The ISME Journal (2015) 9, 256–264

www.nature.com/ismej

ORIGINAL ARTICLE

The niche of an invasive marine microbe in a subtropical freshwater impoundment

K David Hambright[1,2,3], Jessica E Beyer[1,2], James D Easton[1,3], Richard M Zamor[1,2], Anne C Easton[1,3] and Thayer C Hallidayschult[1,2]

[1]*Plankton Ecology and Limnology Laboratory, Department of Biology, University of Oklahoma, Norman, OK, USA;* [2]*Program in Ecology and Evolutionary Biology, Department of Biology, University of Oklahoma, Norman, OK, USA and* [3]*Biological Station, Department of Biology, University of Oklahoma, Norman, OK, USA*

Growing attention in aquatic ecology is focusing on biogeographic patterns in microorganisms and whether these potential patterns can be explained within the framework of general ecology. The long-standing microbiologist's credo 'Everything is everywhere, but, the environment selects' suggests that dispersal is not limiting for microbes, but that the environment is the primary determining factor in microbial community composition. Advances in molecular techniques have provided new evidence that biogeographic patterns exist in microbes and that dispersal limitation may actually have an important role, yet more recent study using extremely deep sequencing predicts that indeed everything is everywhere. Using a long-term field study of the 'invasive' marine haptophyte *Prymnesium parvum*, we characterize the environmental niche of *P. parvum* in a subtropical impoundment in the southern United States. Our analysis contributes to a growing body of evidence that indicates a primary role for environmental conditions, but not dispersal, in the lake-wide abundances and seasonal bloom patterns in this globally important microbe.
The ISME Journal (2015) **9,** 256–264; doi:10.1038/ismej.2014.103; published online 20 June 2014

Introduction

'Everything is everywhere, but, the environment selects' (Baas-Becking, 1934; de Wit and Bouvier, 2006) suggests that dispersal is not limiting for microbes, but that the environment is the primary factor determining whether a particular microbe is actively participating in a given community. If true, microbial biogeographic distributions may be reflective only of technical limitations in detection, whereas their active or meaningful participation in a community will depend on the suitability of a given habitat to foster positive population growth (Gibbons *et al.*, 2013; Hambright *et al.*, 2014). The marine haptophyte *Prymnesium parvum* Carter is considered to be a globally invasive species in many freshwater systems in which it is now a community member and often dominant. However, recent study has suggested that *P. parvum* is not dispersal limited, as extremely low-density populations have been detected in habitats that do not experience *P. parvum* blooms, many of which are directly downstream of *P. parvum* bloom sites (Zamor *et al.*, 2012). This phenomenon reflects the general notion of the rare or dormant microbial biosphere in microbes (Sogin *et al.*, 2006; Caron and Countway, 2009; Jones and Lennon, 2010; Gibbons *et al.*, 2013), and leads to the hypothesis that environmental conditions that foster high growth in *P. parvum* relative to other microbial constituents are the principle determinants of *P. parvum* blooms in inland freshwater systems.

P. parvum blooms and fish kills have been observed in many inland aquatic systems worldwide (Granéli *et al.*, 2012), including waterbodies, mostly reservoirs, in at least 20 US states (Roelke *et al.*, 2011; Hambright, 2012). This apparent incredible range expansion since the first North American report (Pecos River of southern Texas) in the 1980s (James and De La Cruz, 1989) represents an interesting enigma—while *P. parvum* has an ability to thrive across a broad range of environmental conditions (Edvardsen and Paasche, 1998), the conditions found in most North American bloom sites tend to be far removed from optimal conditions described from the laboratory study (Baker *et al.*, 2009). Most inland blooms of *P. parvum* in the southwestern United States have occurred at relatively low salinities (1–3 partial salinity units, psu) during winters when temperatures range from 10 °C to 20 °C, yet laboratory studies suggest that inland strains of *P. parvum* are well suited to high salinities (8–30 psu) and temperatures (20–30 °C) (Baker *et al.*, 2007, 2009; Hambright *et al.*, 2010, 2014; Roelke *et al.*, 2011; Patiño *et al.*, 2014). This apparent

Correspondence: KD Hambright, Department of Biology, University of Oklahoma, 730 Van Vleet Oval, Norman, OK 73072, USA.
E-mail: dhambright@ou.edu
Received 22 March 2014; revised 9 May 2014; accepted 13 May 2014; published online 20 June 2014

paradox has led to speculation that its toxigenic capabilities provide a competitive edge to *P. parvum* over other algae, allowing blooms to develop during periods of stress, such as created by low nutrient availabilities (for a review, see Granéli *et al.*, 2012). However, while toxicity may have important roles in predator avoidance and heterotrophy in this unicellular mixotroph, toxin production is unlikely to provide a competitive advantage to *P. parvum* to the degree necessary to lead to bloom formation under suboptimal environmental conditions (Jonsson *et al.*, 2009; Remmel and Hambright, 2012).

Because it is a microbial eukaryote that can occur in immense numbers, encyst and be passively transported, its dispersal capabilities are potentially unlimited (sensu Finlay, 2002; but see Martiny *et al.*, 2006; Hanson *et al.*, 2012). Indeed, examination of the *P. parvum* distribution in Lake Texoma (Oklahoma–Texoma) reveals that *P. parvum* is dispersed throughout the lake, yet blooms and fish kills are common only in areas in which environmental conditions are conducive for growth (Hambright *et al.*, 2010; Zamor *et al.*, 2012). Thus, the recent expansion of *P. parvum* into new habitats, which is typically noted only after a bloom and fish kill, suggests that habitats with suitable environmental conditions (e.g., elevated nutrients and salinities, see below) may be becoming more abundant, particularly in the southwestern United States, where quality of surface water resources is subjected to the pressures of climate change and increasing freshwater demands that accompany growing human populations and development (Roelke *et al.*, 2011).

