WILEY LIMNOLOGY AND OCEANOGRAPHY

Endemicity and climatic niche differentiation in three marine ciliated protists

Journal:	*Limnology and Oceanography*
Manuscript ID	LO-17-0420.R2
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	04-Jun-2018
Complete List of Authors:	Williams, Richard; Linneuniversitet Fakulteten for halso och livsvetenskap, EEMiS; Universidad Complutense de Madrid, Department of Zoology and Physical Anthropology; University of Kansas, Biodiversity Institute Owens, Hannah; University of Florida, Department of Biology; University of Kansas, Biodiversity Institute Clamp, John; North Carolina Central University, Department of Biological and Biomedical Sciences Peterson, Town; University of Kansas, Biodiversity Institute Warren, Alan; The Natural History Museum, Department of Zoology Martin-Cereceda, Mercedes; Universidad Complutense de Madrid, Department of Microbiology III
Keywords:	Ciliate, Ecological Niche Models, Niche Differentiation
Abstract:	The biogeographic pattern of single-celled eukaryotes (protists), including ciliates, is poorly understood. Most marine species are believed to have a relatively high dispersal potential, such that both globally-distributed and geographically-isolated taxa exist. Primary occurrence data for three large, easily identified ciliate species, Parafavella gigantea, Schmidingerella serrata and Zoothamnium pelagicum, and environmental data drawn from the National Oceanic and Atmospheric Administration's (NOAA) World Ocean Atlas were used to estimate each species' spatial and environmental distributions using Maxent v3.3.3k. The predictive power of the models was tested with a series of spatial stratification studies, which were evaluated using partial receiver operating characteristic statistics. Differences between niches occupied by each taxon were evaluated using background similarity tests. All predictions showed significant ability to anticipate test points. The null hypotheses of niche similarity were rejected in all background similarity tests comparing the niches among the three species. This paper provides a first quantitative assessment of environmental conditions associated with three species of ciliates and a first estimate of their spatial distributions in the North Atlantic, which can serve as a benchmark against which to document distributional shifts. These species follow consistent, predictable patterns related to climate and environmental biochemistry; the importance of climatic conditions as regards protist distributions is noteworthy considering the effects of global climate change.

TITLE

Endemicity and climatic niche differentiation in three marine ciliated protists

AUTHORS

Richard AJ Williams[1,2,3*], Hannah L Owens[2,4], John Clamp[5]†, A Townsend Peterson[2], Alan

Warren[6], Mercedes Martín Cereceda[7]

[1] Centre for Ecology and Evolution in Microbial Model Systems, Linnaeus University, SE-391 82, Kalmar, Sweden

[2] Biodiversity Institute, University of Kansas, Lawrence, KS 66045, USA

[3] Department of Biodiversity, Ecology and Evolution, Universidad Complutense de Madrid, 28040, Madrid, Spain

[4] Department of Biology, University of Florida, Gainesville, FL 32611, USA

[5] Department of Biological and Biomedical Sciences, North Carolina Central University, Durham, NC 27707, USA

[6] Department of Life Sciences, Natural History Museum, London SW7 5BD, UK

[7] Department of Genetics, Physiology and Microbiology, Universidad Complutense de Madrid, 28040, Madrid, Spain

* Corresponding author: richard.williams@lnu.se

† John Clamp passed away on Feb. 6th, 2018, following a short illness

KEYWORDS

Ciliates, Ecological Niche Models, Niche Differentiation, *Parafavella gigantea*, *Schmidingerella serrata, Zoothamnium pelagicum*

RUNNING HEAD

Marine ciliate niche and distribution

Abstract: 213 words

Paper: 4445 words

References: 53

ABSTRACT

The biogeographic pattern of single-celled eukaryotes (protists), including ciliates, is poorly understood. Most marine species are believed to have a relatively high dispersal potential, such that both globally-distributed and geographically-isolated taxa exist. Primary occurrence data for three large, easily identified ciliate species, *Parafavella gigantea*, *Schmidingerella serrata* and *Zoothamnium pelagicum*, and environmental data drawn from the National Oceanic and Atmospheric Administration's (NOAA) World Ocean Atlas were used to estimate each species' spatial and environmental distributions using Maxent v3.3.3k. The predictive power of the models was tested with a series of spatial stratification studies, which were evaluated using partial receiver operating characteristic statistics. Differences between niches occupied by each taxon were evaluated using background similarity tests. All predictions showed significant ability to anticipate test points. The null hypotheses of niche similarity were rejected in all background similarity tests comparing the niches among the three species. This paper provides a first quantitative assessment of environmental conditions associated with three species of ciliates and a first estimate of their spatial distributions in the North Atlantic, which can serve as a benchmark against which to document distributional shifts. These species follow consistent, predictable patterns related to climate and environmental biochemistry; the importance of climatic conditions as regards protist distributions is noteworthy considering the effects of global climate change.

INTRODUCTION

Marine microorganisms are essential to the planet's health, marine biodiversity, and the fishing industry. They carry out nearly 50% of the world's photosynthesis, 70% of which is locked up in long-term carbon storage (Field et al. 1998). Microbes are the base of the marine ecological pyramid, and are vital to the marine food web that transfers energy from primary producers to higher trophic levels (Azam et al. 1983; Fenchel 2008). Understanding the current spatial distributions of marine microorganisms is thus important to improving understanding of their functions, and is fundamental for impact assessment of effects of climate change.

Microorganisms—used here to refer to all microscopic organisms (including protists, archaea, bacteria and fungi)—are generally considered to be cosmopolitan in distribution (Beijerinck 1913): "everything is everywhere, but the environment selects". They are thought to be ubiquitous thanks to huge population sizes and a consequent low probability of extirpation; also, they are limitlessly dispersible, have negligible rates of allopatric speciation compared to larger organisms, and can potentially occur at any place that meets their biological requirements (Finlay et al. 1996). These assumptions suggest that processes driving biogeographic patterns of microbe diversity are fundamentally different from macroscopic organisms, with profound implications for understanding mechanisms driving microbial distributions and evolution. A variant on this view states that whereas most free-living protist species are globally distributed, about one-third of them may be biogeographically restricted (Foissner 1999; Foissner 2006).

