogeographic pattern and habitat associations of seahorses**

Our predicted biogeographic pattern of seahorse species is consistent with a previous analysis of coastal fishes in general (Tittensor et al., 2010). The latitudinal gradient that generally more species live in the tropics may be largely shaped by temperature (Tittensor et al., 2010). Temperature variables can influence food availability and climate suitability (Willig, Kaufman, & Stevens, 2003), and had high importance in our models. But like other shore fishes, more species of seahorses are predicted to occur in the central Indo-Pacific (especially the Philippines) versus other tropical regions (e.g., Caribbean; Tittensor et al., 2010). This longitudinal divergence matches well with the hypothesized footprints of seahorse origin (probably in north-eastern Australia), dispersal (by rafting), segregation (e.g., the closure of the Isthmus of Panama) and evolution over past 20–30 million years (Boehm et al., 2013; Casey, Hall, Stanley, & Vincent, 2004; Teske, Cherry, & Matthee, 2004). Higher availability and heterogeneity of shallow-water habitats in the central Indo-Pacific might facilitate species immigration and diversification (Sanciango, Carpenter, Etnoyer, & Moretzsohn, 2013), as we find that seahorses tend to live in shallow depths and have diverse habitat preferences among species.

Our study provides comprehensive, global evidence to support the hypothesis that seahorse species have different levels of habitat reliance and species-specific habitat preferences (Curtis & Vincent, 2005). We are the first to provide quantitative evidence that habitat dependency varies largely among seahorse species. We demonstrate global divergences on distributions and habitat associations among species, which has been observed in local studies (Curtis & Vincent, 2005; Lourie, Green, & Vincent, 2005; Zhang & Vincent, 2017). Similar divergence patterns among congeneric species are



**FIGURE 7** Species response curves of depth, distance to the nearest sponge and mean sea surface temperature, with different colours and types of line for different species

also common in pipefishes (Malavasi et al., 2007) and other fishes (Fairclough, Clarke, Valesini, & Potter, 2008; Lombarte, Recasens, González, & de Sola, 2000). Such segregation and difference may be important to minimize competition among closely related species (Fairclough et al., 2008). Our evidence suggests that habitat-forming sponges are important to many seahorse species at large spatial scales. These sponges may provide various functional roles (e.g., shelter and prey sources) in seahorse's life (Bell, 2008). Further discussion about seahorse-habitat/environment relationships is presented in Appendix S4.

## 4.2 | Conservation implications for seahorses

The spatially explicit maps and habitat-preference knowledge derived from our models may inform both global-scale and local-scale seahorse conservation. For instance, to minimize the impact of international trade, CITES Parties can use our maps to locate focus areas rich in those heavily traded species at the global scale (Vincent et al., 2011). These maps can further be used with threat maps if available to determine global priority areas for seahorse conservation. Conservation programmes can use and validate our maps and habitat knowledge to study and protect seahorse populations at local scales (Zhang & Vincent, 2017).

Our study provided new geographical information to assess conservation status for seahorses. IUCN requires assessors to evaluate species against all criteria (A–E) with available data and to assign the most severe category to the species (IUCN Standards and Petitions Subcommittee, 2017). Previous assessments of seahorses largely applied Criterion A (i.e., population decline rate). Only one case (*H. capensis*) used Criterion B1 (Extent of Occurrence). We identified six species that met threatened thresholds of Criterion B2 (area of occupancy, AOO). Currently, one (*H. capensis*) has been evaluated as endangered, but other five were either “Least Concern” or “Data Deficient” (Table S3.3, IUCN, 2017). Although the AOO is estimated based on predictive maps and thus contain uncertainties and require further improvement (Guisan et al., 2013), it is likely that we overestimated the AOO. If we include other constraint factors (e.g., anthropogenic impacts) in our models, the AOO might become smaller. Given that, these five species can still justify the IUCN “Vulnerable” threshold of AOO (<2,000 km<sup>2</sup>) for Criterion B2.

