



Environmental optima for common diatoms from Ontario lakes along gradients of lakewater pH, total phosphorus concentration, and depth

Matthew P. Duda · Branaavan Sivarajah · Kathleen M. Rühland · Andrew M. Paterson · Jennifer L. Barrow · Yuanyu Cheng · Elizabeth J. Favot · Kristopher R. Hadley · Roland I. Hall · Kathryn E. Hargan · Clare Nelligan · Euan D. Reavie · Robin E. Valleau · Petra Werner · Christopher Wilkins · John P. Smol

Received: 15 December 2022 / Accepted: 25 May 2023 / Published online: 17 June 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract Diatoms are powerful biomonitoring indicators for contemporary and past limnological conditions, provided that their environmental optima are well-defined. Surface-sediment-calibration (or training) sets are widely used to develop quantitative transfer functions linking species distributions to limnological variables. Accurate estimates of each taxon's environmental optimum benefit from a large dataset spanning a broad gradient of the variable of interest.

Here, we compiled data from surface-sediment-diatom analyses from over 450 lakes in Ontario (Canada) to identify ecological optima for the province's most common taxa. The data were collected from the same laboratory using similar methods and spanned more than three decades of research (1987–2019) on 546 samples (representing 464 unique sites). We describe the ecological responses of 52 common diatom taxa to three limnological variables that are key to lake management: pH, total phosphorus (TP), and maximum lake depth. The gradients were broad (pH: 4.2–9.3; TP: 1.1–96.8 $\mu\text{g L}^{-1}$; maximum depth: 0.7–95.0 m), enabling robust statistical analyses. We used

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10933-023-00288-7>.

M. P. Duda · K. M. Rühland · J. L. Barrow · Y. Cheng · E. J. Favot · C. Nelligan · R. E. Valleau · C. Wilkins · J. P. Smol (✉)
Paleoecological Environmental Assessment and Research Lab (PEARL), Department of Biology, Queen's University, Kingston, ON K7L 3N6, Canada
e-mail: smolj@queensu.ca

B. Sivarajah
Department of Geography and Environment, Mount Allison University, 144 Main St., Sackville, NB E4L 1A7, Canada

A. M. Paterson
Environmental Monitoring and Reporting Branch, Ontario Ministry of Environment, Conservation and Parks, Dorset, ON, Canada

K. R. Hadley
Hutchinson Environmental Sciences Ltd., 1-5 Chancery Ln., Bracebridge, ON P1L 2E3, Canada

R. I. Hall
Department of Biology, University of Waterloo, Waterloo, ON N2L 3G1, Canada

K. E. Hargan
Department of Biology, Memorial University of Newfoundland, St. John's, NL A1B 3X9, Canada

E. D. Reavie
Natural Resources Research Institute (NRRI), University of Minnesota Duluth, 5013 Miller Trunk Highway, Duluth, MN 55811, USA

P. Werner
Diatoms as Bioindicators, Grainauer Str. 8, 10777 Berlin, Germany

hierarchical logistic regression modelling to identify statistically significant ecological response curves for all common taxa, and weighted averaging to identify the optima. In our dataset, 96% of common diatom taxa had significant ecological responses and optima for pH, 81% for TP, and 83% for depth. Aside from the benthic cosmopolitan *Achnanthydium minutissimum* complex (occurring in 88.1% of samples), the most frequently recorded taxa were planktonic, including *Discostella stelligeralpseudostelligera* (83.2%), *Asterionella formosa* (77.8%), and the *Lindavia bodanica* complex (76.6%). Our results illustrate the responses of diatom taxa across key limnological gradients, providing important autecological information to advance their use as bioindicators in environmental assessments. These data can be used to address management and knowledge gaps in understudied areas of Ontario where calibration sets are not available and can support research in regions with similar limnological environments and diatom communities.

Keywords Biomonitoring · Algae · Sediment · Calibration set · Optimum

Introduction

Diatoms (Class: Bacillariophyceae) are unicellular, eukaryotic organisms that are ubiquitous across the aquatic continuum (Smol and Stoermer 2010), occupying diverse aquatic habitats that span broad environmental gradients (e.g., pH, nutrient levels, metal(loid) concentrations, ionic strength, lake depth; Round et al. 1990). These siliceous algae, with at least 30,000 extant species (Mann and Vanormelingen 2013), form the base of most freshwater and marine food webs and play a key role in aquatic biogeochemical processes (Smol and Stoermer 2010). Because diatoms respond rapidly and predictably to changes in the chemical and physical conditions of their surroundings, have fast reproduction rates, and have siliceous frustules which readily preserve in most sediments (Julius and Theriot 2010), they are ideal bioindicators of contemporary and past environmental conditions across timescales spanning seasons to millennia (Smol and Stoermer 2010).

Diatoms have been extensively used as bioindicators in paleolimnological studies, tracking long-term

limnological responses to anthropogenic stressors including acidification, eutrophication, climate impacts, industrial contamination, and salinization (Smol 2008; Smol and Stoermer 2010). For example, during the acid rain debates of the 1970s to 1990s, diatom-based approaches provided unequivocal qualitative (i.e., species changes) and quantitative (i.e., reconstructions of lakewater pH) evidence that many low-alkalinity lakes in northeastern North America and Europe acidified because of anthropogenic activities (Battarbee 1990; Cumming et al. 1992; Dixit et al. 1992).

Diatom autecological information is commonly derived from surface-sediment calibration sets (also known as training sets) that relate the modern distribution of taxa to current environmental measurements (Birks et al. 1990; Fritz et al. 1993; Wilson et al. 1994; Hall and Smol 1996; Ramstack et al. 2003; Chen et al. 2008; Juggins and Birks 2012). This approach provides estimates of taxon optima and tolerances, which can then be used to infer past limnological conditions related to shifts in the composition of downcore sedimentary diatom assemblages. Generally, a minimum of 40 to 70 sample sites are used in calibration datasets to encompass the gradient of limnological characteristics observed within a study area (Reavie and Juggins 2011). Localized calibration datasets can also be merged to create larger calibration sets, capturing broader environmental conditions that can yield a more accurate estimate of a taxon's ecological range (Juggins and Birks 2012). For example, to assess eutrophication trends in a lake in the province of British Columbia (Canada), Cumming et al. (2015) developed a 251-lake-calibration dataset along a total phosphorus (TP) gradient of 2–227 $\mu\text{g L}^{-1}$ and quantified TP optima and tolerances for 48 common diatom taxa. Similarly, large databases such the European Diatom Database (EDDI; Battarbee et al. 2001a, b), as well as the “Omnidia” software (Lecointe et al. 1993), provide diatom optima and tolerances that are widely applied throughout Europe. Optima and tolerances calculated using large datasets are often more robust and have the potential to be applied in regions where localized calibration sets are not available because these analyses are time consuming and require taxonomic expertise. Diatom autecological information derived from large calibration sets can be an attractive and useful tool for biomonitoring assessments and lake management.