Nonetheless, microbial geographic patterns are still poorly understood. Description of the biogeography of protists is hampered by a scarcity of taxonomists, high frequency of misidentification and dramatic under-sampling (Foissner 2006). Traditionally, protists have been identified by morphological features, though this approach risks lumping organisms with discrete biogeography, ecology and genetics into single morphospecies (Gentekaki and Lynn 2010; Katz et al. 2011). The high rates of synonymy and existence of polymorphic life stages (phenotypic variability) in ciliates can exacerbate the problem (Dolan 2016). Protist biogeography has benefited from molecular studies, as environmental sequencing can provide abundant distributional data. Several contributions have described geographic circumscription in 18 S rDNA sequences from diverse aquatic protist groups—diatoms (Evans et al. 2009), Cercozoa (Bass et al. 2007), Amoebozoa (Aguilar et al. 2014), heterotrophic flagellates (Boenigk et al. 2006), and ciliates (Bass et al. 2009). Several recent morphological and molecular studies have found that many protists, particularly the rarer ones, follow discrete biogeographic patterns (Filker et al. 2016; Logares et al. 2015; Segovia et al. 2017). Further studies relating morphology, functionality and molecular information are needed to describe the dispersal ability of these and other protist species.

Ciliophora is a diverse phylum of heterotrophic or mixotrophic protists—the ciliates—with ~4500 taxonomically valid, free-living species (Foissner et al. 2008), characterized by the presence of hair-like organelles called cilia, used for locomotion, attachment, feeding, sensation, etc., and nuclear dimorphism (Gao et al. 2016). They have been sampled from marine environments; indeed, the Ocean Biogeographic Information System (OBIS) has locality data for nearly 200,000 Ciliophora observations, albeit with a clear bias towards sampling along major trade routes. Some microbial species are more conspicuous and easier to identify with confidence than others, leading some authors to propose these “flagship” species as the ultimate proof of endemism (Foissner 2006). Such species, the “elephants” of the microbial world, cannot be missed if they are present, because of their distinctive morphological features and/or significantly large size, with the caveat that traditional morphological-based identification of microbial species risks lumping diverse ecological and genetic species into a single morphospecies complex and underestimating cryptic diversity (Gentekaki and Lynn 2010).

Ecological niche models (ENM) infer suitable abiotic habitat conditions for non-model organisms by generating a correlational model that unites occurrence information and environmental data for the taxon of interest to determine the geographic distribution of habitat conditions correlated with species occurrences (Soberón and Peterson 2005). They have been used to explore diverse topics in distributional ecology, including species' geographic distributions, niche conservatism, spread of invasive macrospecies and diseases, and effects of climate change on species distributions; see Peterson et al. (2011) for a summary. ENMs have long been used to infer distributions of species in the marine environment (Wiley et al. 2003). To date, marine ENMs have been used mainly to create models for macrofaunal distributions (Bentlage et al. 2013; Saupe et al. 2014), although some authors have applied ENM approaches to characterize phytoplankton (Brun et al. 2015) and Foraminifera (Langer et al. 2013). Relatively few attempts have been made to describe the ecological dimensions of microbial distributions for free-living terrestrial microbes (Aguilar and Lado 2012). One noteworthy recent contribution explored the phylogeography of 18S rDNA variants of the myxomycete *Badhamia melanospora* (phylum Amoebozoa) (Aguilar et al. 2014).

The aim of this contribution is to show, for the first time in ciliates, the utility of ENMs in putting ciliate morphospecies on the map, that is, in analysing and understanding their biogeography. We focus on three flagship species from the phylum Ciliophora: *Parafavella* *gigantea* (Brandt, 1896), *Schmidingerella* (= *Favella*) *serrata* (Möbius, 1887), and *Zoothamnium pelagicum* Du Plessis, 1891. These species are particularly obvious thanks to their large body size. *Parafavella gigantea* (up to 750 µm in length) and *Schmidingerella serrata* (up to 350 µm long) are relatively obvious within plankton samples as they are much larger than many other protists. *Zoothamnium pelagicum*, though somewhat smaller (up to 120 µm), is colonial, and often forms discoidal clusters 2-3 mm thick that are clearly visible to the human eye (Laval 1968). These were the only species for which sample size and taxonomic clarity were sufficient to permit developing ecological niche models (ENMs).

METHODS

Data on ciliate distributions and ecological dimensions

Primary occurrence data for Ciliophora species were downloaded from the Global Biodiversity Information Facility (GBIF; http://www.gbif.org/) and Ocean Biogeographic Information System (OBIS; http://www.iobis.org/) on 10 December 2015. We obtained more than 200,000 observations, but most were not usable for development of ENMs: many were not supported by voucher specimens or sequences (which we consider to be necessary documentation), and taxonomic resolution for many taxa was poor. Observations were often duplicated from the same locality; and sample sizes were low for most taxa.

We found only three ciliate species with sufficient sample size, taxonomic clarity, and geographic spread to develop ENMs: *P. gigantea* (*N* = 368), *S. serrata* (*N* = 135), and *Z. pelagicum* (*N* = 279). Records for these taxa were collected between 1903 and 2008 (see Fig. 1); duplicate records falling within the same 1° grid square were eliminated. The area of coverage was limited to the North Atlantic (20°E-80°W; 0°-90°N) centred on the best-sampled ocean regions worldwide. Occurrence data were visualized in ESRI ARCGIS 9.3.1

(ESRI, Redlands, CA, USA), and showed broadly consistent distributional patterns. The few data (<2% for each taxon) that fell outside this region were nonetheless included in development of models to avoid prejudging limits of unknown distributional patterns (see Fig. 1).

Environmental data layers were drawn from National Oceanic and Atmospheric Administration's (NOAA) World Ocean Atlas (NOAA 1999), at a native spatial resolution of 1°: ~110 km at the Equator. These data layers represent long-term annual means from oceanographic datasets covering 1900-1997. We developed two models: Model 1, based on six annual surface temperature and salinity variables (mean, maximum and minimum, for each), and Model 2, based on the same six variables for surface temperature and salinity, as well as six annual average surface biochemical variables (Supplementary Table 15). The data for each environmental coverage layer were converted into standard normal variates in ArcGIS 9.3.1 prior to analysis.

