ECOGRAPHY

Research

Stream diatoms exhibit weak niche conservation along global environmental and climatic gradients

Janne Soininen, Aurélien Jamoneau, Juliette Rosebery, Thibault Leboucher, Jianjun Wang, Mikolaj Kokociński and Sophia I. Passy

J. Soininen (http://orcid.org/0000-0002-8583-3137) (janne.soininen@helsinki.fi) and J. Wang (http://orcid.org/0000-0001-7039-7136), Dept of Geosciences and Geography, Univ. of Helsinki, Helsinki, Finland. JW also at: State Key Laboratory of Lake Science and Environment, Nanjing Inst. of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China, and Univ. of Chinese Academy of Sciences, Beijing, China. – A. Jamoneau, J. Rosebery and T. Leboucher, Aquatic Ecosystems and Global Changes Research Unit, IRSTEA, Cestas, France. – M. Kokociński, Dept of Hydrobiology, Adam Mickiewicz Univ., Poznań, Poland. – S. I. Passy, Dept of Biology, Univ. of Texas at Arlington, Arlington, TX, USA.

Ecography
41: 1–8, 2018
doi: 10.1111/ecog.03828

Subject Editor: Jason Pither
Editor-in-Chief: Miguel Araújo
Accepted 31 May 2018

Niche conservatism (NC) describes the scenario in which species retain similar characteristics or traits over time and space, and thus has potentially important implications for understanding their biogeographic distributions. Evidence consistent with NC includes similar niche properties across geographically distant regions. We investigated whether NC was evident in stream diatom morphospecies by modeling species responses to environmental and climatic variables in a set of calibration sites (from the US) and then evaluated the models with test sets (from France, Finland, New Zealand, Antilles and La Réunion). We also examined whether diatom species showed congruency in environmental niche optima and niche breadths between the study regions, and whether species occupancy and functional traits influenced the observed patterns. We used boosted regression tree models with local environmental variables and climatic variables as predictors. We detected low NC in both environmental and climate models and a lack of consistent differences in NC between widely distributed and regionally rare species and among functional groups. For all species, diatom environmental and climatic optima varied clearly between the regions but showed some positive relationships especially for pH and total phosphorus. Diatom niche breadths were only weakly correlated between the US and the other regions. We demonstrated that diatoms showed overall relatively little NC globally, and NC was especially low for climatic variables. Collectively, these findings suggest that there may exist locally adapted lineages within the diatom morphospecies or diatoms possess some adaptation potential for differences in temperature. We argue that in diatoms, environmental and especially climate models may not be transferrable in space globally but need regional diatom data for calibration because species niches seem to differ among geographical regions.

Keywords: streams, species distribution models, diatoms

NORDIC SOCIETY OIKOS

www.ecography.org

Introduction

Examination of past and present species distributions has attracted ecologists for a long time. While biotic interactions and contemporary environmental factors are among the key sources of variation in present distributions, evolution, long-term climatic changes and species dispersal moderate species' ranges over long time periods (Vellend 2010). The retention of niche characteristics over time and space, i.e. niche conservation (NC), is also a fundamental cause for large-scale species distributions and diversity gradients on Earth (reviewed by Wiens and Graham 2005, Wiens et al. 2010). Niche conservation relates to the responses of species to climate changes (Parmesan and Yohe 2003), affects ecosystem functioning (Cadotte et al. 2008) and influences the success of invasive species in novel environments (Sax 2001). Therefore, NC underlies large-scale diversity patterns, which in turn have functional implications for ecosystems.

Niche conservatism is moderated by natural selection, gene flow, pleiotropy and the lack of genetic variability in traits (Wiens and Graham 2005). For example, efficient gene flow across sites may prevent local adaptation and associated speciation under strong dispersal rates (Anderson 1994, Finlay 2002) leading thus to NC. In addition, species may not be able to expand their ranges and niches simply because they lack genetic variation in traits (Case and Taper 2000). Recently, NC has gained much attention since it constrains the changes in species ranges due to climate change. Given that if species niche is conserved, climate change may result in local extinctions because species cannot maintain viable populations in the altered conditions. Conversely, if species can adapt to climate changes, they may retain their present ranges even in a changing climate.

Global patterns in NC may give insights into the underlying mechanisms generating NC. If NC is globally pervasive, it is possible that there is a strong gene flow among populations even at global scales. Therefore, an intriguing approach would be to examine NC across continents using model taxa that have comparatively broad distributions and efficient dispersal. Freshwater diatoms are appropriate target taxa for such analyses as they typically have relatively clear environmental niches (Stoermer and Smol 1999), high dispersal rates and potentially global species pools (Finlay 2002, but see Vyverman et al. 2007). Bennett et al. (2010) studied NC in lacustrine diatoms along pH and found that pH niche appeared to be conserved across continents. The generality of diatom NC along multiple environmental gradients including climate is, however, not well resolved.