In this study, we synthesized surface-sediment diatom relative abundance data from 20 multi-lake studies conducted between 1987 and 2019 at the Paleocological Environmental Assessment and Research Lab (PEARL) at Queen’s University (Kingston, Ontario, Canada) to estimate optima for lakewater pH, TP concentration, and maximum lake depth for common diatom taxa in Ontario lakes (Table 1). This dataset is comprised of 546 samples from 464 unique sites (some lakes were resampled in varying years and, at some lakes, multiple depositional basins were sampled). The sampled lakes span Ontario’s three ecozones (Hudson Bay Lowlands, Ontario Shield, Mixedwood Plains; Fig. 1), and are therefore diverse in terms of limnological and physical features.

We focused on the relationships between the most common diatom taxa in Ontario and lakewater pH, TP concentrations, and maximum depth. These limnological variables were selected as they have well-established relationships with diatoms, are often correlated to other important limnological variables (e.g., buffering capacity, light, etc.), and are commonly

measured when conducting spatial surveys. Moreover, lakewater pH and TP concentration are important water-quality variables to measure lake acidification and eutrophication, respectively, which are major stressors for freshwater ecosystems (Schindler 1988; Anderson et al. 2002; Schindler et al. 2008; Keller 2009; Van Staden et al. 2022). The strong response of diatom taxa to changes in TP concentrations and pH makes them excellent bioindicators and form the basis of quantitative inference models for reconstructing past environmental conditions (Birks et al. 1990; Hall and Smol 1996; Battarbee et al. 2001a, b; Reavie and Smol 2001; Tremblay et al. 2014; Cumming et al. 2015). In contrast, quantitatively reconstructing lake depth using diatoms from a set of calibration lakes has many challenges and is therefore not a common practice. However, quantitative reconstructions are not the goal of this study, and we rationalize including an exploration of diatom distributions across maximum lake depth as the relationship between lake depth and key taxa can help clarify diatom responses to acidification and eutrophication across a diverse

Table 1 Details of the studies included in this synthesis, separated by lead researcher

Ecozone(s)	Study region(s)	No. of sample sites	Year(s) collected	Reference(s)
Hudson Bay Lowlands	Sub-Arctic	13	2009–2010	Rühland et al. (2014)
Ontario Shield/Hudson Bay Lowlands	Ring of Fire	58	2011	Hargan, unpublished
Ontario Shield	Sudbury and Killarney	81	1987–1992; 1997–1999	Dixit et al. (1989, 1991, 2002)
Ontario Shield	Muskoka-Haliburton	54	1992	Hall and Smol (1996)
Ontario Shield	Experimental Lakes Area	34	2006	Enache et al. (2011)
Ontario Shield	Wawa, Algoma District	6	2006	Greenaway et al. (2012)
Ontario Shield	Muskoka-Haliburton	45	2009	Hadley et al. (2013)
Ontario Shield	Muskoka-Haliburton	31	2007–2008	Barrow et al. (2014)
Ontario Shield	Algonquin Park, Algoma District, Lake Nipissing, Muskoka	15	2015, 2017	Favot et al. (2019), Favot (2021)
Ontario Shield	Algonquin Park	5	2015	Wilkins (2021)
Ontario Shield	Sudbury and Killarney	80	2018	Cheng et al. (2022)
Ontario Shield/Mixedwood Plains	Southeast Ontario	64	1998	Reavie and Smol (2001)
Ontario Shield/Mixedwood Plains	Muskoka-Haliburton; Lake Manitou	8	2013, 2015	Nelligan et al. (2016), (2020)
Ontario Shield/Mixedwood Plains	Muskoka Haliburton, Oak Ridges Moraine	8	2016, 2019	Valleau (2021), Valleau et al. (2022)
Mixedwood Plains	Bruce Peninsula	44	1999	Werner and Smol (2005)
Full lake set		546	1987–2019	