## 4.3 | Utility of distance-based habitat predictors in SDMs

We are the first to demonstrate that “proximity to habitats” is more informative than “habitat presence/absence” to predict distributions of low-mobility organisms. Proximity to habitats has been employed and proved useful in SDMs for high-mobility animals, including reef fishes (Shelton, Thorson, Ward, & Feist, 2014), Bonelli's eagles (Balbontín, 2005) and bats (Rainho & Palmeirim, 2011). One underlying assumption is that high-mobility animals choose to live close to important resource patches (e.g.,

feeding grounds). But this assumption might be questionable for low-mobility species, as they were believed unlikely able to select habitats at large spatial scales. Instead, site-level habitat characteristics (e.g., habitat presence/absence) were used in low-mobility species including Madagascar geckos (Pearson et al., 2007) and Juliana's golden mole (Jackson & Robertson, 2011). Although a few studies have used proximity to habitats for low-mobility species (Dillard, Russell, & Ford, 2008; Dorrough & Ash, 1999), ours is the first to indicate that proximity to habitats can be more useful than habitat presence/absence for low-mobility animals.

The correlations between sedentary animals and the proximity to habitats may result from two factors: (1) behaviour and ecology of the animal and (2) coarse resolution of original data. First, low-mobility organisms can disperse across large spatial scales through natural disturbance (e.g., ocean currents) and hitchhiking (Luiz, Allen, Robertson, Floeter, & Madin, 2015). They may then choose “stops” in preferred habitats or be “dropped” in unsuitable ones. In the second case, the animal may move over a distance longer than its home range to find suitable habitats (Caldwell & Vincent, 2013; Matthews, 1990). Some sightings of the species might be recorded during this “habitat-finding” process and thus distorted the species-habitat spatial relations. Second, there might be spatial mismatch between species occurrences and habitat locations due to coarse resolutions of the original records. As a result, some species occurrences might not overlap with suitable habitats, but they are still close in space.

## 4.4 | Utility of different sources of species data in SDMs

This study is among the first to demonstrate that adding citizen science (CS) and museum collections (MC) to research-grade data (RG) can help to derive more accurate SDMs. We encourage the integration of MC and/or CS data with RG data as MC and CS can be more sufficient than RG and helpful for improving model accuracy. MC are usually the most data-rich source for many organisms (Ponder, Carter, Flemons, & Chapman, 2001), as is also shown in our seahorses. MC have been commonly used in SDM research (Newbold, 2010), although we demonstrate, compared with RG, MC may derive lower quality models with quite different predictions (Aubry et al., 2017). This is also true for citizen science. Our results suggest that combining quality-unknown data (MC and CS) without RG is risky, as this may accumulate errors and result in lower-accuracy models (Zhang & Vincent, 2017).

Additionally, compared with MC, CS might have some advantages. First, CS provide more recent georeferenced information than historical collections to reflect the current distributions of species. Second, validating CS data might be easier than checking the historical specimens from worldwide museums. These advantages may partly explain why CS derived more similar predictions to those of RG data than MC did in our study. A recent study on a rare snow quail species also indicated that CS data could derive similar predictions

to those based on RG data (Jackson et al., 2015), encouraging the use of citizen-science data sources.

#### 4.5 | Applying SDMs to IUCN Red List assessment

Our study highlights that applying SDMs can derive useful geographical information for the assessment of conservation status for rare species (Aubry et al., 2017). Rare species (especially habitat specialists) that have low population density are often patchily distributed along with their key habitats (Marcer et al., 2013). Therefore, even though they may have large extent of occurrence (EOO, IUCN Criterion B1), their area of occupancy (AOO, IUCN Criterion B2) could be very small. For example, in our study, *H. fisheri* (a species endemic to Hawaii) has an EOO that does not meet a threatened threshold, but its AOO is small enough to pass the “Endangered” threshold of AOO according to IUCN Criterion B. Therefore, estimating AOO based on SDMs such as Maxent may add essential information to assess conservation status for rare species.