There are multiple methods of investigating the patterns in NC. Among these, phylogenetic analyses and paleontological findings would need either high resolution data about species phylogenetics or ample fossil record, which are both missing for many taxonomic groups at present. Another useful tool to examine large-scale patterns in NC are species distribution models (SDMs). Generally, predicting species distributions is one of the great challenges in ecology (Guisan and Zimmermann 2000, Guisan and Thuiller 2005). Species distribution models relate occurrence data to climatic and/or local environmental predictors and have been employed to predict the distribution of a wide range of taxa such as birds (Brotons et al. 2004), butterflies (Luoto et al. 2005), trees (Thuiller et al. 2003), fish (Buisson et al. 2008), freshwater plankton (Soininen et al. 2013), and even diatoms at a regional scale in Finland (Pajunen et al. 2016). In NC research, SDMs can be used to quantify model transferability across study regions, and test whether the model calibrated in one region could be successfully applied in another geographical region, which would indicate NC.

Our aim here is to investigate NC in freshwater diatom morphospecies (hereafter 'diatoms') globally using two approaches: 1) SDMs along local environmental and climatic gradients and 2) the comparison of species` niche properties among study regions. We first model diatom responses with SDMs using a set of calibration sites (here, US sites) and then evaluate the model performances with independent five test sets from France, Finland, New Zealand, Antilles and La Réunion.

We further consider consistency in niche properties (that is, species environmental optima and niche breadths) among regions to be indicative of NC (Bennett et al. 2010). We first hypothesize that species growth forms or guilds may affect diatom occupancy and dispersal capacity and thus have influence on the degree of NC. For example, Heino and Soininen (2006) showed that attached stream diatom species had higher occupancy than non-attached diatom species in boreal streams in Finland. Passy (2016) also documented that the low profile diatom species, which grow close to the substratum, had significantly higher occupancy than the high profile (i.e. species that occupy the overstorey of the biofilm) and motile species (i.e. species that can move across the biofilm) in the US streams. Assuming that NC is promoted by high dispersal capacity (Anderson 1994, Finlay 2002), and that low profile and attached species have greater occupancy and dispersal capacity than others (Heino and Soininen 2006, Passy 2016), we predict (H_1) that NC will be most evident among low profile species, which are often attached, compared to other species.

In addition to growth forms, we hypothesize that there may be differences in the degree of NC between species' occupancy groups. We predict that core (i.e. regionally widely distributed species) species may have higher NC than satellite (i.e. regionally rare species) species (H_2). Core species have high occurrence and cell densities, broader niches and often good dispersal capacity (Passy 2012), being able to establish populations in a wider range of environments. In contrast, satellite species with typically more specific environmental preferences may need to adapt locally to new environmental conditions in order to establish populations when dispersing to new localities. Thus, the overall objectives of this study are to examine whether global NC exists among diatom species and whether there are differences in NC among ecological guilds and species occupancy groups.

Material and methods

Datasets

We included diatom datasets from United States (1201 sites), France (2640), Finland (196), New Zealand (104), Antilles (125) and La Réunion (53) with a total of 4319 study sites, sampled for diatoms and water chemistry. Datasets are described in detail in Soininen et al. (2016). After cleaning of diatom samples with acid or hydrogen peroxide, about 400–600 diatom frustules per sample were identified using a light microscope. Main diatom identification keys were Krammer and Lange-Bertalot (1986–1991) and Lange-Bertalot (1995–2015, 2000–2013). Taxonomic consistency was ensured by expert taxonomists from the Patrick Center for Environmental Research in Philadelphia, from Irstea in France and from the Univ. of Helsinki in Finland. Datasets were also taxonomically homogenized according to Omnidia 5.3 (Lecointe et al. 1993), in order to obtain a final list of taxa validated by all the experts involved.

We considered four environmental variables including pH, conductivity (µS cm^{-1}), total phosphorus (mg l^{-1}) and altitude (m) available for all regions except for New Zealand (without total phosphorus). Altitude was included as environmental variable here as it correlates well with current velocity and substrate size and thus indicates stream physical conditions for diatoms (Wang et al. 2011). Altitude was extrapolated from the Global Multi-resolution Terrain Elevation Data 2010 at 7.5 arc-second resolution (Danielson and Gesch 2011). We also considered four climatic variables including annual precipitation (mm), seasonality in precipitation (%), annual temperature (°C) and annual temperature range (°C) drawn from WorldClim database (Hijmans et al. 2005). Climate data covered years 1950–2000 and average values were included in our analyses. All pairwise correlations among environmental and climatic variables had $r_p < 0.65$, see Soininen et al. (2016) for details.

Diatom guilds

We assigned all diatom species to guilds (low profile, high profile and motile) according to Passy (2007) and Rimet and Bouchez (2012). We also distinguished acid-tolerant species, which can withstand acid stress, as a separate guild. Acid-tolerant forms (that is, acidophilous and acidobiontic species in Van Dam et al. 1994) thrive in $pH < 7$.