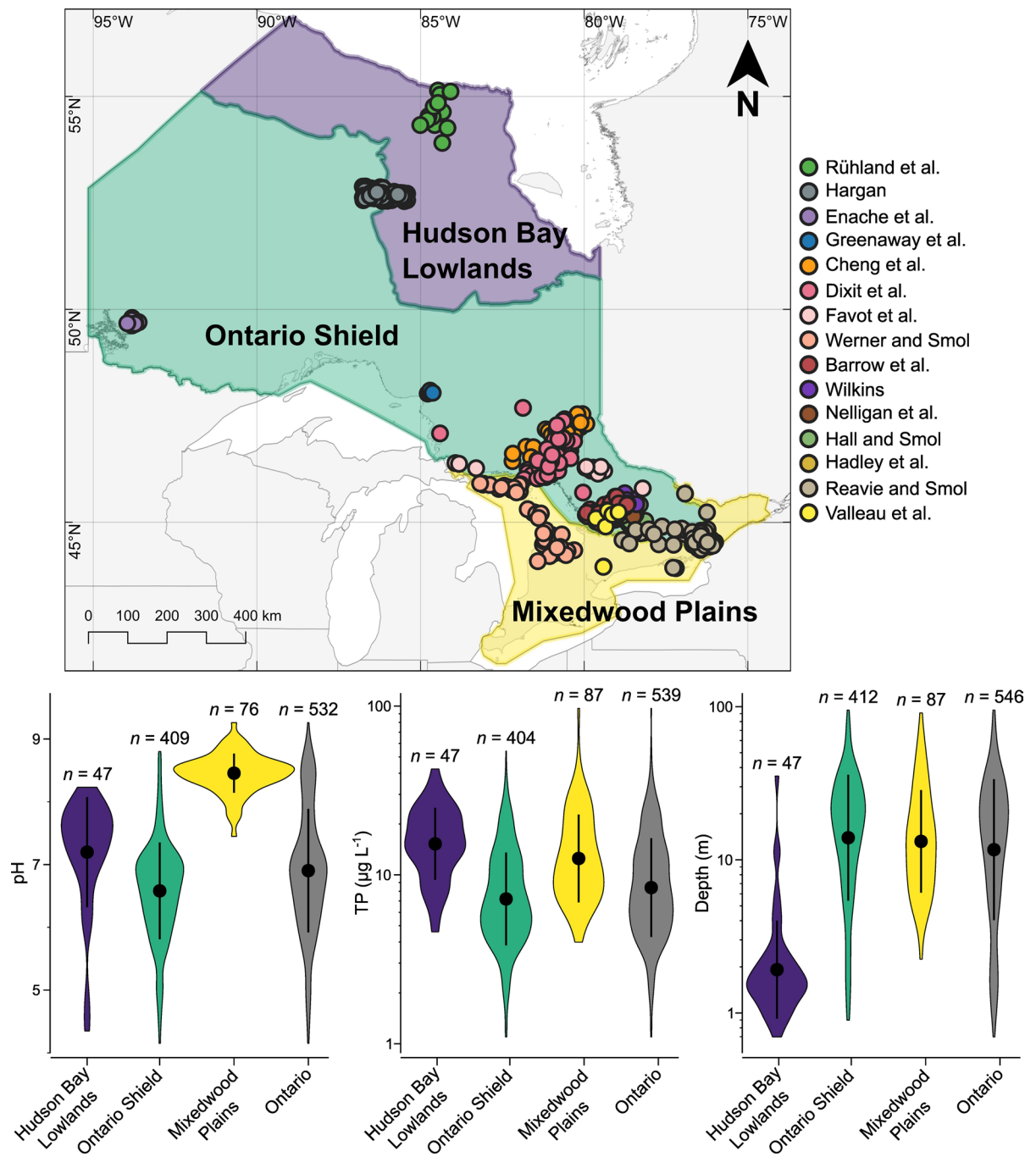


Fig. 1 Map of Ontario showing the ecozones and locations of the study lakes, and corresponding violin plots depicting the distribution of lakewater pH, total phosphorus (TP) concentra-

tion, and maximum depth. Note that TP and depth are shown on a log₁₀ scale. The violin plots illustrate the median as a dot and the interquartile range as whiskers

suite of lakes. For instance, diatom-assemblage responses to increasing TP concentrations could be vastly different in a shallow lake relative to a deeper

lake, because environmental factors associated with lake depth (e.g., light availability, habitat structure) are substantially different between the two systems.

Our objectives were three-fold: 1) to identify the most common diatom taxa observed in the surface sediment of Ontario lakes; 2) to estimate the environmental optima for lakewater pH, TP concentration, and maximum depth for the most common taxa; and 3) to estimate the environmental range across Ontario where the common diatoms occur and flourish. By refining diatom autecology and advancing our understanding of the environmental conditions in which diatom taxa are distributed, this study will contribute to the more effective use of diatoms as bioindicators in environmental assessments.

Methods

Site selection and sample collection

We merged 20 Ontario surface sediment calibration datasets and multi-lake studies collected by a single research program (PEARL at Queen's University, Kingston, ON), thereby minimizing methodological variability (Table 1). All studies used comparable methods for water chemistry and sediment collection, as well as for diatom preparation and enumeration. Sediment cores were retrieved from the deepest basin of each lake (or where the deepest point was identified using a depth sounder near the centre of the lake) using a Glew-type gravity corer (Glew 1989, 1991) or a modified Kajak-Brinkhurst gravity corer (used in Dixit et al. 1991, 1992). Occasionally, multiple cores were collected from lakes with complex bathymetry (i.e., with two or more depositional areas such as Lake Manitou (Nelligan et al. 2020) and Lake Nipissing (Favot 2021)). The top-most sediment samples were sectioned at 0.25, 0.5, or 1 cm intervals using a vertical extruder (Glew 1988) and correspond to limnological data collected from the surface waters of the same lake at the time of coring, or as part of ongoing water-quality-monitoring programs. Water-chemistry measurements, including TP concentration and pH, were analyzed in a variety of laboratories, but all used similar well-established methodologies (OME 1979, 1981, 1983; Environment Canada 1994a, b; Janhust 1998). In some cases, lakewater pH was measured while collecting the sediment core using calibrated handheld pH meters. We categorized the limnetic TP concentrations following the classification scheme for Canadian waters by Wetzel (2001), where lakes

with TP concentrations $< 12 \mu\text{g L}^{-1}$ are oligotrophic, between 12 and $24 \mu\text{g L}^{-1}$ are mesotrophic, between 24 and $100 \mu\text{g L}^{-1}$ are eutrophic, and $> 100 \mu\text{g L}^{-1}$ are hypereutrophic. Lake-depth measurements were taken concurrent with sediment-core retrieval. In most cases, sediment cores were collected from the deepest point of the lake, thus corresponding with maximum depth.

Diatoms from all waterbodies used in this study (546 samples, 464 unique sites) were prepared following standard techniques used at PEARL as described in Rühland and Smol (2002). Briefly, 0.2 to 0.5 g wet sediment (or 0.02 to 0.05 g freeze-dried sediment) was subsampled and placed into glass scintillation vials. To remove the organic components, ~15 ml of a 1:1 molecular weight ratio of concentrated sulphuric and nitric acid was added to each sample, heated to ~80 °C for at least 2 h, and rinsed repeatedly with deionized water to achieve circumneutral pH before preparing permanent microscope slides using Hyrax® or Naphrax® mounting medium.

Diatoms were identified to at least the species level (often to variety) using a research-grade compound microscope under oil immersion at 1000× magnification. For each sample, at least 250 diatom valves, but often more than 400, were enumerated. Taxonomic identifications were based on a variety of standard sources including Krammer and Lange-Bertalot (1986–1991), Patrick and Reimer (1966), Camburn and Charles (2000), Reavie and Kireta (2015) and published articles (e.g., Koppen 1975; Siver and Kling 1997).