Ecological Niche Modelling

ENMs were calibrated for each species using the maximum entropy algorithm Maxent v3.3.3k (Phillips et al. 2006). Maxent is a correlational algorithm that uses presence-only species occurrence data to estimate ecological niche parameters (Phillips et al. 2006). Maxent fits a suitability surface for the species of interest to the set of pixels across the study region, maximising the entropy of the probability distribution but constrained to return higher per-pixel suitability scores for pixels with environmental variable combinations most like those at which the species has been detected. That is, the resulting suitability surface is a raster map in which each pixel is scored regarding its similarity to climatic and biochemical conditions at sites where the species has been observed.

From Maxent suitability surfaces, distributions of suitable areas for each species were defined using a minimum training occurrence threshold (Pearson et al. 2007). We defined two thresholds for suitable and unsuitable areas for each species: the least presence threshold (LPT)—equal to the lowest probability at any occurrence location—and a more conservative 10th percentile training presence threshold, which is the highest suitability value that includes 90% of the calibration data. LPT is appropriate when there is no error in occurrence data that may be introduced during geo-referencing or identification, whereas the 10th percentile approach anticipates up to 10% error among occurrences. More complex methods of thresholding have demonstrated better skill in classifying suitable and unsuitable environments (Jímenez-Valverde and Lobo 2007; Liu et al. 2005), but our dataset lacks true absence data, so a simple omission-based technique was preferable (Bean et al. 2012). In addition, under this approach, the map of suitable conditions for a species is defined to be inclusive of all habitable conditions, and as such should include all known populations (Peterson 2014).

Testing predictive power of ENMs

This study examines whether the occurrences of three ciliate taxa follow a consistent and predictable environmental pattern in the Atlantic Ocean. As such, we tested whether models could predict independent subsets of occurrence data reliably. These tests covered the entire study area, using replicate random subsets of the occurrence data, and spatially stratified subsets of the occurrence data, 50% of localities for model calibration and 50% for testing for both models under default parameters, and choosing logistic output format with suitability values from 0 (unsuitable) to 1 (suitable). To avoid extrapolation in model features, no clamping or extrapolation was permitted (Owens et al. 2013); to avoid overly complex models, no hinge or threshold features were permitted.

For prediction across the study area, 10 bootstrap replicate runs with a maximum of 10,000 iterations were conducted using a random seed with 70% of occurrence points. For spatial stratification challenges, occurrence data were split spatially into quadrants above and below the median longitude and latitude of the occurrence data. From this spatial stratification, we developed the following three pairs of quadrants: west versus east of the median longitude, north versus south of the median latitude, and on-diagonal (upper left-hand and lower right-hand quadrants) versus off-diagonal (lower left-hand and upper right-hand quadrants). In each case, we developed both reciprocal predictions, testing the ability of ENMs to anticipate the spatial distribution of occurrence data in areas for which no sampling is available. Models were evaluated by applying partial receiver operating characteristic (ROC) statistics to the 50% subset of occurrences withheld from model development for testing. Area under the curve (AUC) ratios were calculated using the Partial ROC function available in CONABIO's NicheToolBox (http://shiny.conabio.gob.mx:3838/nichetoolb2/). Final models were developed using 50% of available data.

One concern in comparisons of models based on different sets of environmental data is that increased model complexity may lead to overfitting (Peterson et al. 2007; Radosavljevic and Anderson 2014). A recent tendency, as a consequence, has been to use the Akaike information criterion (AIC) as a means of comparing model likelihood values while penalizing complex (and ostensibly overfit) models (Warren and Seifert 2011). Specifically, we calculated the sample-size-corrected AICc statistic using ENMTools version 1.3.3 (Warren et al. 2010), and chose as the "best" model the one that had the lowest AICc values.

Results based on AIC (without the sample-size correction) and the alternative Bayesian information criterion (BIC) metric were similar, and so are not presented.

Niche difference and range restriction

To assess whether niches differed among taxa (which would imply range limits), we evaluated differences in niches occupied by each taxon using background similarity tests (Warren et al. 2008) available in the ENMTools R package version 0.1 (https://github.com/danlwarren/ENMTools). This test considers similarity between predicted geographic distributions among species, using two statistics, Schoener's *D* and a modified Hellinger's *I* metric, to quantify similarity. This test evaluates whether ENMs generated from two species are more different than expected when occurrences are drawn from the same underlying distribution across the region accessible to each species. The test allows specification of an area of analysis ("the background"), which we equate with the area accessible to a species over relevant time periods (Barve et al. 2011; Soberón and Peterson 2005). Numbers of points sampled from the background were set at observed sample sizes. In each test, 100 replicate analyses were performed to estimate probabilities associated with null hypotheses of niche similarity. The hypothesis that species were no more like each other than if points had randomly been drawn from the study area was rejected if observed similarity between models fell below the 5th of the null distribution.

RESULTS

The area under the curve (AUC) ratios for the independent *Parafavella gigantea* testing data were 1.77 and 1.79 for Model 1 and Model 2, respectively ($P < 0.05$; Table 1). The 10% threshold for both models predicted a broad potential distribution across the cold temperate zone of the North Atlantic north to about 79 / 81° N (Model 1 / Model 2) in the eastern

Atlantic, and about 67° N (both models) off the Greenland coast (Fig. 2) and the Labrador Sea. The southerly limit was at about 48° / 47° N in the eastern Atlantic, and 36° / 38° N off the US coast. Both models showed a more northerly prediction in the eastern Atlantic than in the western Atlantic, owing to the ameliorating effect of the Gulf Stream / North Atlantic Drift / Norwegian Current. Locality points with a borderline prediction for both models were principally from the poorly represented southeastern quadrant of the study area, the Greenland Sea in the northeast, and Labrador Sea and Hudson Bay in the northwest. The 10th percentile training presence fractional area predicted was about 22% of the study area for each model, whereas that predicted under least presence threshold (LPT) was about 43 / 44% of the study area.

The area under the curve ratios for the independent *Schmidingerella serrata* testing data were 1.83 and 1.81 for Model 1 and Model 2, respectively (Table 1). Both models predicted a broad distribution in the North Atlantic from about 74° / 71° N in the eastern Atlantic, around 65° N off the Greenland coast, and 52° N off the Canadian coast (Fig. 2**)**. The southern predicted limit extended from about 45° / 46° N in the European coast, curving slightly to the north in the mid-Atlantic, and about 36° N off the US coast. Most "borderline" points were either in the Labrador Sea and the northwest Atlantic or the southeast part of the distribution, including the Mediterranean. The least presence threshold prediction was similar between the two models, covering about 30% and 28% of the study area, respectively. The proportion of the study area predicted by the 10th percentile area was 18.3% by Model 1, compared to 18.9% in Model 2.