Data analyses

We used US diatom data as a training set because they exhibited the greatest gradients in local environmental and climatic conditions. Since climatic conditions in the US and the focal region differed to some extent, we selected the suite of climatic variables that maximized the overlap between the US and the focal region. Thus, we retained temperature range only for Finland models, and did not use annual precipitation and annual temperature for New Zealand and Antilles models, respectively. According to principal component analysis, the climatic and environmental gradient across the US covered relatively well the variation in the other regions except for climatic conditions in Antilles (Supplementary material Appendix 1 Fig. A1). As the results of predictive models between the US and Antilles did not differ notably from the modeling results between the US and other countries, we also included Antilles data in climatic models. We also ran the analyses with US sites having environmental and climatic conditions strictly overlapping with the corresponding conditions in the test sets. However, as these results were very similar to the results shown here, they are not discussed further.

In our SDM procedure, the five focal data sets were used as test sets for US one at a time. We included species that occurred both in the training set and in the respective test set and excluded the species that were present at fewer than 25 sites in the US data. This was done because the model performance of rare species may be misleadingly high (Brotons et al. 2004). Thus, the suite of species included in the analyses differed among the five model comparisons (e.g. in US vs Finland, US vs France and so on). The number of species included in the model comparisons ranged between 202 (US vs France) and 46 (US vs La Réunion).

We modelled the species occurrences with three types of models: 1) with environmental variables, 2) with climatic variables and 3) full models comprising all variables. We used boosted regression trees (BRT) as a SDM method in the analyses. BRT is a machine learning technique, proven to be a robust method for creating SDMs for microorganisms (Pajunen et al. 2016) due to its high efficiency in fitting nonparametric data, and ability to manage various types of predictor variables. It does not require prior data transformation and takes into account the interaction effects between predictors (Elith et al. 2008). BRTs were performed with a Bernoulli distribution using maximum number of 3000 trees, an interaction depth of 4 and a learning rate of 0.001.

The observed and predicted occurrences of species were compared by calculating the area under the curve of a receiver operating characteristic plot (AUC) (Fielding and Bell 1997). AUC provides an evaluation of the agreement between the observed presence/absence records over a range of probability thresholds above which the model predicts presence. Models have intermediate predictive performance at AUC values of > 0.7 and excellent performance at AUC values of > 0.9 (Heikkinen et al. 2012). We compared the AUC values between models in the following manner: the model trained and tested in US (AUC_{us}) vs the model of the same species trained in US but predicted in another region (AUC_{pred}). If a species niche is well conserved, the values of AUC_{us} and AUC_{pred} should be comparable. Thus, we used the ratio between AUC_{us} and AUC_{pred} as a measure of niche conservation. We first considered a niche to be conserved if this ratio is between 0.9 and 1.1. We also used a wider range of ratios (e.g. between 0.8 and 1.2) to examine if changing the ratio would affect the model outcomes. Finally, we related this ratio to species occupancy and guild affiliation. Based on occupancy, species were classified as core (present at more than 5% of the sites) or satellite (present at 5% or fewer of all sites). These limits were deemed the most appropriate for our data with relatively limited number of sites for some of the regions. Based on species traits, species were placed into diatom guilds, as described above. We used χ^2 test to investigate if there are significant differences between the number of 'conserved' species (ratio between 0.9 and 1.1) and 'non-conserved' species among species occupancy groups and guilds.

We then compared species environmental optima and niche breadths between the US and the other study regions. Environmental optimum of each species was estimated with the mean value of environmental variable in sites where the species was present, weighted by its abundance. To estimate niche optima variation, we calculated for each species the difference between optima in the focal region vs the US. We used a Wilcoxon signed rank test to examine if mean species optima in the focal region were significantly different from species optima in the US. To facilitate the comparison with earlier papers (Bennett et al. 2010), we also calculated the relationships of the species' environmental optima between the US and other regions using regression analyses. For significant relationships, we examined 95% confidence interval around the slope and the intercept and used ANOVA to test whether regression slopes and intercepts differ significantly from 1 and 0, respectively, in order to investigate if species show consistency in their environmental optima between the US and other regions.

Species niche breadths were calculated using outlying mean index analysis (OMI, Dolédec et al. 2000) using species abundance data. OMI is a method for quantifying habitat niches with regard to niche position and niche breadth of species along multiple environmental gradients. The niche breadth is the amplitude in the distribution of species along the measured environmental gradients. A species that has a wide niche breadth is defined as a generalist, whereas a species confined to only a narrow range of environmental conditions is considered as a specialist. In OMI analysis, we considered species tolerance as a metric of niche breadth. We computed species tolerances with respect to local environmental variables, climate variables and all variables, and then compared species tolerance between the US and the focal region with linear regression. Given that it is possible that more widely distributed species have larger niche breadth, we also tested whether niche breadths were related to the number of occupied sites. Moreover, we tested whether niche breadth was related to the AUC value of the model as species with a narrow niche is more likely to have better predictability.