Here we report results of a long-term study of *P. parvum* in subtropical Lake Texoma (Oklahoma–Texas, USA), in which *P. parvum* blooms are now commonplace. We use these data to test the hypothesis that environmental conditions that foster high growth in *P. parvum* are the principle determinants of *P. parvum* blooms in the lake. Our analysis provides further support for the Baas-Becking hypothesis 'Everything is everywhere, but, the environment selects', as we identify a primary role for environmental conditions, but not dispersal, in the lake-wide distributions and bloom patterns in *P. parvum*.

Materials and methods

Study site

Lake Texoma (Figure 1), an impoundment of the Red and Washita Rivers, was constructed in 1944 for flood control, hydropower generation and recreation. The lake is the 12th largest reservoir in the United States (at normal pool elevation) with a surface area of 360 km^2, and mean and maximum depths of 8.7 and 26 m. The lake watershed occupies 87 500 km^2 of the high plains of Texas and the rolling plains of Texas and Oklahoma. Owing to this extremely large watershed (watershed area: lake surface area = 243), nutrient loading to the lake is high and the lake is eutrophic to hypertrophic, depending on season and location within this complex, dendritic reservoir (Oklahoma Water Resources Board, 2010). The watershed also contains abundant deposits of calcium carbonate, halite, gypsum, anhydrite and other Permian-Salado evaporites (Ground and Groeger, 1994), which lead to salinities that often exceed 1 psu (defined here as having a specific conductivity equivalent to 1 g l^{-1} NaCl), the general limit for fresh water. The phytoplankton are often dominated by filamentous and colonial cyanobacteria and the lake is home to a diverse array of zooplankton, including multiple daphniids, numerous copepod and rotifer species (Franks *et al.*, 2001; Hambright *et al.*, 2010), and more than 50 fish species, many of which are recreationally important (Matthews *et al.*, 2004).

Lake sampling

We sampled eight littoral sites along the northern Oklahoma side of Lake Texoma in marinas and coves, and five pelagic stations in the Red and Washita River arms of the lake, as well as at the dam (Figure 1). Littoral samples were taken in shallow (usually <1 m) waters, usually from boat ramps or docks, if present. Temperature and salinity (as specific conductance), and dissolved oxygen, chlorophyll and phycocyanin concentrations were measured *in situ* with YSI and Hydrolab sondes, using a single mid-water point for littoral sites, and surface-to-bottom water column profiles for pelagic sites. Water samples (1 l; mid-water samples for littoral sites or 6 to 10 m, depending on depth, integrated samples for pelagic sites) were collected in acid-washed, deionized- and sample-rinsed Nalgene bottles, stored on ice in the field and refrigerated in the laboratory for subsequent analyses of *P. parvum* abundances by quantitative PCR (Zamor *et al.*, 2012) (microscopy (hemocytometer, 6–12 fields per sample) was used in 2006–2007), and for concentrations of chlorophyll (acetone extraction), total nitrogen (TN) and total phosphorus (TP) (flow-injection autoanalysis) (American Public Health Association, 1998) and pH. Lake surface elevation data were taken from the 0800 hours measurements recorded at the Denison Dam (Army Corps of Engineers–Tulsa District, 2012). Further details of sampling and sample analyses and additional parameters monitored can be found in Hambright *et al.* (2010).

Statistical analyses

All statistical analyses were carried out in R (version 3.0.2; R Development Core Team, 2013). We used logistic regression analysis to examine relationships between *Prymnesium* distributions within the lake (categorized as either presence–absence or

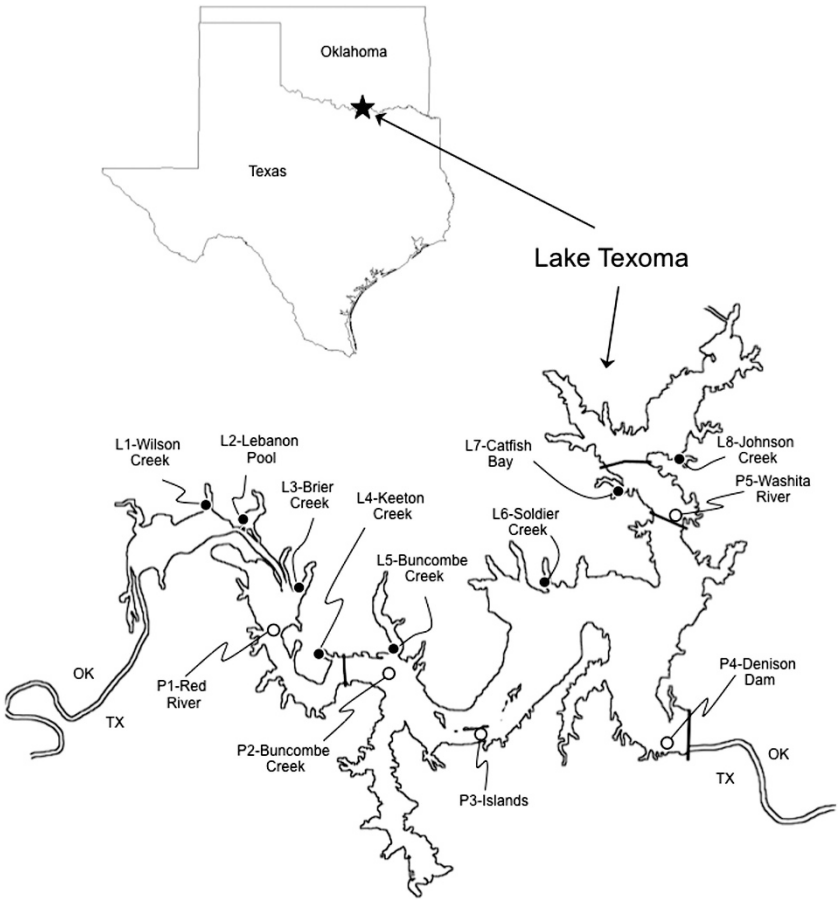

Figure 1 Schematic of Lake Texoma location and sampling sites. L1–L8 are littoral sites and P1–P5 are pelagic sites.