#### ACKNOWLEDGEMENTS

This is a contribution from Project Seahorse. We thank Dr. Xue Xinxin, Francesca Fogliata, Kelly Wang, Marie McCallum, Myra Lam, Naman Siddique and Sammi Conard in assisting seahorse data collection from peer-reviewed literature. Many thanks also to those hundreds of citizen scientists who submitted seahorse sightings around the world on our iSeahorse website. We are grateful to Dr. Christopher Harley, Dr. Sarah Gergel, Dr. William Cheung, Riley Pollom, Kyle Gillespie and Muhammed Oyinola for their constructive comments. XZ was supported by a Chinese Doctoral Scholarship from the China Scholarship Council. Project Seahorse is supported by Guyliau Chocolates (Belgium) in a partnership for marine conservation.

#### DATA ACCESSIBILITY

All collated seahorse species occurrences, compiled habitat GIS layers, and the Maxent results generated for the 33 species are available from the figshare database under the project folder “Seahorse Global Distribution.”

#### ORCID

Xiong Zhang  <http://orcid.org/0000-0003-0132-734X>

#### REFERENCES

- Alder, J. (2003). Putting the coast in the “Sea Around Us”. The Sea Around Us Newsletter, 15, 1-2. URL: <http://seararoundus.org/newsletter/Issue15.pdf>; <http://data.unep-wcmc.org/datasets/23> (version 2.0).
- Altman, D. G., & Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *British Medical Journal*, 308, 1552. <https://doi.org/10.1136/bmj.308.6943.1552>
- Aubry, K. B., Raley, C. M., & McKelvey, K. S. (2017). The importance of data quality for generating reliable distribution models for rare, elusive, and cryptic species. *PLoS ONE*, 12, e0179152. <https://doi.org/10.1371/journal.pone.0179152>
- Aylesworth, L., Xavier, J., Oliveira, T., Tenorio, G., Diniz, A., & Rosa, I. (2015). Regional-scale patterns of habitat preference for the seahorse *Hippocampus reidi* in the tropical estuarine environment. *Aquatic Ecology*, 49, 499–512. <https://doi.org/10.1007/s10452-015-9542-3>
- Balbontín, J. (2005). Identifying suitable habitat for dispersal in Bonelli's eagle: An important issue in halting its decline in Europe. *Biological Conservation*, 126, 74–83. <https://doi.org/10.1016/j.biocon.2005.04.023>
- Bell, J. J. (2008). The functional roles of marine sponges. *Estuarine, Coastal and Shelf Science*, 79, 341–353. <https://doi.org/10.1016/j.ecss.2008.05.002>
- Boehm, J., Woodall, L., Teske, P. R., Lourie, S. A., Baldwin, C., Waldman, J., & Hickerson, M. (2013). Marine dispersal and barriers drive Atlantic seahorse diversification. *Journal of Biogeography*, 40, 1839–1849.
- Brooks, T. M., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., Rylands, A. B., Konstant, W. R., & Magin, G. (2002). Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, 16, 909–923. <https://doi.org/10.1046/j.1523-1739.2002.00530.x>
- Calabrese, J. M., Certain, G., Kraan, C., & Dormann, C. F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, 23, 99–112. <https://doi.org/10.1111/geb.12102>
- Caldwell, I. R., & Vincent, A. C. (2013). A sedentary fish on the move: effects of displacement on long-snouted seahorse (*Hippocampus guttulatus* Cuvier) movement and habitat use. *Environmental Biology of Fishes*, 96, 67–75. <https://doi.org/10.1007/s10641-012-0023-4>
- Casey, S. P., Hall, H. J., Stanley, H. F., & Vincent, A. C. (2004). The origin and evolution of seahorses (genus *Hippocampus*): a phylogenetic study using the cytochrome b gene of mitochondrial DNA. *Molecular Phylogenetics and Evolution*, 30, 261–272. <https://doi.org/10.1016/j.ympev.2003.08.018>
- Cohen, F. P., Valenti, W. C., Planas, M., & Calado, R. (2017). Seahorse aquaculture, biology and conservation: Knowledge gaps and research opportunities. *Reviews in Fisheries Science & Aquaculture*, 25, 100–111. <https://doi.org/10.1080/23308249.2016.1237469>
- Curtis, J. (2006). A case of mistaken identity: Skin filaments are unreliable for identifying *Hippocampus guttulatus* and *Hippocampus hippocampus*. *Journal of Fish Biology*, 69, 1855–1859. <https://doi.org/10.1111/j.1095-8649.2006.01228.x>
- Curtis, J. M., & Vincent, A. C. (2005). Distribution of sympatric seahorse species along a gradient of habitat complexity in a seagrass-dominated community. *Marine Ecology Progress Series*, 291, 81–91. <https://doi.org/10.3354/meps291081>
- Dillard, L. O., Russell, K. R., & Ford, M. W. (2008). Habitat models of occurrence for the threatened Cheat Mountain salamander, *Plethodon nettingi*. *Applied Herpetology*, 5, 201–224. <https://doi.org/10.1163/157075408785911057>
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., & Leitão, P. J. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Dorrough, J., & Ash, J. E. (1999). Using past and present habitat to predict the current distribution and abundance of a rare cryptic lizard, *Delma impar* (Pygopodidae). *Austral Ecology*, 24, 614–624. <https://doi.org/10.1046/j.1442-9993.1999.00995.x>
- Fairclough, D., Clarke, K., Valesini, F., & Potter, I. (2008). Habitat partitioning by five congeneric and abundant Choerodon species (Labridae) in a large subtropical marine embayment. *Estuarine, Coastal and Shelf Science*, 77, 446–456. <https://doi.org/10.1016/j.ecss.2007.10.004>