Data analyses were conducted with R software ver. 3.4.1 (R Development Core Team) using packages gbm (Ridgeway et al. 2017), caTools (Tuszynski 2014) and vioplot (Adler 2005).

Data deposition

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.v1v7856> (Soininen et al. 2018).

Results

Virtually all of the BRT models (local, climate and full models) had at least intermediate performance with AUC values > 0.7 (Fig. 1). Models with local variables only had generally better performances in the training set in the US compared with the test sets (Fig. 1). However, there were some conserved species (those that fall within a AUC_{us}:AUC_{pred} ratio between 0.9 and 1.1, indicated by a gray zone in Fig. 1) in all comparisons between regions. The highest proportion of conserved species out of all species was found between the US and La Réunion (39.1%) and the lowest proportion between the US and France (5.5%).

In climate BRT models (Fig. 1), there was in general an even lower number of conserved species than for local variable models or full models. The exceptions were the climatic models between the US and Antilles (24.1%) and those between the US and New Zealand (15.7%), which had slightly more conserved species than the local or full models. Even if we used a wider range of ratios (e.g. between 0.8 and 1.2), the climatic models showed, with very few exceptions (for Antilles), lower number of conserved species than the local models (Supplementary material Appendix 1 Fig. A2).

In a few cases, diatom guilds differed from each other in NC as acid tolerant species in France were significantly ($p < 0.05$, χ^2 test) more conserved according to local and full models than the other species (Supplementary material Appendix 1 Fig. A3). There was no consistent pattern of NC between the core and satellite species. However, core species were more poorly conserved ($p < 0.01$, χ^2 test) than the other species in France (local models) and New Zealand (all models) (Fig. 1, Supplementary material Appendix 1 Fig. A4, A5). Core species had significantly ($p < 0.001$, Kruskal–Wallis test) larger niche breadth than satellite species in New Zealand, La Réunion and Antilles, but in the other study regions niche breadths did not differ between core and satellite species.

The violin plots and associated Wilcoxon signed rank tests showed clearly that environmental and climatic optima typically differed between the US and the other countries (Fig. 2). However, for example, for pH and total phosphorus, La Réunion and Antilles did not differ significantly from the US, while France did not differ from the US in terms of conductivity. Compared to local variables, climatic optima varied more between the US and the other five regions (except for precipitation seasonality in Finland). Diatom species showed some significant ($p < 0.05$) positive relationships especially in their pH and TP optima between the US and other study regions (Supplementary material Appendix 1 Fig. A6). For pH and annual temperature, the regression slopes and intercepts did not differ significantly from the identity line