bloom–no bloom, where bloom was quantified as $\geqslant 10\,000$ cells per ml) and environmental conditions, in which the predictor variables (identified from earlier studies; (Hambright *et al.*, 2010; Zamor *et al.*, 2012) consisted of temperature (°C), salinity (psu), TN (mg N l^{-1}), TP (mg P l^{-1}), molar TN:TP and site type (littoral or pelagic); $N = 1181$ (rms package; version 4.0-0; Harrell, 2013). This approach of dichotomizing the response variable and using logistic as opposed to a continuous generalized linear modeling approach was taken because these response variables best fit our question (i.e., comparing variables predicting presence and blooms) and because our data are zero inflated owing to a preponderance of absence and non-bloom situations. Preliminary analysis of environmental predictors indicated correlations between TN and TP and between TP and TN:TP (see Supplementary Information and Supplementary Figure S1), but variance inflation factors were sufficiently low (all <5) to rule out potential effects of multicollinearity on parameter estimates or goodness-of-fit metrics (Davis *et al.*, 1986). We quantified goodness of fit of logistic regression models using the model likelihood ratio χ^2; Nagelkerke's R^2_N as a measure of explained variance; the C index, a measure of concordance of model prediction, which ranges from 0.5 (random) to 1 (perfect), >0.8 indicates a useful model; and Somer's D_{xy}, which measures the rank correlation difference between concordance and discordance of predictions to model, and ranges from 0 (random) to 1 (perfect). Logistic regression models were further validated using bootstrapping ($n = 100$) to measure the degree of overfitting; slope ranges from 0 (extreme overfitting) to 1 (no overfitting). Predictors of *P. parvum* presence and blooms were further elucidated using classification tree analyses (rpart package; version 4.1-1; Therneau *et al.*, 2013). Classification trees complement logistic regression well because they can capture complex interactions and non-monotonic relationships between predictors and response variables, and display these complex relationships in easily interpreted plots (De'ath and Fabricius, 2000). Classification trees predicting either presence or blooms of *P. parvum* were assembled using the default settings of rpart. To minimize overfitting, trees were pruned to minimize cross-validated error (for details, see De'ath and Fabricius, 2000).

Although chlorophyll, dissolved oxygen and pH are routinely measured in water quality monitoring programs owing to ease of measurement, any relationship between *P. parvum* and these predictors may be confounded by the direct influence of algal

growth and abundances on these three variables. These variables were strongly correlated with other variables used in the models (Supplementary Figure S1), although, as before, the variance inflation factors were sufficiently low (all <5) to suggest multicollinearity was not likely a substantial problem. In addition to the previously described predictors, our previous analysis (Hambright *et al.*, 2010) suggested that water level elevation might be a good predictor of *P. parvum* abundances in Lake Texoma, and Grover *et al.* (2012) have suggested a potential negative relationship between cyanobacterial and golden algal abundances. Therefore, we also ran logistic regression and classification tree analyses in which chlorophyll, DO, pH and elevation ($N=1181$), and chlorophyll, DO, pH, elevation and cyanobacterial abundances, as phycocyanin concentrations ($N=560$) were included as predictor variables in an effort to ascertain any increase in predictive capabilities for *P. parvum* presence and blooms by the inclusion of these additional variables. Full and reduced models were compared using the corrected Akaike Information Criterion (for details, see Burnham and Anderson, 2002).

Results

P. parvum bloomed in at least one cove of Lake Texoma in seven of the nine winters during 2004–2012 (Figure 2). Highest densities were observed at littoral sites, such as Lebanon Pool, on the western Red River arm of the lake, and lower densities further downstream and on the Washita River arm of the lake (see also Hambright *et al.*, 2010; Zamor *et al.*, 2012). As demonstrated previously, prevalent environmental conditions during the bloom years included relatively high salinities and nutrients. Net maximum population growth rates of *P. parvum* during the initial bloom phase in Lebanon Pool (calculated as the slope of the log of *P. parvum* densities over time during the initial growth period of each bloom) scaled positively with salinity, but were not significantly related (all $P>0.05$) to TN:TP, pH or temperature (Figure 3).

Logistic regression analysis revealed that the presence of *P. parvum* was best predicted by salinity, site type, temperature, TP and TN:TP; TN was not a significant predictor (Table 1). Addition of chlorophyll, DO, pH and water elevation to the model resulted in a slight improvement of the overall model fit, with chlorophyll, pH and water elevation being added as significant predictors of *P. parvum* presence (assessed as Nagelkerke's R^2; Supplementary Table S1). Model comparison using corrected Akaike Information Criterion suggested that the second, 10-predictor model (including chlorophyll, DO, pH and water elevation) was strongly preferred over the 6-predictor model excluding these factors (Supplementary Table S3). Addition of phycocyanin to the analysis resulted in a smaller data set ($n=560$) and did not improve either 10- or 6-predictor model results (Supplementary Table S4).