- Fernandez, M., Blum, S., Reichle, S., Guo, Q., Holzman, B., & Hamilton, H. (2009) Locality uncertainty and the differential performance of four common niche-based modeling techniques. *Biodiversity Informatics*, 6, 36–52.
- Foster, S. J., & Vincent, A. C. J. (2004). Life history and ecology of seahorses: Implications for conservation and management. *Journal of Fish Biology*, 65, 1–61. <https://doi.org/10.1111/j.0022-1112.2004.00429.x>
- Franklin, J. (2010) *Mapping species distributions: spatial inference and prediction*. New York: Cambridge University Press. page 21–32. <https://doi.org/10.1017/CBO9780511810602>
- Gaston, K. J., & Fuller, R. A. (2009). The sizes of species' geographic ranges. *Journal of Applied Ecology*, 46, 1–9. <https://doi.org/10.1111/j.1365-2664.2008.01596.x>
- Giri, C., Ochieng, E., Tieszen, L. L., Zhu, Z., Singh, A., Loveland, T., ... Duke, N. (2011). Status and distribution of mangrove forests of the world using earth observation satellite data (version 1.3, updated by UNEP-WCMC). *Global Ecology and Biogeography*, 20, 154–159. <https://doi.org/10.1111/j.1466-8238.2010.00584.x>  
Data URL: <http://data.unep-wcmc.org/datasets/4> <https://doi.org/10.1111/j.1466-8238.2010.00584.x>
- Graham, C. H., Elith, J., Hijmans, R. J., Guisan, A., Townsend Peterson, A., & Loiselle, B. A. (2008). The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, 45, 239–247.
- Graham, C. H., Ferrier, S., Huettman, F., Moritz, C., & Peterson, A. T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, 19, 497–503. <https://doi.org/10.1016/j.tree.2004.07.006>
- Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I., ... Mantyka-Pringle, C. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, 16, 1424–1435. <https://doi.org/10.1111/ele.12189>
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29–36. <https://doi.org/10.1148/radiology.143.1.7063747>
- Hanski, I. (2011). Habitat loss, the dynamics of biodiversity, and a perspective on conservation. *AMBIO: A Journal of the Human Environment*, 40, 248–255. <https://doi.org/10.1007/s13280-011-0147-3>
- Harasti, D., Martin-Smith, K., & Gladstone, W. (2014). Ontogenetic and sex-based differences in habitat preferences and site fidelity of White's seahorse *Hippocampus whitei*. *Journal of Fish Biology*, 85, 1413–1428. <https://doi.org/10.1111/jfb.12492>
- Harris, G. M., Jenkins, C. N., & Pimm, S. L. (2005). Refining biodiversity conservation priorities. *Conservation Biology*, 19, 1957–1968. <https://doi.org/10.1111/j.1523-1739.2005.00307.x>
- Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29, 773–785. <https://doi.org/10.1111/j.0906-7590.2006.04700.x>
- Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E., Hungate, B. A., Matulich, K. L., & O'Connor, M. I. (2012). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, 486, 105. <https://doi.org/10.1038/nature11118>
- IUCN (2017) The IUCN Red List of Threatened Species. Version 2017-3. [www.iucnredlist.org](http://www.iucnredlist.org). Downloaded on 30 January 2018.
- IUCN Standards and Petitions Subcommittee (2017). Guidelines for application of IUCN red list criteria at regional and national levels. Version 4.
- Jackson, M. M., Gergel, S. E., & Martin, K. (2015). Citizen science and field survey observations provide comparable results for mapping Vancouver Island White-tailed Ptarmigan (*Lagopus leucura saxatilis*) distributions. *Biological Conservation*, 181, 162–172. <https://doi.org/10.1016/j.biocon.2014.11.010>
- Jackson, C. R., & Robertson, M. P. (2011). Predicting the potential distribution of an endangered cryptic subterranean mammal from few occurrence records. *Journal for Nature Conservation*, 19, 87–94. <https://doi.org/10.1016/j.jnc.2010.06.006>