The area under the curve ratios for the independent *Z. pelagicum* testing data were 1.76 and 1.75 for Model 1 and Model 2, respectively (Table 1). Both predicted (at the 10th percentile training presence fractional area) a broad distribution in the North Cold Temperate Atlantic from about 72° / 71° N in the eastern Atlantic, respectively, to around 52° N off the Canadian coast (Fig. 2**)**. The southern predicted limit included much of the North Temperate Atlantic, to about 36° N for both models off the US coast, almost all the Mediterranean, and even some Tropical areas (to 18° / 15° N in the southeast Atlantic). Model 1's least presence threshold prediction was broader than that of Model 2 (66.7% / 61.4% of the study area, respectively). By contrast, the 10th percentile prediction was rather small for each, and, again, more restricted for Model 2: 30.0% / 24.5% of the study area. In each case, it was restricted to a central band (most of the North Temperate Atlantic). All 10 replicates of spatial stratification tests for each taxon and for both models showed significant ability to predict test points ($P <$ 0.05; see Table 1 and Supplementary Material).

All tests comparing niches of the three species rejected the null hypothesis of niche similarity between pairs of species when compared to a null distribution generated from the background region ($P < 0.05$; Figure 3, Supplementary Material Table 13). *Parafavella gigantea* was predicted to range in the cold temperate zone of the North Atlantic, extending to the Arctic Ocean, whereas the predicted distribution of *Z. pelagicum* was more southerly, extending to the Caribbean and the Equator; the predicted distribution of *S. serrata* was intermediate. The sample-size corrected Akaike information criterion (AICc) statistics, in all three cases, indicated that the simpler models based only on six annual surface temperature and salinity variables (i.e., Model 1) were preferable to the more complex models that included six biochemical variables as well (Model 2), as differentials in AICc were >90 in all three cases (Supplementary Material Table 14).

Because all major areas of the predicted-suitable area for each of the three species appeared to be inhabited--at least as far as limited sampling permitted us to conclude (e.g., in parts of the North Sea)--we conclude that these species likely have quite-excellent dispersal abilities.

DISCUSSION

Documentation of spatial patterns of microbial species remains undefined and contentious, owing to the complicated nature of detecting them over wide areas. Evidence for cosmopolitan distributions has been demonstrated for several protistan lineages (Cermeño and Falkowski 2009; Richards et al. 2005). However, many microbial species have broad but restricted distributional patterns (Bass et al. 2009; Bass et al. 2007; Foissner 2006); see Bass and Boenigk (2011) for a comprehensive review. This pattern of moderate endemicity is also seen in prokaryotic microbes (Noguez et al. 2005; Tamames et al. 2010). Some species are even endemic to discrete geographic areas and ecosystems (Foissner 2006). It is unclear to what extent microbial biogeography is obscured by poor understanding of species limits. Only one study (Aguilar et al. 2014) explored the spatial distribution of a single morphospecies of amoeba, *Badhamia melanospora*, using molecular genotyping and ENMs. The authors detected two geographically-structured groups of ribotypes for *B. melanospora*, each of which showed limited distributions, and concluded that this species is not cosmopolitan. Thus, it may be the case that morphospecies mask microbial diversity and biogeographic patterns (Gentekaki and Lynn 2010; Katz et al. 2011).

Our ecological niche models for three ciliophoran species clearly detected an environmental signal unique to each, such that each species occupies a distinct fundamental ecological niche. The area identified as suitable for each species varied greatly on broad spatial scales, ranging from about 18% of the study area for *S. serrata* to 25-30% for *Z. pelagicum* (10%

threshold). None of the species is likely to be ubiquitous across the entire study area and they showed statistically significant differences in their environmental characteristics. Our models demonstrated that the distribution of each species is constrained by environment, particularly maximum and minimum temperatures.

Our locality data were drawn from specimens from the North Atlantic Ocean, the region for which sampling is most dense and complete, although we are conscious that these species' ranges may extend more broadly. Our models were consistently able to predict non-North Atlantic locality data, with the following exception: *S. serrata* beyond the study area are found off the coast of Ecuador and Peru, the North Pacific, and the eastern Mediterranean. Although the latter two sets of points were well predicted by our models, the Ecuador-Peru points were in an area with higher temperature and salinity than those used to train our *S. serrata* model, and thus fell outside the ecological niche estimated in our models. Still, the *S. serrata* models showed a good fit for the North Atlantic locality data, and provide a parsimonious prediction. The non-Atlantic *S. serrata* points may form part of the natural distribution of the species, or represent recent, perhaps human-mediated, invasion. For instance, the points from the Ecuador-Peru coast might be explained by natural dispersal and colonization by the species following the opening of the Panama Canal (similar to the Lesseps immigrant Indo-Pacific species that entered the Mediterranean via the Suez Canal), or direct transport in a ship's ballast water (Foissner 2011). However, we know of no evidence that can clarify which is the case.

Additional documented populations for the other two ciliates comprised occurrences that were highly consistent with our predictions. *Parafavella gigantea* has been detected in the North Pacific and along the Russian northern coastal areas, consistent with our North Atlantic prediction. The *P. gigantea* models showed a good fit for the North Atlantic locality data, in that they recovered the areas where the species has been detected, without predicting an area too broad to be credible; however, the North Pacific was not included in the model predictions. Additional locality data for *P. gigantea* compiled in a recent study (Dolan et al. 2017) comprised 38 additional locality points for the species, and all were anticipated in our least presence threshold model (significant $P < 0.05$). Few locality points for *Z. pelagicum* were available from outside the North Atlantic ($N = 6$), all from the North Pacific. An exhaustive literature search for additional records revealed no additional locality data for *Z. pelagicum*. Our model does not provide a prediction for the North Pacific distributions of these three species. However, climatic and biochemical characteristics of the North pacific are broadly similar to those of the North Atlantic, and we anticipate that these locality points would be predicted by broader analyses.