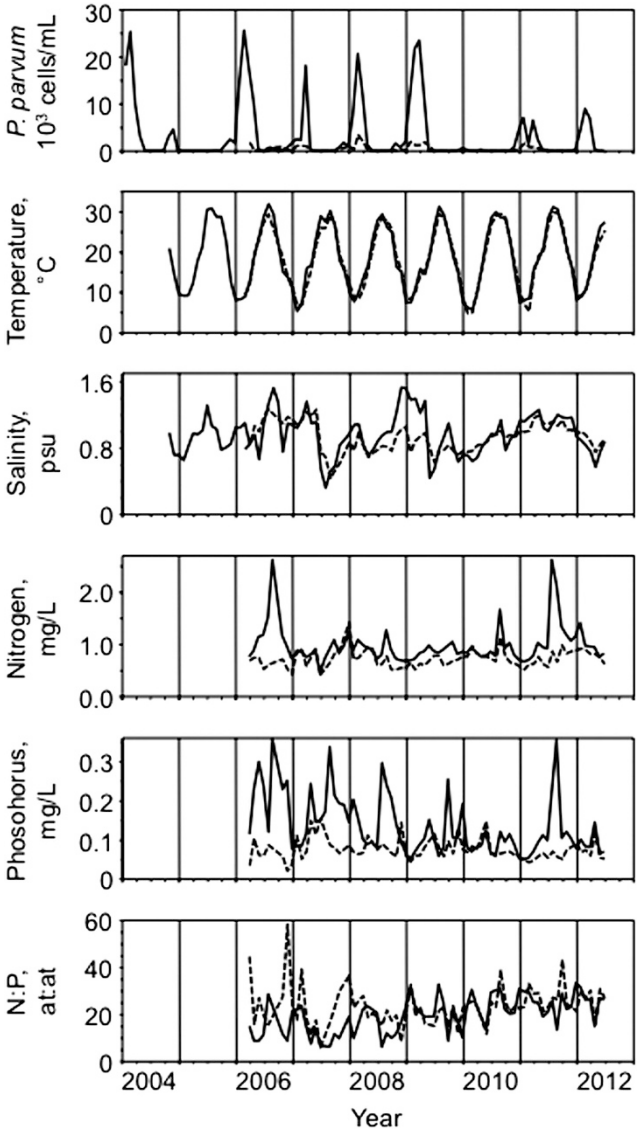

Figure 2 Monthly mean abundances of *P. parvum* and temperature, salinity (as psu), total nitrogen and total phosphorus concentrations, and molar N:P in Lake Texoma littoral (black line) and pelagic (dashed line) stations during 2004–2012.

Blooms of *P. parvum* (defined as $\geqslant 10\,000$ cells per ml) were best predicted by salinity, site, temperature, TN, TP and TN:TP (Table 2). Addition of chlorophyll, DO, pH and water elevation to the model also resulted in a slight improvement of the overall bloom model, with chlorophyll and water elevation being added as significant predictors of *P. parvum* blooms (assessed as Nagelkerke's R^2; Supplementary Table S2). Corrected Akaike Information Criterion model comparison suggested that the second, 10-predictor model (including chlorophyll, DO, pH and elevation) was preferred over the reduced, 6-predictor model for prediction of *P. parvum* blooms (Supplementary Table S5). Addition of cyanobacteria

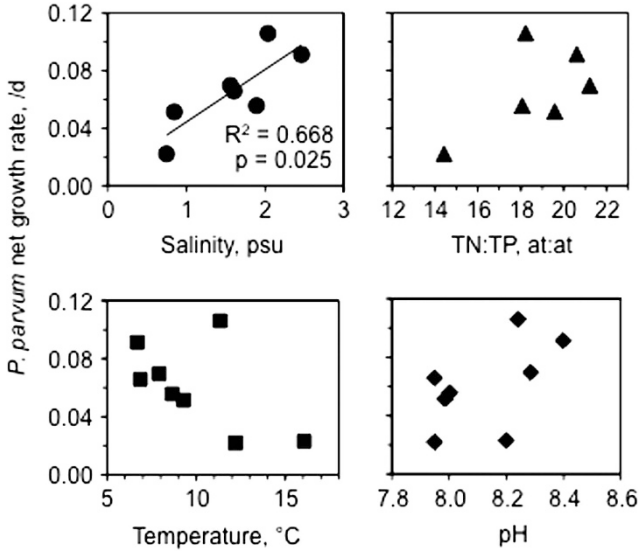

Figure 3 Net growth rates of *P. parvum* during initial bloom development in Lebanon Pool (L2) during the winters of 2004–2005 through 2011–2012 in relation to mean salinity, TN:TP, temperature and pH during the period of population increase. Note that the relationship between growth rate and pH is confounded by an expected increase in CO_2 consumption (and concomitant increase in pH) with increased growth rates.

Table 1 Results from logistic regression analysis showing the independent variables, partial logistic regression coefficients (*B*), standard errors of the partial slope coefficients (s.e.), Wald test and the significance level (*P*)

Independent variables	B	*S.e.*	*Wald* Z	P-*value* (>|Z|)
Salinity (psu)	2.6553	0.2431	10.92	**<0.0001**
Site (littoral or pelagic)	−0.6409	0.1679	−3.82	**0.0001**
Temperature (°C)	−0.1075	0.0108	−9.98	**<0.0001**
TN (mg l^{-1})	−0.2730	0.2629	−1.04	0.2991
TP (mg l^{-1})	−3.2344	1.3448	−2.41	**0.0162**
TN:TP (at:at)	−0.0264	0.0071	−3.74	**0.0002**
Intercept	−0.1884	0.3672	−0.51	0.6079

Abbreviations: TN, total nitrogen; TP, total phosphorus.
Model goodness of fit is shown by the model likelihood ratio χ^2, Nagelkerke R^2_N, the *C* index, Somer's D_{xy} and the bootstrapping slope.
The dependent variable was coded so that 0 = *P. parvum* absence and 1 = *P. parvum* presence.
Model $\chi^2 = 356.30$, $P < 0.0001$.
$R^2_N = 0.363$; C = 0.816, $D_{xy} = 0.632$; Slope = 0.976.
$N = 1181$.
All *P*-values ≤ 0.05 are indicated in bold.

Table 2 Results from logistic regression analysis showing the independent variables, partial logistic regression coefficients (*B*), standard errors of the partial slope coefficients (s.e.), Wald test and the significance level (*P*)

Independent variables	B	*S.e.*	*Wald* Z	P-*value* (>|Z|)
Salinity (psu)	2.3751	0.2527	9.4	**<0.0001**
Site (littoral or pelagic)	−2.3976	0.7353	−3.26	**0.0011**
Temperature (°C)	−0.1049	0.0234	−4.48	**<0.0001**
TN (mg l^{-1})	2.1341	0.5686	3.75	**0.0002**
TP (mg/l^{-1})	−15.5889	4.4775	−3.48	**0.0005**
TN:TP (at:at)	−0.1213	0.0322	−3.77	**0.0002**
Intercept	−1.0392	0.9167	−1.13	0.2569

Abbreviations: TN, total nitrogen; TP, total phosphorus.
Model goodness of fit is shown by the model likelihood ratio χ^2, Nagelkerke R^2_N, the *C* index, Somer's D_{xy} and the bootstrapping slope.
The dependent variable was coded so that 0 = no bloom (*P. parvum* <10 000 cells per ml) and 1 = bloom (*P. parvum* ⩾10 000 cells per ml).
Model $\chi^2 = 258.81$, $P < 0.0001$.
$R^2_N = 0.481$; $C = 0.921$, $D_{xy} = 0.843$; slope = 0.947.
$N = 1181$.
All *P*-values ≤ 0.05 are indicated in bold.