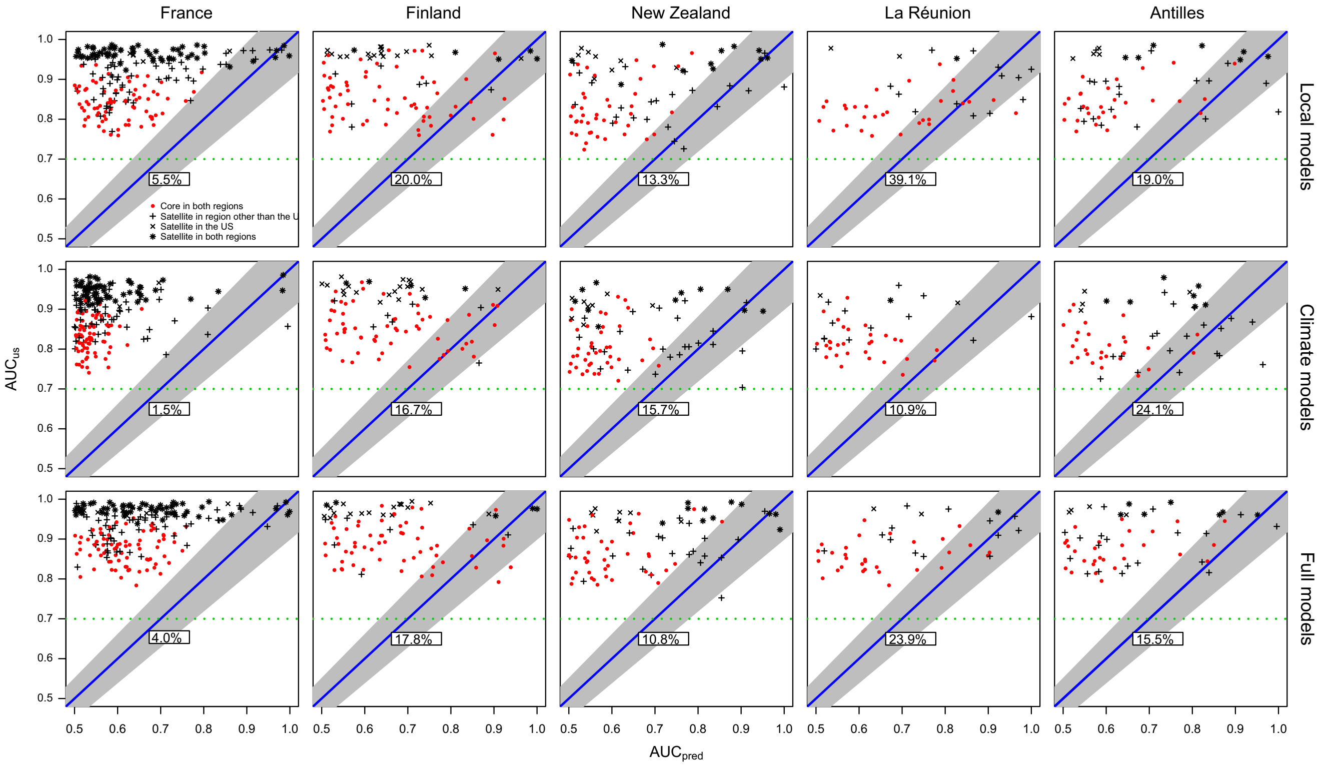

Figure 1. The correlation plots between area under the curve (AUC) values of species distribution model in the US (AUC_{us}) and predicted values in the other study regions (AUC_{pred}). The models are based on local environmental variables (upper panels), climatic variables (middle panels) and both groups of variables (full models, lower panels). The gray zones represent the AUC_{us}:AUC_{pred} ratio of > 0.9 but < 1.1 with the proportion (%) of the conserved species out of the total number of modelled species. The figure shows separately the AUC values of core (> 5% of sites) and satellite (≤ 5% of sites) species. The horizontal line at AUC 0.7 shows the limit for modelling species distribution at least at satisfactory level.

(slope = 1, intercept = 0) for Finland and France, respectively, but the residuals of the regressions were significantly different from a normal distribution (Shapiro–Wilk normality test, $p < 0.05$).

Diatom niche breadths were generally not correlated between the US and the other regions (results not shown). Niche breadths were larger in the US than in other regions in some comparisons, especially for satellite species in New Zealand and Antilles (Supplementary material Appendix 1 Fig. A7) and for different guilds, especially the motile guild (Supplementary material Appendix 1 Fig. A8). Niche breadths were typically positively related to the number of occupied sites in the test models (Supplementary material Appendix 1 Fig. A9) while in more than a third of the comparisons niche breadths were negatively related to AUC values (Supplementary material Appendix 1 Fig. A10).

Discussion

Using a global data set, we studied NC among stream diatom species with respect to local environmental and climatic variables. We first found that diatom species showed relatively little NC across study regions, evident from the SDM results and differences in species environmental optima and niche breadths across regions. Second, NC was in general even weaker for climatic variables than local environmental variables. Next, we will discuss these main findings in more detail.

The lack of strong NC in stream diatoms adds to a growing body of knowledge, suggesting that NC may be species- and context-dependent. For example, clear NC with respect to climatic variables was detected in arctic-alpine plants (Wasof et al. 2016) and high correlation ($r = 0.85$) between pH optima was observed in lake diatoms (Bennett et al. 2010). Conversely, Alahuhta et al. (2017) did not find evidence for NC in aquatic macrophytes. With regards to water pH, our results seem at least partly agree with Bennett et al. (2010) as we found that pH optima showed significant positive relationships between the US and other regions (highest $r^2 = 0.47$ for France and lowest $r^2 = 0.07$ for Réunion). However, according to regressions and optimum differences between the US and the focal region (Fig. 2), pH optima were typically not consistent. Given that Bennett et al. (2010) did not consider diatom niches other than that of pH, we cannot compare our findings for nutrient and climatic niches directly. Moreover, as the five test regions have shorter gradients in environment and climate than the US, it is possible that the species in the other regions may not contain the true species optima.