The *S. serrata* populations occurring between the Galapagos and the Peruvian coast occupy a tropical niche, quite distinct from the cold temperate distribution of populations detected in the North Atlantic. The true environmental tolerances of this species may indeed be broader than the model we have developed here. If so, our ENM requires locality data from areas with higher temperature and salinity to reflect this ecological tolerance, and our model would thus fail to represent the full fundamental ecological niche. Alternately, the tropical Pacific population may have ecological tolerances distinct from those of the North Atlantic population, and may even represent a distinct species within the morphospecies, suggesting a need for investigation of the ecophysical constraints and genetic distinctiveness of the two populations. Multiple phylogeographic studies on diverse protists, but not ciliate groups (Aguilar et al. 2014; Bass et al. 2007; Evans et al. 2009), have shown distinct geographic ribotypes in morphological species. To the best of our knowledge, no studies have assessed phylogeographic patterns in any of the taxa investigated here. This study demonstrates that three common, morphologically conspicuous and widespread morphospecies occupy distinct geographic distributions and ecological niches in the North Atlantic.

Our occurrence data were collected over a relatively long time period, just over a century (1903-2008). Global climate change and short-term regional climatic phenomena, including the North Atlantic Oscillation, the Atlantic Multidecadal Oscillation, and the El Niño-Southern Oscillation, may have significant effects on distributions of the three ciliate species analyzed. The environmental data layers used as input to our models were averaged over much of the twentieth century, which overlaps well with the temporal provenance of the occurrence data. These considerations may introduce minor biases, nonetheless, owing to a rise in global mean surface temperature of about 0.1°C toward the end of the century (Hansen et al. 2010). More generally, the environmental signature for a given pixel was assigned an average value in our analyses, rather than the values for the year in which the sample was collected. Such models with finer temporal resolution can and should be developed to resolve this issue, considering the appropriate environmental regime over the period of specimen collection. We aim to take this "next step" when we can develop both the relevant data layers and occurrence data that are sufficiently rich and dense to permit such analyses. For now, however, our models provide a first estimate of environmental envelopes for these three ciliate species, with the caveat that predictions may involve a slight northern bias that is unlikely to affect the argument regarding the ubiquity hypothesis versus moderate endemicity.

Predictions across the study area based on different environmental data sets were similar for each species. In all models, minimum and maximum temperature contributed the most to the prediction, although the percent contribution of these two temperature coverages was reduced by the introduction of additional data layers for the Model 2 series. The additional coverages of phosphate and silicate content were most important for *P. gigantea* and *S. serrata*, whereas oxygen saturation and silicate content were most important for *Z. pelagicum*. In contrast, the salinity layers and the apparent oxygen usage layer were relatively unimportant in all models. Model 2 provided finer resolution for each taxon, predicting less of the study area and omitting fewer of the independent evaluation points at the least presence threshold, suggesting a better model. However, the model complexity evaluation indicated clearly that the increased detail of Model 2 for each species did not outweigh the negative effects of increasing model complexity and dimensionality, such that Model 1 was preferable for each species. A clear link exists between the biogeography of each of the species and climatic conditions, so distributions of the species will likely change with climate change. This possibility merits further investigation, considering the current scientific focus on anthropogenic global climate change, and other short-term regional climatic phenomena. This study is a first reference point for the distribution of each species, and can serve as a benchmark against which to compare future distributional patterns.

CONCLUSIONS

This study represents a first effort to describe the spatial and environmental distributions, using ecological niche models, of three species of ciliated protists, a group of microorganisms that are essential to marine carbon cycling and trophic chains, marine biodiversity, and even to the fishing industry (Caron and Countway 2009; Lom and Dyková 1992) but for which geographic distributions are poorly documented. This study thus serves as a first reference for the distribution of each species, and as a benchmark against which to compare potential future distributional patterns. Although future work remains to be done to refine these models, particularly to consider climate variability, our findings point clearly to a situation in which each of the studied species has a unique environmental signature and geographic distribution.

REFERENCES

Aguilar, M., A. M. Fiore-Donno, C. Lado, and T. Cavalier-Smith. 2014. Using environmental niche models to test the 'everything is everywhere' hypothesis for *Badhamia*. ISME J. **8:** 737-745.

Aguilar, M., and C. Lado. 2012. Ecological niche models reveal the importance of climate variability for the biogeography of protosteloid amoebae. ISME J **6:** 1506-1514.

Azam, F., T. Fenchel, J. G. Field, J. S. Gray, L. A. Meyerreil, and F. Thingstad. 1983. The ecological role of water-column microbes in the sea. Mar. Ecol.-Prog. Ser. **10:** 257-263.

Barve, N. and others 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol. Model. **222:** 1810-1819.

Bass, D., and J. Boenigk. 2011. Everything is everywhere: a twenty-first century de/reconstruction with respect to protists., p. 88-110. *In* D. Fontaneto [ed.], Biogeography of microscopic organisms: is everything small everywhere? Cambridge University Press.

Bass, D. and others 2009. A molecular perspective on ecological differentiation and biogeography of Cyclotrichiid Ciliates. J. Eukaryot. Microbiol. **56:** 559-567.

Bass, D., T. A. Richards, L. Matthai, V. Marsh, and T. Cavalier-Smith. 2007. DNA evidence for global dispersal and probable endemicity of protozoa. BMC Evol. Biol. **7:** e162.

Bean, W. T., R. Stafford, and J. S. Brashares. 2012. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. Ecography **35:** 250-258.

Beijerinck, M. W. 1913. De infusies en de ontdekking der backteriën. Müller (Reprinted in Verzamelde geschriften van M.W. Beijerinck, vijfde deel, pp. 119–140. Delft, 1921).

Bentlage, B., A. T. Peterson, N. Barve, and P. Cartwright. 2013. Plumbing the depths: extending ecological niche modelling and species distribution modelling in three dimensions. Glob. Ecol. Biogeogr. **22:** 952-961.

Boenigk, J., K. Pfandl, T. Garstecki, H. Harms, G. Novarino, and A. Chatzinotas. 2006. Evidence for geographic isolation and signs of endemism within a protistan morphospecies. **72:** 5159-5164.

Brun, P. and others 2015. Ecological niches of open ocean phytoplankton taxa. Limnol. Oceanogr. **60:** 1020-1038.

Caron, D. A., and P. D. Countway. 2009. Hypotheses on the role of the protistan rare biosphere in a changing world. Aquat. Microb. Ecol. **57:** 227-238.

Cermeño, P., and P. G. Falkowski. 2009. Controls on diatom biogeography in the ocean. Science **325:** 1539-1541.