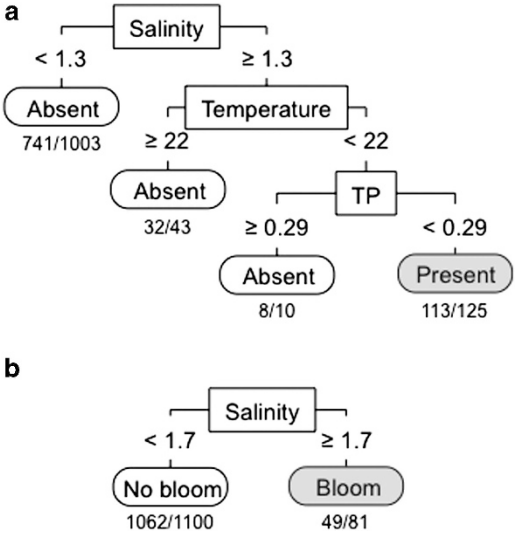

Figure 4 Pruned classification trees for *P. parvum* presence or absence (**a**) and bloom or no bloom (**b**) using the predictors salinity (psu), site type (littoral, pelagic), temperature (°C), TN (mg l^{-1}), TP (mg l^{-1}) and molar TN:TP. Presence was defined as detectable by microscope (⩾166 cells per ml; 2006–2007) and quantitative PCR (⩾26 cells per ml; 2008–2012) analyses; misclassification rate = 0.243 (presence) and 0.059 (bloom); $N = 1181$.

to the analysis reduced the size of the data set ($n = 560$) and slightly weakened both 10- and 6-predictor model results (Supplementary Table S6).

The classification tree analysis of *P. parvum* presence generated a pruned tree with three splits and four terminal nodes with a misclassification rate of 0.243 (Figure 4a). Only one terminal node was classified by environmental conditions (salinity ⩾1.3 psu, temperature <21.7 °C and TP <0.29 mg l^{-1}) in which *P. parvum* was likely to be present. The classification tree analyses of *P. parvum* blooms generated a pruned tree with only one split and two terminal nodes with a misclassification rate of 0.059 (Figure 4b). In this case, *P. parvum* blooms were predicted to be most common when salinity ⩾1.7 psu.

Discussion

Previously, based on a 3-year analysis of Lake Texoma littoral sites during winter (January–April) only, we concluded that salinity and TN:TP were the primary predictors of *P. parvum* densities

(Hambright *et al.*, 2010). Although all parameters examined in that study were correlated with winter *P. parvum* densities, a general lack of any significant relationships between *P. parvum* and environmental factors across the complete data set led to the conclusion that there was a great deal of stochasticity behind *P. parvum* dynamics in Lake Texoma. Now with an additional 6 years of data, much of the apparent stochasticity has become predictable. As before, but now with the complete data set, including both pelagic and littoral sites, *P. parvum* presence and blooms are related positively to salinity and negatively to TN:TP (Tables 1 and 2 and Supplementary Tables S1 and S2). Both presence and blooms are also negatively related to site type and temperature, indicating that *P. parvum* is more likely to be present or to bloom in littoral sites during the winter. While winter *P. parvum* abundances had scaled positively with both TN and TP (Hambright *et al.*, 2010), in this study, *P. parvum* presence scaled negatively with both TN and TP, and *P. parvum* blooms scaled positively with TN, but negatively with TP. These negative relationships with nutrients seem contradictory to the nutrient-bloom paradigm embodied by the general concept of eutrophication. However, further analysis of the data reveals that the negative and positive relationships are due to unimodal relationships between the probabilities of *P. parvum* presence and *P. parvum* blooms with TN and TP (Supplementary Figure S2). Probabilities for presence and blooms would be predicted to occur at relatively high TN (1.2–1.7 $mg\,l^{-1}$) and TP (100–250 $\mu g\,l^{-1}$) concentrations, with both probabilities declining markedly at lower and higher values.

Additions of chlorophyll, DO, pH and water elevation resulted in improved model fits for both *P. parvum* presence and *P. parvum* blooms. Hambright *et al.* (2010) also noted that the inclusion of chlorophyll improved their multiple regression model of winter *P. parvum* densities. However, the potential correlation between *P. parvum* and its primary photosynthetic pigment is uninformative with respect to understanding environmental regulation of *P. parvum* dynamics. The predictive capability of pH, which has been reported to affect toxicity of *P. parvum* toxins (Valenti *et al.*, 2010, but see, Cichewicz and Hambright, 2010), and of DO are similarly compromised, as elevated pH and O_2 concentrations are indicative of high algal growth rates and blooms. However, pH was not a significant predictor of *P. parvum* blooms generally (Supplementary Table S2). Previously, we (Hambright *et al.*, 2010) hypothesized that water elevation could be used to predict *P. parvum* blooms because of a strong negative relationship between salinity in the littoral regions of the lake and water levels, which reflected the balance between inflows and evaporation. Roelke *et al.* (2011) report a similar role for inflows in three Brazos River system reservoirs (Lakes Possum Kingdom, Granbury and Whitney), with blooms occurring only at low inflows and high salinities. Our present study also reveals relationships between water level elevation and both *P. parvum* presence and blooms (Supplementary Tables S2 and S3), as well as with salinity (Supplementary Figure S3), but unlike the Brazos system in which inflows and salinities are more tightly coupled, these relationships in Lake Texoma are nonlinear and very noisy, making useful interpretation difficult. For example, the counter-intuitive positive relationships between water level and *P. parvum* presence or blooms detected by logistic regression analyses are driven by unimodal relationships (Supplementary Figure S2), the left-hand side of which resulted from a period of relative drought and lower than normal water levels during seasons when other conditions conducive to *P. parvum* (e.g., cooler temperatures, higher salinities) were absent. Finally, there has been considerable discussion pertaining to potential negative (via allelopathy; see references in Granéli *et al.*, 2012) and positive (via indirect proliferation of bacterial prey for *P. parvum*; see references in Granéli *et al.*, 2012) impacts of cyanobacteria on *P. parvum*. Although we only monitored cyanobacterial abundances during 2009–2012, addition of phycocyanin concentrations (a proxy for cyanobacterial abundances) to our logistic regression models did not provide any improvement in predictive capabilities, even though cyanobacteria are very common in Lake Texoma during much of the year.