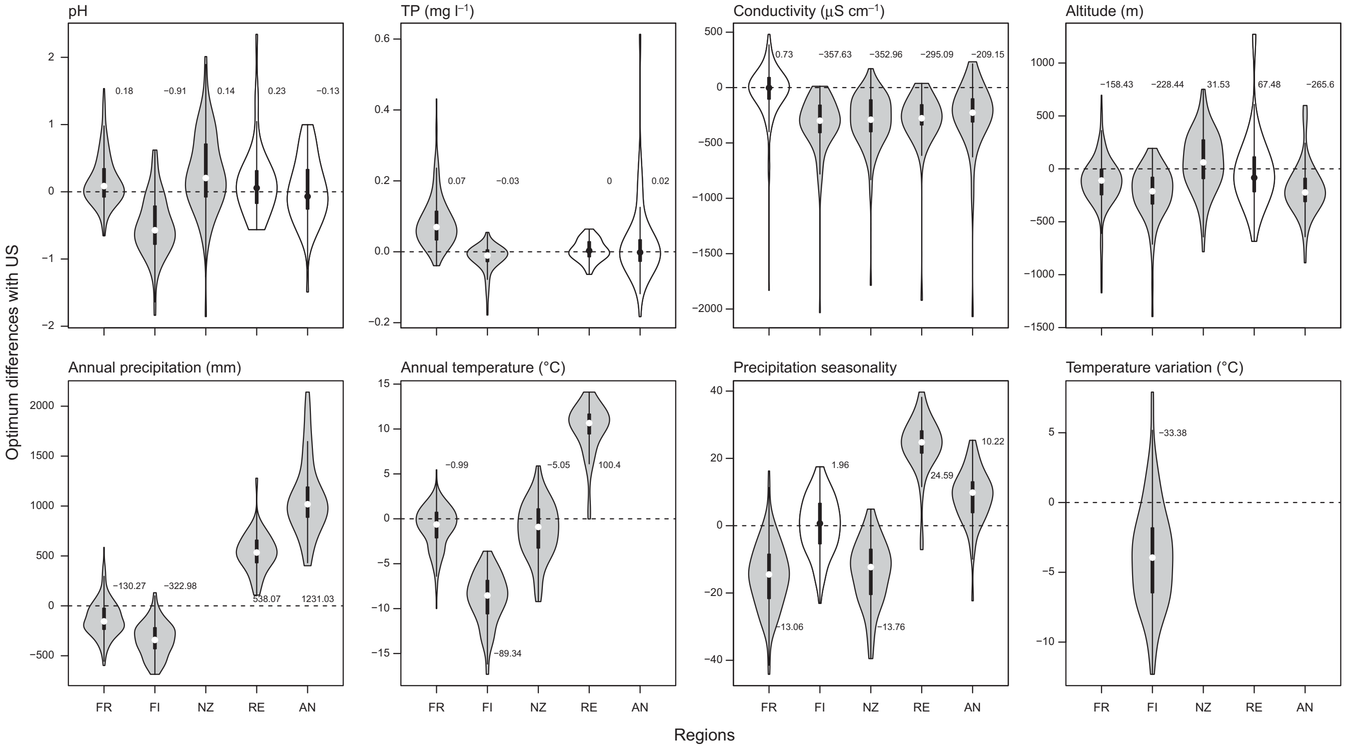

Figure 2. The violin plots showing the differences in species optima in terms of local environmental variables and climatic variables in each five study regions in comparison with the US optima values. Numbers represent the differences between the means of environmental or climate variables for the focal region and US within the whole dataset. White violin indicates non-significant ($p > 0.05$) differences between the US optima and the focal region according to a Wilcoxon signed rank test. Regions: FR = France, FI = Finland, NZ = New Zealand, RE = La Réunion, AN = Antilles. TP = total phosphorus.

Nevertheless, our regression results about species optima also show that NC is perhaps even weaker in climatic niches than in water chemistry niches.

The low NC especially with respect to climatic niches is a noteworthy finding of our study, indicating that diatoms may show some adaptation to local conditions. This notion is supported by the observation of diatom environmental optima matching the prevailing mean and median conditions in the respective study region. For example, temperature optima were the highest for La Réunion, which is located at low latitudes with high mean and median temperatures. Likewise, mean pH optimum for Finnish diatoms was notably the lowest because mean and median water pH in Finland were the lowest compared to the other regions (Soininen et al. 2016). It is possible that diatoms either undergo local adaptation (Sjöqvist et al. 2015) or comprise cryptic species (Beszteri et al. 2007, Souffreau et al. 2013) with potentially different environmental preferences among lineages. What is most interesting here is that such within species variability in diatoms is perhaps more pronounced with respect to climate than the local environment. Traditionally, the variation in diatom community compositions is considered to be driven primarily by local water chemistry variables (Soininen 2007). However, a recent study from boreal streams indicated that diatom distributions are also related to climatic variables, especially energy or temperature-related factors, such as growing degree days (Pajunen et al. 2016). Nonetheless, our present results of lower NC with respect to climate imply that diatoms may adapt better to climate change than organisms with more conserved climate niches, and therefore, be less prone to extinction by climate change.

Our analysis did not support the hypothesis H_1 that diatom guilds would have consistent differences in NC. Our prediction that NC would be the most evident among low profile species was not supported even if earlier observational studies from the same study regions have indicated that diatom traits may be related to occurrences and perhaps to dispersal ability and the degree of NC (Heino and Soininen 2006, Passy 2016). It, therefore, seems that diatom traits conferring ability to tolerate harsh conditions and acquire resources (Passy 2007) are not clearly related to the degree of NC. Likewise, we did not find evidence for core species with broader distributions to be more conserved than the regionally rarer satellite species. This disagrees with our prediction H_2 for higher NC in core species with large niche breadth. This outcome may stem from the fact that core and satellite species did not differ from each other in terms of niche breadth in all study regions. Moreover, it may be that specialist species may have not adapted well in new environments as e.g. acid-tolerant diatom species are found practically only in acid waters globally.