Dolan, J. R. 2016. Planktonic protists: little bugs pose big problems for biodiversity assessments. **38:** 1044-1051.

Dolan, J. R., R. W. Pierce, and E. J. Yang. 2017. Tintinnid ciliates of the marine microzooplankton in Arctic Seas: a compilation and analysis of species records. Polar Biol. **40:** 1247-1260.

Evans, K. M., V. A. Chepurnov, H. J. Sluiman, S. J. Thomas, B. M. Spears, and D. G. Mann. 2009. Highly differentiated populations of the freshwater diatom *Sellaphora capitata* suggest limited dispersal and opportunities for allopatric speciation. Protist **160:** 386-396.

Fenchel, T. 2008. The microbial loop-25 years later. J. Exp. Mar. Biol. Ecol. **366:** 99-103.

Field, C. B., M. J. Behrenfeld, J. T. Randerson, and P. Falkowski. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. Science **281:** 237-240.

Filker, S., R. Sommaruga, I. Vila, and T. Stoeck. 2016. Microbial eukaryote plankton communities of high-mountain lakes from three continents exhibit strong biogeographic patterns. Mol. Ecol. **25:** 2286-2301.

Finlay, B. J., J. O. Corliss, G. Esteban, and T. Fenchel. 1996. Biodiversity at the microbial level: the number of free-living ciliates in the biosphere. Q. Rev. Biol. **71:** 221-237.

Foissner, W. 1999. Protist diversity: estimates of the near-imponderable. Protist **150:** 363-368.

---. 2006. Biogeography and dispersal of micro-organisms: a review emphasizing protists. Acta Protozool. **45:** 111-136.

---. 2011. Dispersal of protists: the role of cysts and human introductions, p. 61 – 87. *In* D. Fontaneto [ed.], Biogeography of microscopic organisms: is everything small everywhere? Cambridge University Press.

Foissner, W., A. Chao, and L. A. Katz. 2008. Diversity and geographic distribution of ciliates (Protista: Ciliophora). Biodivers. Conserv. **17:** 345-363.

Gao, F. and others 2016. The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the phylum Ciliophora (Eukaryota, Alveolata). Sci. Rep. **6:** e24874.

Gentekaki, E., and D. H. Lynn. 2010. Evidence for cryptic speciation in *Carchesium polypinum* Linnaeus, 1758 (Ciliophora: Peritrichia) inferred from mitochondrial, nuclear, and morphological markers. J. Eukaryot. Microbiol. **57:** 508-519.

Hansen, J., R. Ruedy, M. Sato, and K. Lo. 2010. Global surface temperature change. Rev. Geophys. **48:** RG4004.

Jímenez-Valverde, A., and J. M. Lobo. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. Acta Oecol.-Int. J. Ecol. **31:** 361-369.

Katz, L. A., J. DeBerardinis, M. S. Hall, A. M. Kovner, M. Dunthorn, and S. V. Muse. 2011. Heterogeneous rates of molecular evolution among cryptic species of the ciliate morphospecies *Chilodonella uncinata*. J. Mol. Evol. **73:** 266-272.

Langer, M. R., A. E. Weinmann, S. Lotters, J. M. Bernhard, and D. Rodder. 2013. Climate-driven range extension of *Amphistegina* (Protista, Foraminiferida): models of current and predicted future ranges. PLoS ONE **8:** e54443.

Laval, M. 1968. Zoothamnium pelagicum du Plessis. Cilié péritricheplanctonique: morphologie, croissance et comportament. Protistologica **4:** 333-363.

Liu, C. R., P. M. Berry, T. P. Dawson, and R. G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography **28:** 385-393.

Logares, R., J. F. Mangot, and R. Massana. 2015. Rarity in aquatic microbes: placing protists on the map. Res Microbiol **166:** 831-841.

Lom, J., and I. Dyková. 1992. Protozoan parasites of fishes, p. 315. Developments in aquaculture and fisheries science. Elsevier Science.

NOAA. 1999. World Ocean Atlas 1998.

Noguez, A. M., H. T. Arita, A. E. Escalante, L. J. Forney, F. García-Oliva, and V. Souza. 2005. Microbial macroecology: highly structured prokaryotic soil assemblages in a tropical deciduous forest. Glob. Ecol. Biogeogr. **14:** 241-248.

Owens, H. L. and others 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. Ecol. Model. **263:** 10-18.

Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. T. Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J. Biogeogr. **34:** 102-117.

Peterson, A. and others 2011. Ecological niches and geographic distributions. Princeton University Press.

Peterson, A. T. 2014. Mapping disease transmission risk: enriching models using biogeography and ecology. Johns Hopkins University Press.

Peterson, A. T., M. Papeş, and M. Eaton. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography **30:** 550-560.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. **190:** 231-259.

Radosavljevic, A., and R. P. Anderson. 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. J. Biogeogr. **41:** 629-643.

Richards, T. A., A. A. Vepritskiy, D. E. Gouliamova, and S. A. Nierzwicki-Bauer. 2005. The molecular diversity of freshwater picoeukaryotes from an oligotrophic lake reveals diverse, distinctive and globally dispersed lineages. Environ. Microbiol. **7:** 1413-1425.

Saupe, E. E., J. R. Hendricks, A. T. Peterson, and B. S. Lieberman. 2014. Climate change and marine molluscs of the western North Atlantic: future prospects and perils. J. Biogeogr. **41:** 1352-1366.

Segovia, B. T. and others 2017. Common and rare taxa of planktonic ciliates: influence of flood events and biogeographic patterns in neotropical floodplains. Microb. Ecol. **74:** 522-533.

Soberón, J., and A. T. Peterson. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodiversity Informatics **2:** 1-10.

Tamames, J., J. J. Abellán, M. Pignatelli, A. Camacho, and A. Moya. 2010. Environmental distribution of prokaryotic taxa. BMC Microbiol. **10:** article 85.

Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution **62:** 2868-2883.

Warren, D. L., R. E. Glor, and M. Turelli. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. Ecography **33:** 607-611.

Warren, D. L., and S. N. Seifert. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecol. Appl. **21:** 335-342.

Wiley, E. O., K. M. McNyset, A. T. Peterson, C. R. Robins, and A. M. Stewart. 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography **16:** 120–127.