Taxonomic harmonization

All contributing taxonomists were trained at the Paleoeological Environmental Assessment and Research Laboratory (PEARL) at Queen's University, Kingston, Ontario. However, since diatom identification has evolved and advanced over the study period, we used a series of conservative harmonization techniques to establish taxonomic units that helped minimize Type I errors. We consider taxonomic units to be a group of morphologically similar taxa that are generally representative of comparable ecological conditions.

First, we re-calculated relative abundances from raw count data to avoid errors that may have been introduced through data transformations. Next,

imprecise taxa identifications were omitted from further analyses (e.g., those that were identified only to genus level, aff. (*affinis*), cf. (*conferre*), unknowns). Taxon names were then updated to the most current nomenclature based on recent publications and online databases (e.g., Spaulding et al. 2022; Guiry and Guiry 2022). Finally, some taxa were grouped into taxonomic complexes to address potential concerns that may arise due to differences in: 1) taxonomic precision; 2) identification of taxa that are morphologically similar and challenging to distinguish using light microscopy; and 3) nomenclature over time. Diatom taxa that were grouped with their varieties include: *Achnantheidium minutissimum*, *Asterionella ralfsii*, *Aulacoseira perglabra*, *Cocconeis placentula*, *Fragilaria capucina*, *Frustulia rhomboides*, *Pinnularia microstauron*, *Sellaphora pupula*, and *Staurisirella pinnata*. Other taxon groupings based on identification imprecision include *Discostella stelligera/pseudostelligera*, *Pantocsekiella comensis/gordonensis*, *Fragilaria delicatissimaltenera/nanana*, *Navicula cryptocephala/cryptotenella*, the *Cymbella gaeumannii* complex (*Encyonopsis falaisensis*, *Encyonema gaeumannii*, *E. perpusillum*), the *Fragilaria ulna* complex (*F. ulna* varieties, *Ulnaria acus*, *Synedra acus* var. *angustissima*), the *Fragilaria virescens* complex (*Fragilariforma virescens* varieties, *Stauroforma exiguiformis*), and the *Lindavia bodanica* complex (*Lindavia bodanica* varieties, *L. comta* varieties, *L. affinis*, *L. intermedia*).

After harmonization, taxon names were updated to the most recently accepted entity. Notable updates include all *Synedra* spp. renamed to *Fragilaria* spp. (as per Alexson et al. 2022), *Cyclotella tripartita* and *Lindavia tripartita* to *Pantocsekiella tripartita* (as per Ács et al. 2016), *Lindavia ocellata* to *Pantocsekiella ocellata* (as per Ács et al. 2016), *Cyclotella michiganiana* and *Lindavia michiganiana* to *Pantocsekiella michiganiana* (as per Schultz and Dreßler 2022), and *Cymbella microcephala* to *Encyonopsis microcephala* (as per Krammer 1997). The species authorities for the common diatom taxa, and the individual taxa within a given complex, are provided in Supplementary Table 1.

Statistical analyses

Histograms of the key limnological variables were used to visualize the distribution of lakewater pH,

TP concentration, and depth in the surface-sediment-calibration set (Supplementary Figures S1–S5). Total phosphorus concentration and maximum depth were right-skewed and therefore were normalized using a \log_{10} transformation. We used violin plots to visualize the distribution of the three environmental variables across ecozones (Fig. 1), created using the R package ‘ggplot2’ v. 3.3.5 (Wickham 2016) in the R workspace v.1.4.1717 (R Core Team 2020).

We focused subsequent analyses on “common taxa” in the harmonized dataset. We defined “common taxa” as those with a Hill’s N2 diversity index (Hill 1973) greater than 10, occurring in at least 10% of samples (i.e., 55 occurrences), and having a relative abundance of at least 10% in at least one sample. These cut-off criteria were selected to create an informative and broadly applicable dataset that considers taxonomic changes and potential misidentification of rare taxa resulting from multiple researchers over three decades.

We used hierarchical logistic regression modeling with niche types and species responses, *sensu* Huisman et al. (1993), to visualize the distribution of taxa along gradients of lakewater pH, TP concentration, and depth. Huisman et al. (1993) outlined five potential response shapes (termed models) applied to an explanatory variable: (I) flat response (i.e., no relationship), which we define as a null response; (II) monotone sigmoid reaching the peak at one end of the gradient; (III) monotone sigmoid with plateau; (IV) unimodal symmetric; and (V) unimodal skewed. Two additional response curves were added by Jansen and Oksanen (2013), representing bimodal ecological response curves. However, considering that unimodal response curves are more representative of our sites across ecozones in Ontario, we restricted our results to models I–V. The most parsimonious model for each response variable was determined using Akaike Information Criterion (AIC), following bootstrapping with 100 permutations to ensure model stability. Model selection was performed using the R package ‘eHOF’ v.1.11 (Jansen and Oksanen 2013).

Lakewater pH, TP concentration, and depth optima were calculated for the common diatom taxa with significant ecological responses (i.e., models II–V) using a weighted-average (WA) approach (ter Braak and Looman 1986) and assessed through bootstrapping with 1000 Monte Carlo permutations. Optima were not calculated for the model I responses as this

model represents a lack of response across the gradient of a limnological variable. We also calculated niche borders to identify a taxon’s range of ecological responses in place of using a Gaussian standard deviation approach. Because ecological responses are often skewed in distribution (e.g., models II, III, V), borders associated with curve maxima provide a more representative description of a taxon’s tolerance to a specific variable (Jansen and Oksanen 2013). Here, the outer borders were calculated by multiplying the maximum response with the coefficient e^{-2} , following Heegaard (2002). All niche borders were truncated to the range of the measured gradient of the dataset and did not extrapolate beyond the presented data. Optima and borders were calculated using the R packages ‘analogue’ v.0.17–5 (Simpson and Oksanen 2020) and ‘eHOF’ (Jansen and Oksanen 2013), respectively.