Broad-scale random dispersal allows diatom species to colonize new sites. However, if dispersal is moderate rather than very strong, species may adapt locally and show divergence in niche breadth and optima. Whether local adaptation is associated with genetic and physiological differences that remain undetectable in morphospecies often comprising cryptic species and species complexes or with phenotypic plasticity, is an important question that should be addressed further by future research (Souffreau et al. 2013, Rose and Cox 2014). We emphasize here that the species concept for diatoms is complex and many cryptic species may complicate the investigation of diatom biogeographical patterns (Mann 1999). Likewise, as long as species identification is primarily based on the features of cell morphology and not molecular data, subjective decisions in species identification may also add some noise to the results. Therefore, diatom bioassessment tools based on morphospecies should be applied very carefully especially if study extent is large and thus encompasses long gradients in environmental and climatic conditions.

In conclusion, we demonstrated that diatom species show overall relatively little NC globally. We also documented that NC was even lower for climatic variables than for local environmental variables. This shows that climate models using diatoms e.g. in paleoecological reconstructions may not be directly transferrable in space because species climatic niches differ among geographical regions. Thus, climate models would need regional diatom data sets for calibrating the models. Finally, the low NC with respect to climatic variables indicates that freshwater diatoms may be less sensitive to global change. We encourage researchers to study NC further in other microorganisms and in different environmental settings in order to determine if low NC is a global pattern in microorganisms, which would suggest that global change may have a weaker impact on their distribution compared to macroorganisms.

Acknowledgements – AJ and JR thank Sébastien Boutry for interesting discussion on some statistical analyses. We also thank ONEMA (Office National de l'Eau et des Milieux Aquatiques), all French Water Agencies and Asconit for data contribution and financial support and Anna Astorga and Riku Paavola for collecting the diatom data in New Zealand.

Funding – We gratefully acknowledge financial support from the Emil Aaltonen Foundation (to JS and JW), the National Key Research and Development Program of China (2017YFA0605203 to JW), the CAS Key Research Program of Frontier Sciences (QYZDB-SSW-DQC043 to JW), NSFC (41571058, 41273088 to JW) and the National Science Foundation (grant NSF DEB-1745348 to SP).

References

Adler, D. 2005. vioplot: violin plot. – R package ver. 0.2.

Alahuhta, J. et al. 2017. A comparative analysis reveals little evidence for niche conservatism in aquatic macrophytes among four areas on two continents. – Oikos 126: 136–148.

Anderson, S. 1994. Area and endemism. – Q. Rev. Biol. 69: 451–471.

Bennett, J. R. et al. 2010. Broad-scale environmental response and niche conservatism in lacustrine diatom communities. – Global Ecol. Biogeogr. 19: 724–732.

Beszteri, B. et al. 2007. An assessment of cryptic genetic diversity within the *Cyclotella meneghiniana* species complex (Bacillariophyta) based on nuclear and plastid genes, and amplified fragment length polymorphisms. – Eur. J. Phycol. 42: 47–60.

Brotons, L. et al. 2004. Presence–absence versus presence-only modeling methods for predicting bird habitat suitability. – Ecography 27: 437–448.

Buisson, L. et al. 2008. Climate change hastens the turnover of stream fish assemblages. – Global Change Biol. 14: 2232–2248.

Cadotte, M. W. et al. 2008. Evolutionary history and the effect of biodiversity on plant productivity. – Proc. Natl Acad. Sci. USA 105: 17012–17017.

Case, T. J. and Taper, M. L. 2000. Interspecific competition, environmental gradients, gene flow, and the coevolution of species' borders. – Am. Nat. 155: 583–605.

Danielson, J. J. and Gesch, D. B. 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010). – U.S. Geological Survey Open-File Report 2011-1073.

Dolédec, S. et al. 2000. Niche separation in community analysis: a new method. – Ecology 81: 2914–2927.

Elith, J. et al. 2008. A working guide to boosted regression trees. – J. Anim. Ecol. 77: 802–813.

Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – Environ. Conserv. 24: 38–49.

Finlay, B. 2002. Global dispersal of free-living microbial eukaryote species. – Science 296: 1061–1063.

Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – Ecol. Model. 135: 147–186.

Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – Ecol. Lett. 8: 993–1009.

Heikkinen, R. K. et al. 2012 Does the interpolation accuracy of species distribution models come at the expense of transferability? – Ecography 35: 276–288.

Heino, J. and Soininen, J. 2006. Regional occupancy in unicellular eukaryotes: a reflection of niche breadth, habitat availability, or size-related dispersal capacity? – Freshwater Biol. 51: 672–685.

Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – Int. J. Clim. 25: 1965–1978.

Krammer, K. and Lange-Bertalot, H. 1986–1991. Bacillariophyceae. Süβwasserflora von Mitteleuropa, 2 (1–4). – Gustav Fischer.

Lange-Bertalot, H. (ed.) 1995–2015. Iconographia diatomologica. Annotated diatom micrographs. Vol. 1–24. – Koeltz Scientific Books.

Lange-Bertalot, H. (ed.) 2000–2013. Diatoms of Europe – diatoms of the European inland waters and comparable habitats. Vol. 1–7. – Koeltz Scientific Books.

Lecointe, C. et al. 1993. "Omnidia": software for taxonomy, calculation of diatom indices and inventories management. – Hydrobiologia 269–270: 509–513.

Luoto, M. et al. 2005. Uncertainty of bioclimate envelope models based on geographical distribution of species. – Global Ecol. Biogeogr. 14: 575–584.

Mann, D. G. 1999. The species concept in diatoms. – Phycologia 38: 437–495.

Pajunen, V. et al. 2016. Climate is an important driver for stream diatom distributions. – Global Ecol. Biogeogr. 25: 198–206.

Parmesan, C. and Yohe, G. 2003. A globally coherent fingerprint of climate change impacts across natural systems. – Nature 421: 37–42.

Passy, S. I. 2007. Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters. – Aquat. Bot. 86: 171–178.

Passy, S. I. 2012. A hierarchical theory of macroecology. – Ecol. Lett. 15: 923–934.

Passy, S. I. 2016. Abundance inequality in freshwater communities has an ecological origin. – Am. Nat. 187: 502–516.

Ridgeway, G. et al. 2017. gbm: generalized boosted regression models. – R package ver. 2.1.3.

Rimet, F. and Bouchez, A. 2012. Biomonitoring river diatoms: implications of taxonomic resolution. – Ecol. Indic. 15: 92–99.

Rose, D. T. and Cox, E. J. 2014. What constitutes *Gomphonema parvulum*? Long-term culture studies show that some varieties of *G. parvulum* belong with other *Gomphonema* species. – Plant Ecol. Evol. 147: 366–373.

Sax, D. F. 2001. Latitudinal gradients and geographic ranges of exotic species: implications for biogeography. – J. Biogeogr. 28: 139–150.

Sjöqvist, C. et al. 2015. Local adaptation and oceanographic connectivity patterns explain genetic differentiation of a marine diatom across the North Sea–Baltic Sea salinity gradient. – Mol. Ecol. 24: 2871–2885.

Soininen, J. 2007. Environmental and spatial control of freshwater diatoms – a review. – Diatom Res. 22: 473–490.

Soininen, J. et al. 2013. Stochastic species distributions are driven by organism size. – Ecology 94: 660–670.

Soininen, J. et al. 2016. Global patterns and drivers of species and trait composition in diatoms. – Global Ecol. Biogeogr. 25: 940–950.

Soininen, J. et al. 2018. Data from: Stream diatoms exhibit weak niche conservation along global environmental and climatic gradients. – Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.v1v7856>.

Souffreau, C. et al. 2013. Molecular evidence for distinct Antarctic lineages in the cosmopolitan terrestrial diatoms *Pinnularia borealis* and *Hantzschia amphioxys*. – Protist 164: 101–115.

Stoermer, E. F. and Smol, J. P. (eds) 1999. The diatoms: applications for the environmental and earth sciences. – Cambridge Univ. Press.

Thuiller, W. et al. 2003. Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). – Global Ecol. Biogeogr. 12: 313–325.

Tuszynski, J. 2014. caTools: Tools: moving window statistics, GIF, Base64, ROC AUC, etc. – R package ver. 1.17.1.

Van Dam, H. et al. 1994. A coded checklist and ecological indicator values of freshwater diatoms from Netherlands. – Neth. J. Aquat. Ecol. 28: 117–133.

Vellend, M. 2010. Conceptual synthesis in community ecology. – Q. Rev. Biol. 85: 183–206.

Vyverman, W. et al. 2007. Historical processes constrain patterns in global diatom diversity. – Ecology 88: 1924–1931.

Wang, J. et al. 2011. Contrasting patterns in elevational diversity between microorganisms and macroorganisms. – J. Biogeogr. 38: 595–603

Wasof, S. et al. 2016. Disjoint populations of European vascular plant species keep the same climatic niches. – Global Ecol. Biogeogr. 24: 1401–1412.

Wiens, J. J. and Graham, C. H. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. – Annu. Rev. Ecol. Evol. Syst. 36: 519–539.

Wiens, J. J. et al. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. – Ecol. Lett. 13: 1310–1324.

Supplementary material (Appendix ECOG-03828 at <www.ecography.org/appendix/ecog-03828>). Appendix 1.