ACKNOWLEDGEMENTS

We dedicate this paper to Dr. John Clamp, who was an integral part of the research and writing of this paper. Dr. Clamp passed while this manuscript under review. He was a professor at North Carolina Central University for more than 30 years, devoting his outstanding research to evolutionary biology, and particularly the study of ciliated protists. He was a larger-than-life character and will be greatly missed as friend and mentor. This work was supported by the Ministerio de Economía y Competitividad, Española (MINECO, grant number CGL2013-40851-P/ BOS) to MMC, and travel support from the Research Coordination Network for Biodiversity of Ciliates (National Science Foundation grant number DEB 1136580) from the U.S. National Science Foundation to JC. RAJW was supported by the Crafoord Foundation, Sweden (grant numbers 20160971 and 20170671), while completing this manuscript.

SUPPLEMENTARY MATERIAL

This document provides a summary for all data management and spatial stratification procedures followed in developing predictive models for three ciliate species; also provided are heuristic estimate of relative contributions of the environmental variables, maps of all 36 spatial stratification predictions, results of the tests to determine range restriction using ENMTools background similarity, and a summary environmental layers used for model development.

DATA ACCESSIBILITY

Primary occurrence data are fully and openly accessible via the Global Biodiversity Information Facility (GBIF; http://www.gbif.org/) and Ocean Biogeographic Information System (OBIS; http://www.iobis.org/). Environmental coverages were drawn from National

Oceanic and Atmospheric Administration’s (NOAA) World Ocean Atlas (NOAA 1999). Full results are available in Supplementary Material.

FIGURE LEGENDS

Figure 1. Global map showing distribution records of *Parafavella gigantea* (A; *N* = 4495), *Schmidingerella serrata* (B; *N* = 1536), and *Zoothamnium pelagicum* (C; *N* = 1738)] based on data drawn from the Ocean Biogeographic Information System (OBIS; http://www.iobis.org), and Global Biodiversity Information Facility (GBIF; http://www.gbif.org/), both accessed 10.12.2015. The sample sizes that we used for model development were greatly reduced as we included only the locality data from vouchered reference material, and removed all duplicate records of a species at any given site.

Figure 2. Predictions of suitable areas across the study area for each of three species, based on models calibrated using two environmental coverage sets.

Figure 3. Model 1 background similarity tests comparing the niche occupied by each taxon using the background similarity tests available in the ENMTools R package version 0.1. Schoener's *D* and a modified Hellinger's *I* metric, are used to quantify the similarity of two probability distributions. In each case, the niche occupied by each taxon is more different than expected from the study region. Model 2 background similarity tests are shown in Supplementary Material Figure 14.

TABLES

Table 1. Summary table of ecological niche modelling results for both models of three species of ciliates. Results of spatial stratification models are shown in Supplementary Material.

Species	*N* training / testing points [1]	Testing AUC ratio [2]	Fractional predicted area LPT [3]	Fractional predicted area 10% PT [4]
All data predictive models				
P. gigantea 1	184 / 184	1.77	0.430	0.221
P. gigantea 2	184 / 184	1.79	0.442	0.220
S. serrata 1	68 / 67	1.83	0.304	0.183
S. serrata 2	68 / 67	1.81	0.283	0.189
Z. pelagicum 1	140 / 139	1.76	0.667	0.299
Z. pelagicum 2	140 / 139	1.75	0.614	0.245

[1] *N*, number of points used for model training / testing.

[2] Training/ testing AUC data, area under the curve of the receiver operating characteristic calculated using NicheToolBox (http://shiny.conabio.gob.mx:3838/nichetoolb2/).

[3] LPT, Least Prediction Threshold.

[4] 10% PT, predictive threshold that excludes the 10% most outlying points.