Results

The majority of the 546 surface sediment samples in our regional dataset are from the Ontario Shield ecozone (412 lakes; ~75% of total), followed by Mixedwood Plains (87 lakes; ~16%), and finally Hudson Bay Lowlands (47 lakes; ~9%; Fig. 1; Table 2). The number of sites examined differ for each environmental variable because lakewater pH and TP concentration were not measured at each site (etc.). Ontario Shield lakes had the lowest median pH (6.7), followed by Hudson Bay Lowlands (7.4), and Mixedwood Plains (8.5; Fig. 1; Table 2). Median TP concentration was the lowest in Ontario

Shield lakes ($6.7 \mu\text{g L}^{-1}$), followed by Mixedwood Plains ($11.2 \mu\text{g L}^{-1}$), and Hudson Bay Lowlands ($14.8 \mu\text{g L}^{-1}$; Fig. 1; Table 2). Median depth was much shallower in the Hudson Bay Lowlands sites (1.6 m) in comparison to the Mixedwood Plains (13.4 m) and Ontario Shield (17.2 m; Fig. 1; Table 2). Based on median values of the full dataset of 464 unique sites, lakes were generally circumneutral (pH=6.9), oligotrophic (TP= $8.0 \mu\text{g L}^{-1}$), and moderately deep (depth=14.8 m; Fig. 1; Table 2).

Following data harmonization there were 728 unique diatom taxonomic units in the full lake set. After applying the cut-off criteria to identify common taxa (Hill’s $N2 \geq 10$, no. of occurrences ≥ 55 , relative abundance $\geq 10\%$ in at least one sample), 52 diatom taxonomic units were considered common in the lake set and are described in Table 3. The top ten taxa and taxon groups include: (1) the *Achnanthydium minutissimum* complex (occurring in 88.1% of samples); (2) the *Discostella stelligera/pseudostelligera* complex (83.2%); (3) *Asterionella formosa* (77.8%); (4) the *Lindavia bodanica* complex (76.6%); (5) the *Fragilaria delicatissima/tenera/nanana* complex (69.2%); (6) *Staurosirella pinnata* (69.0%); (7) *Tabellaria flocculosa* strain IIIp (65.8%); (8) *Aulacoseira ambigua* (63.4%); (9) the *Fragilaria capucina* complex (59.3%); and (10) *Pseudostaurosira brevistriata* (55.9%).

With the exception of the *Cymbella gaeumannii* complex and *Platessa conspicua*, 96% of common taxa had a significant ecological response to pH (models II–V; Table 3). The most frequent diatom response to pH, as determined using AIC was the symmetrical

Table 2 Comparison across ecozones of the three environmental variables examined in this study

Ecozone	Variable	n	Min	Max	Mean	Median	SD
Hudson Bay Lowlands	pH	47	4.4	8.2	7.2	7.4	0.9
	TP ($\mu\text{g L}^{-1}$)	47	4.6	42.4	17.1	14.8	8.0
	Depth (m)	47	0.7	35.2	3.0	1.6	5.3
Ontario Shield	pH	409	4.2	8.8	6.6	6.7	0.8
	TP ($\mu\text{g L}^{-1}$)	404	1.1	54	8.9	6.7	6.4
	Depth (m)	412	0.9	95	19.8	17.2	15.2
Mixedwood Plains	pH	76	7.5	9.3	8.5	8.5	0.3
	TP ($\mu\text{g L}^{-1}$)	87	4.0	96.8	15.3	11.2	12.4
	Depth (m)	87	2.3	91	17.9	13.4	15.8
Full lake set	pH	532	4.2	9.3	6.9	6.9	1.0
	TP ($\mu\text{g L}^{-1}$)	539	1.1	96.8	10.6	8.0	8.4
	Depth (m)	546	0.7	95	18.0	14.8	15.4

TP is the total phosphorus concentration; depth is maximum lake depth; n is the number of sampling sites

Table 3 Lakewater pH, TP concentration, and depth optima of the 52 most common taxa in our 546-sample diatom dataset

Taxon	n	Max %	Hill's N2	pH		TP ($\mu\text{g L}^{-1}$)		Depth (m)	
				Optimum	Model	Optimum	Model	Optimum	Model
<i>Achnanthydium minutissimum</i> complex	481	52.4	153.0	7.40	V	9.61	II	8.2	II
<i>Amphora libyca</i>	64	10.4	11.4	7.96	IV	N/A	I	N/A	I
<i>Amphora pediculus</i>	75	10.5	25.5	8.16	IV	8.30	II	N/A	I
<i>Asterionella formosa</i>	425	46.3	187.8	6.96	V	7.44	V	20.9	V
<i>Asterionella ralfsii</i>	207	70.8	78.1	6.01	III	5.63	V	20.4	V
<i>Aulacoseira ambigua</i>	346	58.9	115.6	7.17	II	9.40	IV	13.6	IV
<i>Aulacoseira distans</i>	162	17.2	63.5	6.11	V	7.61	II	12.5	V
<i>Aulacoseira lirata</i>	201	13.5	83.5	6.10	V	6.08	V	14.2	IV
<i>Aulacoseira nygaardii</i>	92	10.5	30.3	6.19	IV	N/A	I	9.2	V
<i>Aulacoseira perglabra</i>	154	22.5	52.8	6.16	IV	5.13	II	17.3	IV
<i>Aulacoseira subarctica</i>	232	35.7	82.6	7.29	V	8.84	II	16.9	V
<i>Aulacoseira tenella</i>	127	15.0	57.6	6.53	IV	5.74	IV	15.6	III
<i>Brachysira vitrea</i>	299	26.0	78.7	7.27	II	9.09	IV	8.0	II
<i>Chamaepinnularia mediocris</i>	165	14.0	38.4	6.08	II	N/A	I	4.4	II
<i>Cocconeis placentula</i>	80	13.4	16.8	7.94	IV	12.80	IV	N/A	I
<i>Cymbella gaeumannii</i> complex	138	18.4	23.8	N/A	I	6.22	II	N/A	I
<i>Discostella stelligera/pseudostelligera</i>	454	76.9	218.6	6.68	V	6.23	V	18.5	III
<i>Encyonema hebridicum</i>	121	11.4	33.0	6.36	IV	N/A	I	3.0	II
<i>Encyonopsis microcephala</i>	181	15.6	42.5	8.06	V	8.86	II	7.6	V
<i>Eolimna minima</i>	157	10.5	65.9	7.31	IV	13.50	IV	2.9	V
<i>Eunotia exigua</i>	147	33.8	24.5	6.05	II	5.63	V	15.3	IV
<i>Eunotia incisa</i>	154	10.2	50.9	6.24	IV	N/A	I	6.6	IV
<i>Eunotia rhomboidea</i>	89	23.8	14.4	6.17	IV	N/A	I	N/A	I
<i>Eunotia zasuminensis</i>	83	11.3	25.9	6.09	IV	N/A	I	14.8	V
<i>Fragilaria capucina</i> complex	324	57.4	60.5	7.35	IV	9.88	V	9.8	I
<i>Fragilaria crotonensis</i>	224	46.1	99.4	7.73	III	9.64	V	22.1	III
<i>Fragilaria delicatissima/tenera/nanana</i>	378	55.4	87.7	7.27	III	8.28	V	16.2	V
<i>Fragilaria ulna</i> complex	129	29.7	14.5	7.30	II	10.34	II	11.3	IV
<i>Fragilaria virescens</i> complex	205	47.8	35.5	6.06	II	7.90	V	7.2	V
<i>Frustulia rhomboides</i> complex	250	30.3	66.2	5.80	II	N/A	I	9.2	IV
<i>Lindavia bodanica</i> complex	418	51.2	111.9	7.23	IV	7.67	V	23.2	III
<i>Navicula cryptocephala/cryptotenella</i>	277	15.6	73.9	7.73	V	9.93	IV	5.1	II
<i>Navicula seminuloides</i>	70	11.0	23.2	6.61	IV	N/A	I	N/A	I
<i>Nitzschia fonticola</i>	107	16.3	23.2	7.11	IV	14.29	IV	N/A	I
<i>Nitzschia perminuta</i>	191	15.2	64.6	6.84	IV	12.42	V	3.3	II
<i>Pantocsekiella comensis/gordonensis</i>	88	80.8	41.1	8.44	V	8.45	V	17.0	III
<i>Pantocsekiella michiganiana</i>	152	16.5	58.2	8.47	V	8.01	V	13.8	V
<i>Pantocsekiella ocellata</i>	174	19.0	66.9	6.91	IV	5.77	V	26.8	III
<i>Pantocsekiella tripartita</i>	55	12.1	25.2	6.58	IV	3.38	III	24.7	IV
<i>Pinnularia interrupta</i>	81	13.3	29.5	6.78	IV	13.94	IV	1.9	V
<i>Pinnularia microstauron</i>	146	13.2	45.1	6.15	II	10.19	IV	3.6	IV
<i>Platessa conspicua</i>	65	18.2	11.2	N/A	I	N/A	I	N/A	I
<i>Psammothidium sacculus</i>	67	13.8	25.8	7.92	III	11.45	IV	2.8	V
<i>Pseudostaurosira brevistriata</i>	305	21.6	110.4	7.63	V	10.17	II	N/A	II