As in previous inland studies, we found that *P. parvum* presence and blooms in Lake Texoma seemed to correlate with non-optimal growth conditions, at least with respect to temperature and salinity (Hambright *et al.*, 2014; Patiño *et al.*, 2014). During 2004–2012, net population growth rates in Lebanon Pool were highest at suboptimal cold temperatures, between 5 °C and 10 °C, as well as suboptimal salinities, ranging between 0.5 and 2.5 psu, during the bloom development phase, compared with laboratory optima of 27 °C and 22 psu measured for a *P. parvum* strain (UTEX-LB2792; originally cited using the temporary identification LBZZ181, D Nobles, UTEX Culture Collection, University of Texas, personal communication) isolated from the Colorado River, Texas (Baker *et al.*, 2007). Other experiments using UTEX-LB2797 revealed a logarithmic growth–salinity relationship in which maximum potential growth rates of *P. parvum* declined at salinities of 6 psu or lower compared with 15 and 30 psu (Hambright *et al.*, 2014). Similar findings have been documented for Norwegian and Danish strains of *P. parvum* (Larsen and Bryant, 1998). Not only were *P. parvum* growth rates during blooms in Lebanon Pool, Lake Texoma (Oklahoma–Texas) negatively related to temperatures between 7 °C and 16 °C, these blooms were large (up to 200 000 cells per ml) and quite toxic, as indicated by bioassays and large fish kills (Hambright *et al.*, 2014; Zamor *et al.*, 2014). In Hambright *et al.* (2014), both growth rates and

general toxicity in cultures were found to increase with increasing salinity.

We hypothesized that environmental conditions that foster high growth in *P. parvum* relative to other microbial constituents are the principle determinants of *P. parvum* blooms in inland freshwater systems. The alternative hypothesis suggests dispersal limitation. Two aspects of our data are supportive of the environment side of the hypothesis: (1) downstream flow in Lake Texoma is repeatedly introducing *P. parvum* downlake, yet blooms remain restricted to uplake, Red River stretches of the reservoir (Hambright *et al.*, 2014; Zamor *et al.*, 2014), (2) both our logistic regression models (compare R^2_N, C and Somer's D_{xy}) and classification tree analyses (compare misclassification rates) indicate that environmental data were better at predicting *P. parvum* blooms than *P. parvum* presence. We interpret this as meaning that presence is more dependent on factors such as detection limit or propagule pressure or both, whereas blooms are explicitly associated with appropriate environmental conditions. Because it is a microbial eukaryote that can occur in immense numbers, encyst and be passively transported, *P. parvum*'s dispersal capabilities are potentially unlimited (sensu Finlay, 2002; but see Martiny *et al.*, 2006; Hanson *et al.*, 2012) and subjected to probabilities related to various potential vectors, such as wind, migratory-animal-assisted propagation or even human intervention (Johnson *et al.*, 2008). For example, Hambright *et al.* (2014) estimated that during a *P. parvum* bloom of 10^5 cells per ml in Lebanon Pool (volume $\sim 3 \times 10^6\,m^3$), the total *P. parvum* population would exceed 3×10^{11} individuals and represent an immense pool for downstream transport of invasive propagules via hydraulic flushing and advective downstream flow. Yet, *P. parvum*, which is often detected downstream, has yet to bloom in any area of the lake outside the environmental conditions identified here (Zamor *et al.*, 2012). A similar argument can be made for areas downstream of Lake Texoma, as well as many downstream water bodies in the Canadian River watershed, which also experiences *P. parvum* blooms in upstream systems (Zamor, 2013).

Recently, there has been growing interest focused on biogeographic patterns in the distributions of microorganisms and whether general ecological principles based predominantly on the study of macrobial species also hold for microbes. For example, can microbes be invasive? Much of the discussion has centered on the longstanding credo in microbiology, 'Everything is everywhere, but, the environment selects' (Baas-Becking, 1934; de Wit and Bouvier, 2006), which suggests that dispersal is not limiting for microbes and that the environment is the primary factor determining whether a particular microbe is abundant enough to be detected in a given habitat. Indeed, early morphology-based taxonomic studies of protists tended to conclude that cosmopolitan distributions for microbes were the norm (Finlay, 2002; Fenchel and Finlay, 2004). By contrast, recent taxonomic studies using modern molecular genetics (i.e., pyrosequencing) have provided evidence that indeed some microbes seem to exhibit biogeographic patterns unrelated to environmental conditions (Hanson *et al.*, 2012), although the underlying mechanisms which explain these patterns are not yet understood. Interestingly, a more recent study by Gibbons *et al.* (2013) has provided support for the Baas-Becking hypothesis, as they propose that extremely deep sequencing (i.e., on the order of 10^{11} sequences) of an individual marine bacterial community would reveal global phylogenetic diversity of the oceans. Within the *P. parvum* system, a recent phylogenetic analysis (Lutz-Carrillo *et al.*, 2010) also provides evidence that long-distance dispersal in *P. parvum* is common. There, they found that isolates of *P. parvum* from populations in the US states of Texas, South Carolina and Wyoming share high levels of similarity in the first internal transcribed spacer in the nuclear ribosomal operon (ITS1) with *P. parvum* isolates from Scotland; isolates from a single lake in north Texas were similar to isolates from Denmark and Norway; and isolates from Maine were similar to isolates from England. Isolates collected from the US state of Washington were most similar to a clade comprised of isolates from Scotland, Denmark, Norway, Texas, South Carolina and Wyoming, whereas an isolate from Australia was most similar to the England and Maine clade. Thus, if everything is everywhere (or has the potential to be), perceived microbial invasions and range expansions may indeed be reflective only of technical limitations in detection or changing environmental conditions. Alternatively, spatial barriers to dispersal could vary in strength, from being negligible at local and regional scales, to stronger at continental, intercontinental and global scales (Lutz-Carrillo *et al.*, 2010). These contrasting hypotheses should be testable with population genetic surveys.