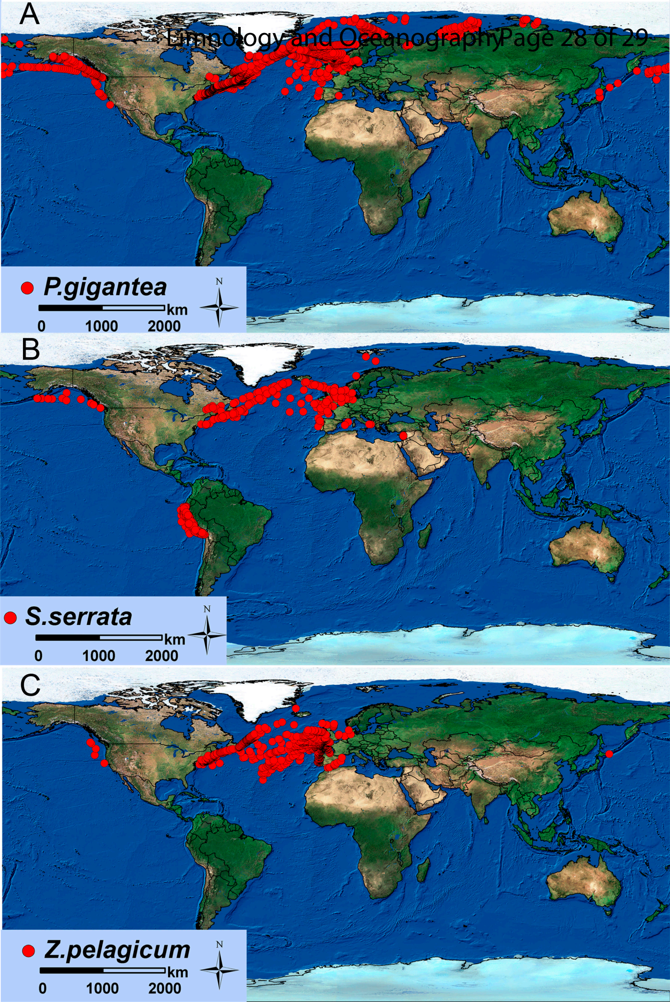
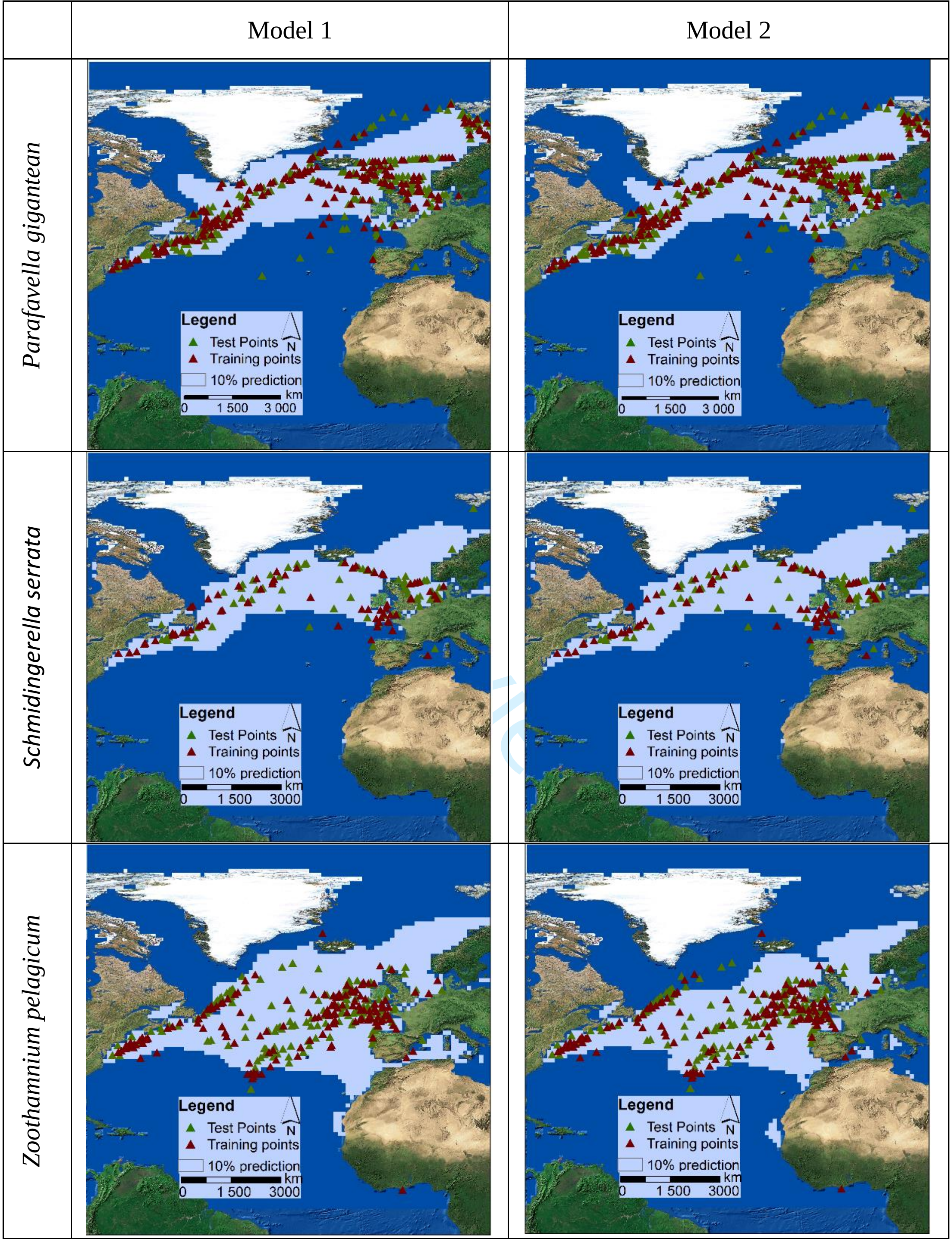

A
● P.gigantea
km
0
1000
2000
N
B
● S.serrata
km
0
1000
2000
N
C
● Z.pelagicum
km
0
1000
2000
N

Figure 2.


Model 1
Model 2
Parafavella gigantean
Schmidingerella serrata
Zoothamnium pelagicum
Legend
Test Points
Training points
10% prediction
N
km
0
1 500
3 000
3000

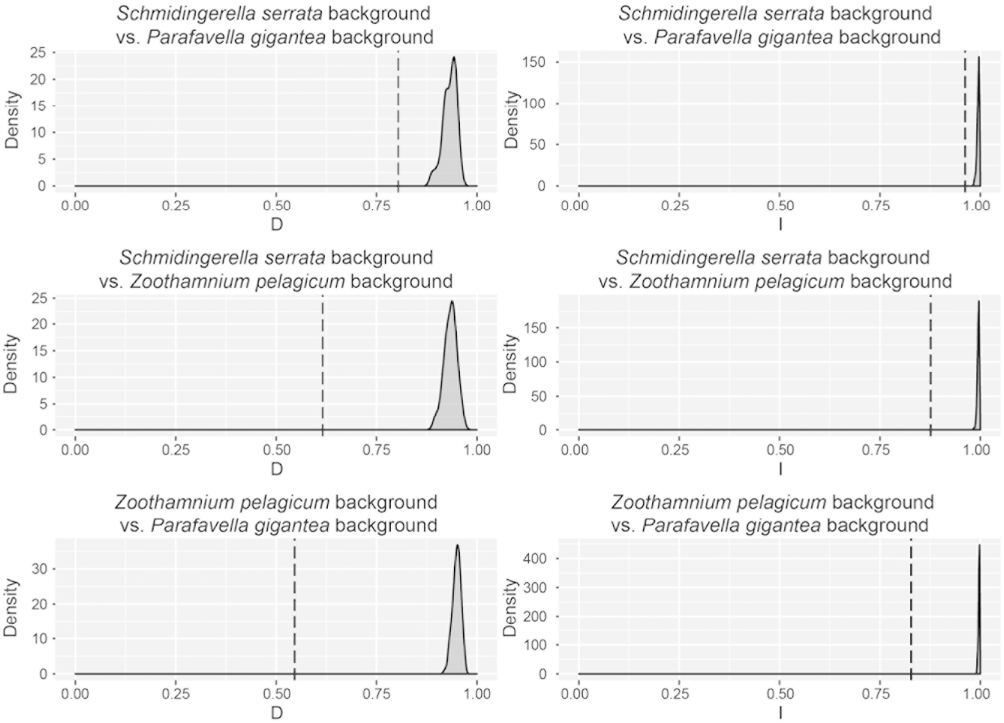


Figure 3. Model 1 background similarity tests comparing the niche occupied by each taxon using the background similarity tests available in the ENMTools R package version 0.1. Schoener’s D and a modified Hellinger's I metric, are used to quantify the similarity of two probability distributions. In each case, the niche occupied by each taxon is more different than expected from the study region. Model 2 background similarity tests are shown in Supplementary Material Figure 14.