- Kleiber, D., Blight, L., Caldwell, I., & Vincent, A. (2011). The importance of seahorses and pipefishes in the diet of marine animals. *Reviews in Fish Biology and Fisheries*, 21, 205–223. <https://doi.org/10.1007/s11160-010-9167-5>
- Lawson, J. M., Foster, S. J., & Vincent, A. C. (2017). Low bycatch rates add up to big numbers for a genus of small fishes. *Fisheries*, 42, 19–33. <https://doi.org/10.1080/03632415.2017.1259944>
- Lombarte, A., Recasens, L., González, M., & de Sola, L. G. (2000). Spatial segregation of two species of Mullidae (*Mullus surmuletus* and *M. barbatus*) in relation to habitat. *Marine Ecology Progress Series*, 206, 239–249. <https://doi.org/10.3354/meps206239>
- Lourie, S., Green, D., & Vincent, A. (2005). Dispersal, habitat differences, and comparative phylogeography of Southeast Asian seahorses (Syngnathidae: Hippocampus). *Molecular Ecology*, 14, 1073–1094. <https://doi.org/10.1111/j.1365-294X.2005.02464.x>
- Lourie, S. A., Pollom, R. A., & Foster, S. J. (2016). A global revision of the Seahorses Hippocampus Rafinesque 1810 (Actinopterygii: Syngnathiformes): Taxonomy and biogeography with recommendations for further research. *Zootaxa*, 4146, 1–66. <https://doi.org/10.11646/zootaxa.4146.1.1>
- Luiz, O. J., Allen, A. P., Robertson, D. R., Floeter, S. R., & Madin, J. S. (2015). Seafarers or castaways: ecological traits associated with rafting dispersal in tropical reef fishes. *Journal of Biogeography*, 42, 2323–2333. <https://doi.org/10.1111/jbi.12574>
- Malavasi, S., Franco, A., Riccato, F., Valerio, C., Torricelli, P., & Franzoi, P. (2007). Habitat selection and spatial segregation in three pipefish species. *Estuarine, Coastal and Shelf Science*, 75, 143–150. <https://doi.org/10.1016/j.ecss.2007.02.022>
- Manning, C.G. (2017) *How is a family of sedentary marine fishes shaped by its habitats, prey, and predators?* University of British Columbia.
- Marcen, A., Sáez, L., Molowny-Horas, R., Pons, X., & Pino, J. (2013). Using species distribution modelling to disentangle realised versus potential distributions for rare species conservation. *Biological Conservation*, 166, 221–230. <https://doi.org/10.1016/j.biocon.2013.07.001>
- Matthews, K. R. (1990). A telemetric study of the home ranges and homing routes of copper and quillback rockfishes on shallow rocky reefs. *Canadian Journal of Zoology*, 68, 2243–2250. <https://doi.org/10.1139/z90-312>
- Mittermeier, R.A., Turner, W.R., Larsen, F.W., Brooks, T.M., & Gascon, C. (2011) Global biodiversity conservation: the critical role of hotspots. In *Biodiversity hotspots* (pp. 3–22). Berlin, Heidelberg: Springer.
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403, 853. <https://doi.org/10.1038/35002501>
- Newbold, T. (2010). Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, 34, 3–22. <https://doi.org/10.1177/0309133309355630>
- Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Townsend Peterson, A. (2007). Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34, 102–117.
- Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. Proceedings of the twenty-first international conference on Machine learning. 83.
- Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, 31, 161–175. <https://doi.org/10.1111/j.0906-7590.2008.5203.x>
- Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, 40(7), 887–893. <https://doi.org/10.1111/ecog.03049>
- Ponder, W. F., Carter, G., Flemons, P., & Chapman, R. (2001). Evaluation of museum collection data for use in