Table 3 (continued)

Taxon	n	Max %	Hill's N2	pH		TP ($\mu\text{g L}^{-1}$)		Depth (m)	
				Optimum	Model	Optimum	Model	Optimum	Model
<i>Sellaphora pupula</i>	286	10.4	90.0	7.02	IV	12.00	IV	3.7	V
<i>Staurosira construens</i>	227	53.6	43.1	7.65	V	12.82	III	5.6	III
<i>Staurosira construens</i> var. <i>venter</i>	261	66.1	63.8	7.22	V	13.73	V	2.5	V
<i>Staurosirella pinnata</i>	377	49.9	111.9	7.41	V	10.77	III	6.2	IV
<i>Stephanodiscus minutulus</i>	64	37.0	19.8	8.41	V	9.78	III	27.0	V
<i>Tabellaria flocculosa</i> strain III	121	32.4	24.1	6.05	IV	7.41	IV	19.1	IV
<i>Tabellaria flocculosa</i> strain IIIp	359	38.4	184.5	6.56	V	7.12	V	18.8	IV
<i>Tabellaria flocculosa</i> strain IV	259	14.3	89.2	6.20	IV	6.56	IV	16.5	IV

n is the number of appearances in the dataset and max % is the maximum relative abundance reported for a given taxon in the dataset. Model I indicates a null response; model II a monotone sigmoid response; model III a monotone sigmoid with plateau response; model IV a unimodal symmetric response; and model V a unimodal skewed response

and unimodal response of model IV (43%; Table 3). Most of the common taxa (83%, 42/52) had a significant ecological response to \log_{10} -transformed TP with the asymmetric and unimodal response of model V being the most frequent (30%; Table 3). For maximum lake depth, 85% (44/52) of the taxa had an ecologically significant response to \log_{10} -transformed depth, with the most common being an asymmetrical and unimodal response of model V (30%; Table 3).

In our Ontario dataset, diatom taxa generally had highly variable pH border ranges, indicating a broad range of tolerance spanning multiple pH units (Fig. 2). The taxa most indicative of acidic conditions were the *Frustulia rhomboides* complex (pH optimum: 5.80), *Asterionella ralfsii* (6.01), *Tabellaria flocculosa* strain III (6.05), *Eunotia exigua* (6.05), and the *Fragilaria virescens* complex (6.06; Fig. 2; Table 3). On the other end of the pH gradient, taxa with the highest pH optima included *Pantocsekiella michiganiana* (8.47), *Pantocsekiella comensis/gordonensis* (8.44), *Stephanodiscus minutulus* (8.41), *Amphora pediculus* (8.16), and *Encyonopsis microcephala* (8.06; Fig. 2; Table 3). Several common taxa had narrow border ranges relative to the dataset and unimodal responses to pH, suggesting that they are effective as pH indicators, including *Pantocsekiella michiganiana* (pH optimum=8.47; range=8.27–8.76), *Staurosira construens* var. *venter* (7.22; 6.42–7.85), and *Staurosirella pinnata* (7.41; 6.37–8.76). Other common taxa displayed relatively wide pH border lengths and a unimodal response (IV, V) to pH, including *Asterionella formosa* (optimum=6.9; range=5.88–9.26),

the *Lindavia bodanica* complex (7.23; 4.15–9.26), and *Tabellaria flocculosa* strain IIIp (6.56; 4.15–8.19) (Fig. 5).