Classical ecologic theory stipulates that the differences in taxonomic diversity and composition among sites will depend on whether species are excluded from meaningful participation in a particular habitat by local biotic and abiotic conditions or by dispersal limitation, in which a species does not arrive at that particular habitat (Shurin, 2000). Unfortunately, in microbial systems, the probability of type 2 error associated with rejecting dispersal as a driving mechanism is heavily dependent on technological detection limits of the commonly used molecular tools. This problem has led some researchers to suggest that dispersal be redefined as being both arrival to, and successful establishment in, a given habitat, as evidenced by metabolic activity and some level of reproduction (Hanson *et al.*, 2012). Thus, mere presence would not suffice as evidence that dispersal had occurred and presumably the problem associated with detection limits would no longer be an issue. Unfortunately, this definition of dispersal

would exclude many microbes that are present in a community but that are not active, biologically relevant participants in the community because they are either extremely rare (Sogin *et al.*, 2006; Caron and Countway, 2009) or dormant (Jones and Lennon, 2010; Gibbons *et al.*, 2013) under the current environmental conditions. However, these microbes could become important community participants if the environment changed to more favorable conditions (Caron and Countway, 2009). Furthermore, the use of active community participation to define dispersal renders 'Everything is everywhere, but, the environment selects' untestable. By contrast, for microbial taxa for which technical limitations of detection have been greatly reduced through development of new methodologies, we would argue that differentiation between the roles of dispersal and habitat suitability in microbial biogeographic patterns can and should be addressed experimentally (e.g., Ehrlén and Eriksson, 2000; Shurin, 2000).

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank Paul Mauck, Jeff Boxrucker and Greg Summers (Oklahoma Department of Wildlife Conservation, ODWC) who were instrumental in initiating this study, and Greg Summers, Matt Mauck and Gene Gilliland for support. We also thank the staff of UOBS for technical and logistical assistance; Karen Glenn, Emily Remmel and Ann Morris for technical assistance; Dave Caron, Adriane Jones and Vikki Campbell for invaluable advice and assistance during every aspect of this project; and Francisco Acosta and Jim Grover for comments and suggestions on earlier drafts of this paper. Financial assistance was provided by the ODWC, through the Sport Fish Restoration Program (Grant F-61-R), the National Science Foundation (DEB-1011454) and the University of Oklahoma Office of the Vice President for Research.

References

American Public Health Association. (1998). *Standard Methods for the Examination of Water and Waste Water*, 20th edn. American Public Health Association: Washington, DC, USA.

Army Corps of Engineers–Tulsa District (2012). DSNT2: Lake Texoma, Denison Dam 2009. Available at http://www.swt-wc.usace.army.mil/DENI.lakepage.html. Last accessed 14 July 2012.

Baas-Becking LGM. (1934). *Geobiologie of inleiding tot de melieukunde*. WP Van Stockum & Zoon: The Hague, The Netherlands.

Baker JW, Grover JP, Brooks BW, Urena-Boeck F, Roelke DL, Errera R *et al.* (2007). Growth and toxicity of *Prymnesium parvum* (Haptophyta) as a function of salinity, light, and temperature. *J Phycol* **43**: 219–227.

Baker JW, Grover JP, Ramachandrannair R, Black C, Valenti TW, Brooks BW *et al.* (2009). Growth at the edge of the niche: an experimental study of the harmful alga *Prymnesium parvum*. *Limnol Oceanogr* **54**: 1679–1687.

Burnham KP, Anderson DR. (2002). *Model Selection and Multimodel Inference: A Practical Information–Theoretic Approach*. Springer: Berlin, Germany.

Caron DA, Countway PD. (2009). The protistan Rare Biosphere in aquatic microbial ecology: hypotheses on the role of diversity in a changing world. *Aquat Microb Ecol* **57**: 227–238.

Cichewicz RH, Hambright KD. (2010). A revised amino group p*K*a for prymnesins does not provide decisive evidence for a pH-dependent mechanism of *Prymnesium parvum*'s toxicity. *Toxicon* **55**: 1035–1037.

Davis CE, Hyde JE, Bangdiwala SI, Nelson JJ. (1986). An example of dependencies among variables in a conditional logistic regression. In: Moolgavkar SH, Prentice RL (eds) *Modern Statistical Methods in Chronic Disease Epidemiology*. Wiley: New York, NY, USA.

de Wit R, Bouvier T. (2006). '*Everything is everywhere*, but, *the environment selects*'; what did Baas Becking and Beijerinck really say? *Environ Microbiol* **8**: 755–758.

De'ath G, Fabricius KE. (2000). Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* **81**: 3178–3192.

Edvardsen B, Paasche E. (1998). Bloom dynamics and physiology of *Prymnesium* and *Chrysochromulina*. In: Anderson DM, Cemballa AD, Hallegraeff GM (eds) *NATO ASI Series, Vol G 41, Physiological Ecology of Harmful Algal Blooms*. Springer: Heildelberg, Germay, pp 193–209.

Ehrlén J, Eriksson O. (2000). Dispersal limitation and patch occupancy in forest herbs. *Ecology* **81**: 1667–1674.

Fenchel T, Finlay BJ. (2004). The ubiquity of small species: patterns of local and global diversity. *BioScience* **54**: 777–784.

