Appendix 1

Figure A1

Plots of principal component analysis summarizing the major variation in local environmental and climatic variables in the regions along the first two axes. Gray symbols represent the US sites and purple symbols the sites from the focal region. Values represent the explained variation on each axis.
Figure A2

Proportions (%) of conserved species when different ratios of AUC\textsubscript{us} and AUC\textsubscript{pred} are used for local, climate and full models. Regions: FR = France, FI = Finland, NZ = New Zealand, RE = La Réunion, AN = Antilles.
Figure A3

Proportions of species in each diatom guild to all conserved species and all (total) species by study region and model. The significance (**p < 0.01, ***p < 0.001) for the species in each guild to be more common among conserved or all (total) species than the other species are given based on $\chi^2$ tests. Guild abbreviations: acid = acid tolerant, high = high profile, low = low profile.
Figure A4

Ratios of species distribution model Area Under the Curve (AUC) values in the US (AUC\textsubscript{us}) and predicted values in the other study regions (AUC\textsubscript{pred}) grouped by core (> 5% of sites) and satellite (≤ 5% of sites) species. Shown are the results for models based on local environmental variables (local models, upper panels), climatic variables (climate models, middle panels) and all variables (full models, lower panels).
Figure A5

Proportion of core vs. satellite species out of all conserved species and all (total) species by study region and model. The significance (**p < 0.01, *p < 0.05) for the core or satellite species to be more common among conserved or all (total) species than the other species are given based on $\chi^2$ tests.
Figure A6

Regression plots for the relationships of species environmental optima between the US and other study regions. The black lines depict the identity (1:1) lines and red lines the significant (p < 0.05) fits to the data. The continuous red lines indicate that the regression slopes and intercepts do not differ significantly (p < 0.05) from the identity line according to ANOVA. The adjusted $R^2$ values are also indicated.
Figure A7

Ratios of niche breadth in the US vs. other regions by core and satellite species and model. The significance levels (**p < 0.01, *p < 0.05) of tests on whether that the species niche breadth ratio > 1 is more common than the ratio of < 1 in core or satellite species are derived from \( \chi^2 \) tests. See Figure A4 for other details.
Figure A8
Ratios of niche breadth in the US vs. other regions by diatom guild and model. We tested whether the niche breadth ratio > 1 is more common than the ratio of < 1 among species in each guild using $\chi^2$ tests ($***p < 0.001$, **$p < 0.01$, *$p < 0.05$). See Figure A3 for other details.
### Figure A9

Relationships between the number of occupied sites (log-transformed) and species niche breadths in the US models (calibration models) and test set models. Regression lines indicate significant (p < 0.05) relationships. Oth = other.
Figure A10

Relationships between species niche breadth and model AUC values in the US models (calibration models) and test set models. Regression lines indicate significant ($p < 0.05$) relationships.