Generally, diatom taxa had broad border ranges for TP concentration (Fig. 3). The taxa indicative of the lowest TP concentrations were *Pantocsekiella tripartita* (TP optimum: $3.38 \mu\text{g L}^{-1}$), the *Aulacoseira perglabra* complex ($5.13 \mu\text{g L}^{-1}$), *Eunotia exigua* ($5.63 \mu\text{g L}^{-1}$), *Asterionella ralfsii* ($5.63 \mu\text{g L}^{-1}$), and *Pantocsekiella ocellata* ($5.77 \mu\text{g L}^{-1}$). The taxa indicative of relatively higher TP concentrations in our lake set were *Nitzschia fonticola* ($14.29 \mu\text{g L}^{-1}$), *Pinnularia interrupta* ($13.94 \mu\text{g L}^{-1}$), *Staurosira construens* var. *venter* ($13.73 \mu\text{g L}^{-1}$), *Eolimna minima* ($13.50 \mu\text{g L}^{-1}$), and *Staurosira construens* ($12.82 \mu\text{g L}^{-1}$) (Fig. 3; Table 3). Common taxa showing a unimodal response with the narrowest borders in the lake set included *Asterionella ralfsii* (TP optimum= $5.63 \mu\text{g L}^{-1}$; range= $1.10\text{--}24.04 \mu\text{g L}^{-1}$) and *Pantocsekiella michiganiana* ($8.01 \mu\text{g L}^{-1}$; $4.08\text{--}21.38 \mu\text{g L}^{-1}$) (Fig. 5). More often, unimodal responses to TP concentration were accompanied with broad border ranges, as observed with *Asterionella formosa* (TP optimum= $7.44 \mu\text{g L}^{-1}$; range= $1.74\text{--}96.8 \mu\text{g L}^{-1}$), the *Lindavia bodanica* complex ($7.67 \mu\text{g L}^{-1}$; $1.10\text{--}96.8 \mu\text{g L}^{-1}$), and *Tabellaria flocculosa* strain IIIp ($7.12 \mu\text{g L}^{-1}$; $1.20\text{--}96.8 \mu\text{g L}^{-1}$; Fig. 5).

The taxa most indicative of shallow water conditions were *Pinnularia interrupta* (depth optimum=1.9 m), *Staurosira construens* var. *venter* (2.5 m), *Psammothidium sacculus* (2.8 m), *Eolimna*

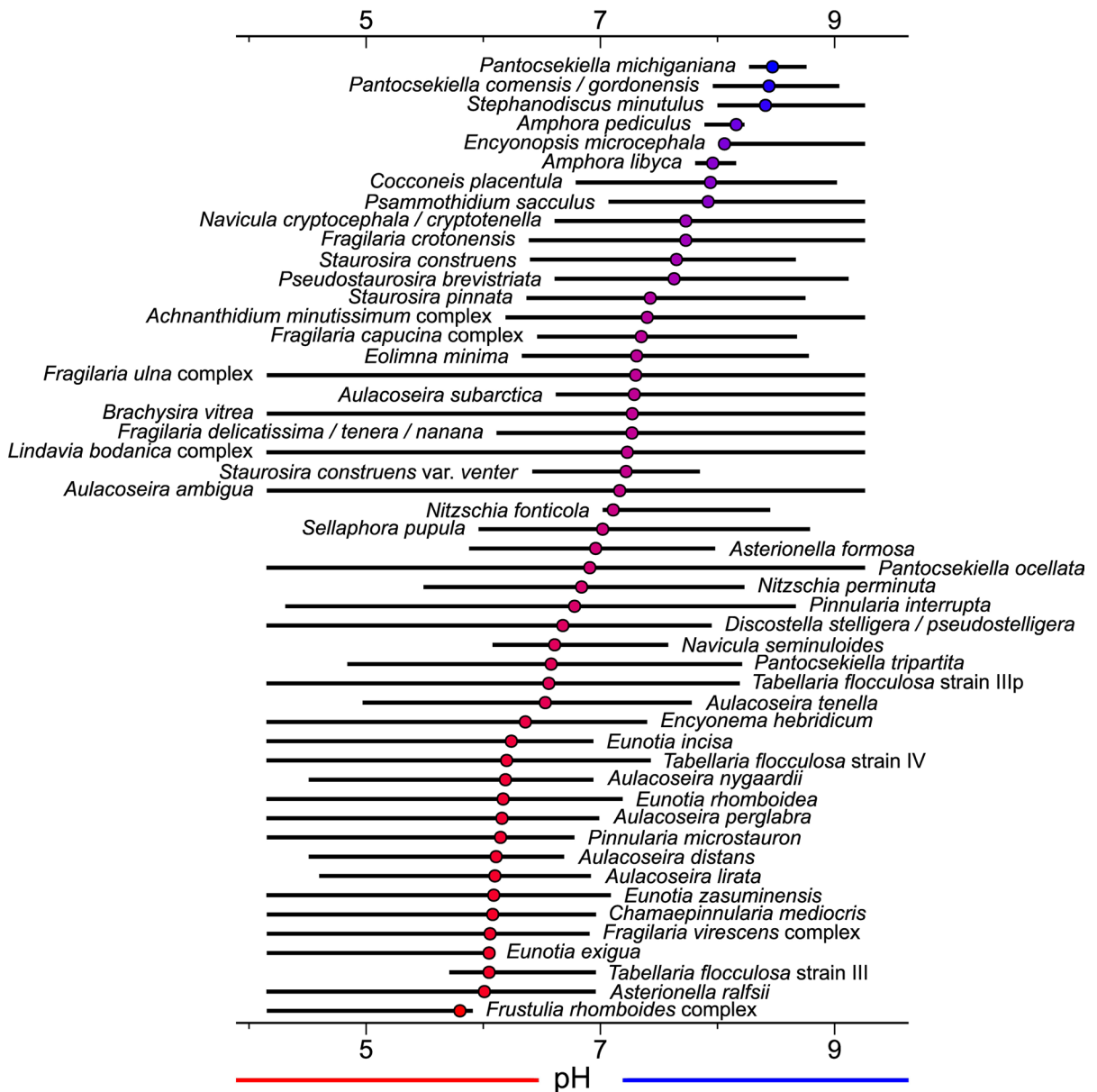


Fig. 2 Caterpillar plot of pH optima, illustrating common taxa with ecologically significant responses (50/52 taxa). The dots represent optima and tails are the outer borders

minima (2.9 m), and *Encyonema hebridicum* (3.0 m). Taxa most indicative of deeper-water conditions were *Stephanodiscus minutulus* (depth optimum: 27.0 m), *Pantocsekiella ocellata* (26.8 m), *Pantocsekiella tripartita* (24.7 m), the *Lindavia bodanica* complex (23.2 m), and *Fragilaria crotonensis* (22.1 m; Fig. 4; Table 3). Common taxa that displayed unimodal responses with narrow borders relative to the dataset

lengths, which suggest they are strong indicators of water depth included *Aulacoseira nygaardii* (depth optimum = 9.2 m; range = 2.2–11.6 m), *Staurosira construens* var. *venter* (2.5 m; 0.8–3.9 m), and *Pantocsekiella tripartita* (24.7 m; 10.0–95.0 m; Fig. 5; Fig. S6). Several taxa displayed unimodal responses with broad depth-border lengths that spanned almost the entire depth gradient of our lake set, including