Finlay BJ. (2002). Global dispersal of free-living microbial eukaryote species. *Science* **296**: 1061–1063.

Franks JL, Clyde GA, Dickson KL. (2001). Zooplankton community structure and seasonal dynamics in Lake Texoma (Oklahoma–Texas). *Tex J Sci* **53**: 203–220.

Gibbons SM, Caporaso JG, Pirrung M, Field D, Knight R, Gilbert JA. (2013). Evidence for a persistent microbial seed bank throughout the global ocean. *Proc Natl Acad Sci USA* **110**: 4651–4655.

Granéli E, Edvardsen B, Roelke DL, Hagström JA. (2012). The ecophysiology and bloom dynamics of *Prymnesium* spp. *Harmful Algae* **14**: 260–270.

Ground TA, Groeger AW. (1994). Chemical classification and trophic characteistics of Texas reservoirs. *Lake Reserv Manage* **10**: 189–201.

Grover JP, Roelke DL, Brooks BW. (2012). Modeling of plankton community dynamics characterized by algal toxicity and allelopathy: a focus on historical *Prymnesium parvum* blooms in a Texas reservoir. *Ecol Modell* **227**: 147–161.

Hambright KD. (2012). Golden algae and the health of Oklahoma's lakes. *LakeLine Fall* **2012**: 33–38.

Hambright KD, Easton JD, Zamor RM, Beyer J, Easton AC, Allison B. (2014). Regulation of growth and toxicity of a mixotrophic microbe: implications for understanding range expansion in *Prymnesium parvum*. *Freshw Sci;* in press.

Hambright KD, Zamor RM, Easton JD, Glenn KL, Remmel EJ, Easton AC. (2010). Temporal and spatial variability of an

invasive toxigenic protist in a North American subtropical reservoir. *Harmful Algae* **9**: 568–577.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.

Harrell FEJ. (2013). rms: Regression Modeling Strategies. R package version 4.0-0. Available at: http://CRAN.R-project.org/package=rms 2013. Last accessed 25 September 2013.

James T, De La Cruz A. (1989). *Prymnesium parvum* Carter (Chrysophyceae) as a suspect of mass mortalities of fish and shellfish communities in western Texas. *Tex J Sci* **41**: 429–430.

Johnson PTJ, Olden JD, Vander Zanden MJ. (2008). Dam invaders: impoundments facilitate biological invasions into fresh waters. *Front Ecol Environ* **6**: 357–363.

Jones SE, Lennon JT. (2010). Dormancy contributes to the maintenance of microbial diversity. *Proc Natl Acad Sci USA* **107**: 5881–5886.

Jonsson PR, Pacvia H, Toth G. (2009). Formation of harmful algal blooms cannot be explained by allelopathic interactions. *Proc Natl Acad Sci USA* **106**: 11177–11182.

Larsen A, Bryant S. (1998). Growth rate and toxicity of *Prymnesium parvum* and *Prymnesium patelliferum* (Haptophyta) in response to changes in salinity, light and temperature. *Sarsia* **83**: 409–418.

Lutz-Carrillo DJ, Southard GM, Fries LT. (2010). Global genetic relationships among isolates of golden alga (*Prymnesium parvum*). *J Am Water Resour Assoc* **46**: 24–32.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL *et al.* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102–112.

Matthews WJ, Gido KB, Gelwick FP. (2004). Fish assemblages of reservoirs, illustrated by Lake Texoma (Oklahoma–Texas, USA) as a representative system. *Lake Reserv Manage* **20**: 219–239.

Oklahoma Water Resources Board (2010). *2008–2009 Oklahoma Lakes Report, Beneficial Use Monitoring Program*. Oklahoma Water Resources Board: Oklahoma City, OK, USA.

Patiño R, Dawson D, VanLandeghem MM. (2014). Retrospective analysis of associations between water quality and toxic blooms of golden alga (*Prymnesium parvum*) in Texas reservoirs: Implications for understanding dispersal mechanisms and impacts of climate change. *Harmful Algae* **33**: 1–11.

R Development Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria. Available at: http://www.R-project.org. Last accessed 25 September 2013.

Remmel EJ, Hambright KD. (2012). Toxin-assisted micropredation: experimental evidence shows that contact micropredation rather than exotoxicity is the role of *Prymnesium* toxins. *Ecol Lett* **15**: 126–132.

Roelke DL, Grover JP, Brooks BW, Glass J, Buzan D, Southard GM *et al.* (2011). A decade of fish-killing *Prymnesium parvum* blooms in Texas: roles of inflow and salinity. *J Plankton Res* **33**: 243–253.

Shurin JB. (2000). Dispersal limitation, invasion resistance, and the the structure of pond zooplankton communities. *Ecology* **81**: 3074–3086.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR *et al.* (2006). Microbial diversity in the deep sea and the underexplored ''rare biosphere''. *Proc Natl Acad Sci USA* **103**: 2115–12120.

Therneau T, B. A, Ripley B. (2013). rpart: Recursive Partitioning. R package version 4.1-3. Available at: http://CRAN.R-project.org/package=rpart 2013. Last accessed 25 September 2013.

Valenti Jr TW, James SV, Lahousse MJ, Schug KA, Roelke DL, Grover JP *et al.* (2010). A mechanistic explanation for pH-dependent ambient aquatic toxicity of *Prymnesium parvum* carter. *Harmful Algae* **55**: 990–998.

Zamor RM. (2013). The ecology of an invasive toxigenic protist. PhD dissertationUniversity of Oklahoma: Norman, OK, USA.

Zamor RM, Glenn KL, Hambright KD. (2012). Incorporating molecular tools into routine HAB monitoring programs: using qPCR to track invasive *Prymnesium*. *Harmful Algae* **15**: 1–7.

Zamor RM, Franssen NR, Porter C, Patton TM, Hambright KD. (2014). Rapid recovery of a fish assemblage following an ecosystem disruptive algal bloom. *Freshw Sci* **33**: 390–401.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)