- biodiversity assessment. *Conservation Biology*, 15, 648–657. <https://doi.org/10.1046/j.1523-1739.2001.015003648.x>
- Rainho, A., & Palmeirim, J. M. (2011). The importance of distance to resources in the spatial modelling of bat foraging habitat. *PLoS ONE*, 6, e19227. <https://doi.org/10.1371/journal.pone.0019227>
- Reynolds, J. D., Dulvy, N. K., Goodwin, N. B., & Hutchings, J. A. (2005). Biology of extinction risk in marine fishes. *Proceedings of the Royal Society of London B: Biological Sciences*, 272, 2337–2344. <https://doi.org/10.1098/rspb.2005.3281>
- Roos, G., Van Wassenbergh, S., Aerts, P., Herrel, A., & Adriaens, D. (2011). Effects of snout dimensions on the hydrodynamics of suction feeding in juvenile and adult seahorses. *Journal of Theoretical Biology*, 269, 307–317. <https://doi.org/10.1016/j.jtbi.2010.10.023>
- Sanciangco, J. C., Carpenter, K. E., Etnoyer, P. J., & Moretzsohn, F. (2013). Habitat availability and heterogeneity and the Indo-Pacific warm pool as predictors of marine species richness in the tropical Indo-Pacific. *PLoS ONE*, 8, e56245. <https://doi.org/10.1371/journal.pone.0056245>
- Searcy, C. A., & Shaffer, H. B. (2016). Do ecological niche models accurately identify climatic determinants of species ranges? *The American Naturalist*, 187, 423–435. <https://doi.org/10.1086/685387>
- Shelton, A. O., Thorson, J. T., Ward, E. J., & Feist, B. E. (2014). Spatial semiparametric models improve estimates of species abundance and distribution. *Canadian Journal of Fisheries and Aquatic Sciences*, 71, 1655–1666. <https://doi.org/10.1139/cjfas-2013-0508>
- Stirling, D. A., Boulcott, P., Scott, B. E., & Wright, P. J. (2016). Using verified species distribution models to inform the conservation of a rare marine species. *Diversity and Distributions*, 22, 808–822. <https://doi.org/10.1111/ddi.12447>
- Teske, P. R., Cherry, M. I., & Matthee, C. A. (2004). The evolutionary history of seahorses (Syngnathidae: Hippocampus): molecular data suggest a West Pacific origin and two invasions of the Atlantic Ocean. *Molecular Phylogenetics and Evolution*, 30, 273–286. [https://doi.org/10.1016/S1055-7903\(03\)00214-8](https://doi.org/10.1016/S1055-7903(03)00214-8)
- Tittensor, D. P., Mora, C., Jetz, W., Lotze, H. K., Ricard, D., Berghe, E. V., & Worm, B. (2010). Global patterns and predictors of marine biodiversity across taxa. *Nature*, 466, 1098–1101. <https://doi.org/10.1038/nature09329>
- Vincent, A., Foster, S., & Koldewey, H. (2011). Conservation and management of seahorses and other Syngnathidae. *Journal of Fish Biology*, 78, 1681–1724. <https://doi.org/10.1111/j.1095-8649.2011.03003.x>
- Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., and De Clerck, O. (2012). Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, 21(2), 272–281.
- UNEP-WCMC, WorldFish Centre, WRI, TNC (2010). Global distribution of warm-water coral reefs, compiled from multiple sources including the Millennium Coral Reef Mapping Project. Version 2.0. Includes contributions from IMaRS-USF and IRD (2005), IMaRS-USF (2005) and Spalding et al. (2001). Cambridge (UK): UN Environment World Conservation Monitoring Centre. URL: <http://data.unep-wcmc.org/datasets/1>
- UNEP-WCMC, Short FT (2017). Global distribution of seagrasses (version 5.0). Fourth update to the data layer used in Green and Short (2003). Cambridge (UK): UN Environment World Conservation Monitoring Centre. URL: <http://data.unep-wcmc.org/datasets/7>
- Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, 62, 2868–2883. <https://doi.org/10.1111/j.1558-5646.2008.00482.x>
- Weatherall, P., Marks, K., Jakobsson, M., Schmitt, T., Tani, S., Arndt, J. E., Rovere, M., Chayes, D., Ferrini, V., & Wigley, R. (2015). A new digital bathymetric model of the world's oceans. *Earth and Space Science*, 2(8), 331–345.
- Whittaker, R. H., Levin, S. A., & Root, R. B. (1973). Niche, habitat, and ecotope. *The American Naturalist*, 107, 321–338. <https://doi.org/10.1086/282837>
- Willig, M. R., Kaufman, D., & Stevens, R. (2003). Latitudinal gradients of biodiversity: Pattern, process, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics*, 34, 273–309. <https://doi.org/10.1146/annurev.ecolsys.34.012103.144032>
- Zhang, Y.-H., Qin, G., Wang, X., & Lin, Q. (2016). A new species of seahorse (Teleostei: Syngnathidae) from the South China Sea. *Zootaxa*, 4170, 384–392. <https://doi.org/10.11646/zootaxa.4170.2.11>
- Zhang, X., & Vincent, A. C. (2017). Integrating multiple datasets with species distribution models to inform conservation of the poorly-recorded Chinese seahorses. *Biological Conservation*, 211, 161–171. <https://doi.org/10.1016/j.biocon.2017.05.020>

## BIOSKETCH

**Project Seahorse** is a small team of scientists and conservationists with offices at the University of British Columbia in Vancouver, Canada, and the Zoological Society of London in the United Kingdom. We work all over the world, with research and conservation projects in Asia, Europe, Australia and the Americas.

**Xiong Zhang** is a team member of Project Seahorse. He is a PhD Candidate in the Zoology Department, The University of British Columbia. His research interests focus on using advanced techniques (e.g., species distribution models) to derive ecological and conservation knowledge for rare and threatened species such as seahorses.

Author contributions: X. Z. conceived the ideas together with A. C. J. V. and X. Z. conducted all the data analyses and wrote the manuscript, with the help of A. C. J. V.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Zhang X, Vincent ACJ. Predicting distributions, habitat preferences and associated conservation implications for a genus of rare fishes, seahorses (*Hippocampus* spp.). *Divers Distrib*. 2018;00:1–13. <https://doi.org/10.1111/ddi.12741>