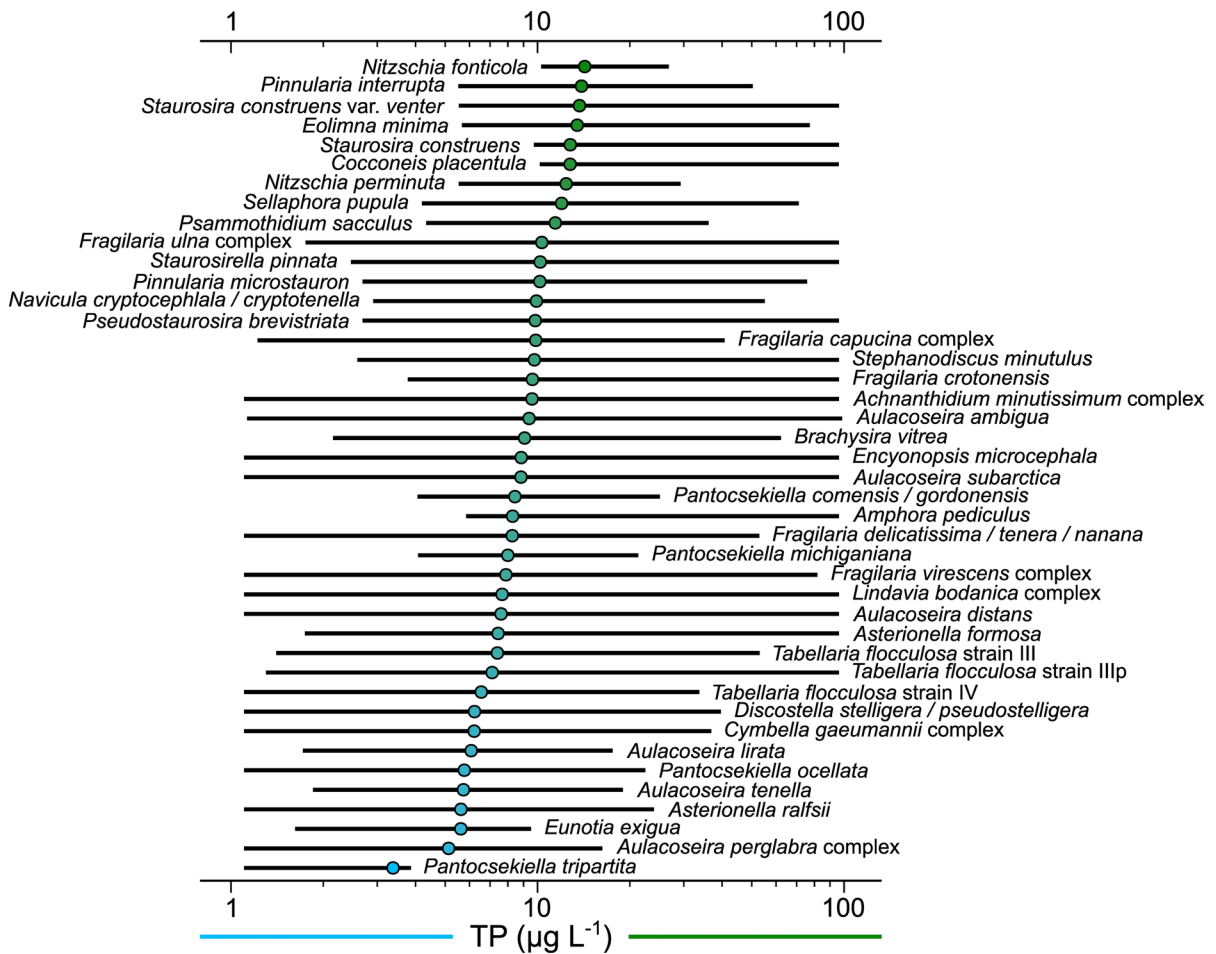


Fig. 3 Caterpillar plot of TP optima, illustrating taxa with ecologically significant responses (43/52 taxa). The dots represent optima and tails are the outer borders. Note that TP is shown on a \log_{10} scale

Aulacoseira ambigua (depth optimum=13.6 m; range=1.2–95.0 m), *Pantocsekiella michiganiana* (13.8 m; 1.0–77.1 m), and *Staurosirella pinnata* (6.18 m; 0.7–63.5 m; Fig. 5).

Discussion

This 546-sample (464 unique sites) dataset, collating diatom information from 20 projects in Ontario (Table 1), incorporates a broad range of samples from remote sub-Arctic sites to lakes in more populated areas in the southernmost part of the province (Fig. 1). The distribution of lakes in the dataset is indicative of the management concerns in Ontario since the 1980s. Nearly 40% of the lakes in our

dataset were selected to examine the effects of acidification caused by smelting operations (Dixit et al. 1989, 1991, 2002; Greenaway et al. 2012), as well as its subsequent recovery (Cheng et al. 2022). Approximately 45% of lakes were selected to describe the effects of eutrophication related to recreation, urbanization, shoreline development, and tourism (Hall and Smol 1996; Reavie and Smol 2001; Werner and Smol 2005; Hadley et al. 2013; Barrow et al. 2014; Wilkins 2021). Around 13% of the lakes were in the sub-Arctic Hudson Bay Lowlands and were originally used to study the recent responses to warming and to retrospectively assess baseline conditions before the onset of major mining operations (Rühland et al. 2014; Hargan et al. 2016; Hargan unpublished). Recent diatom-based studies have aimed to address newly developing

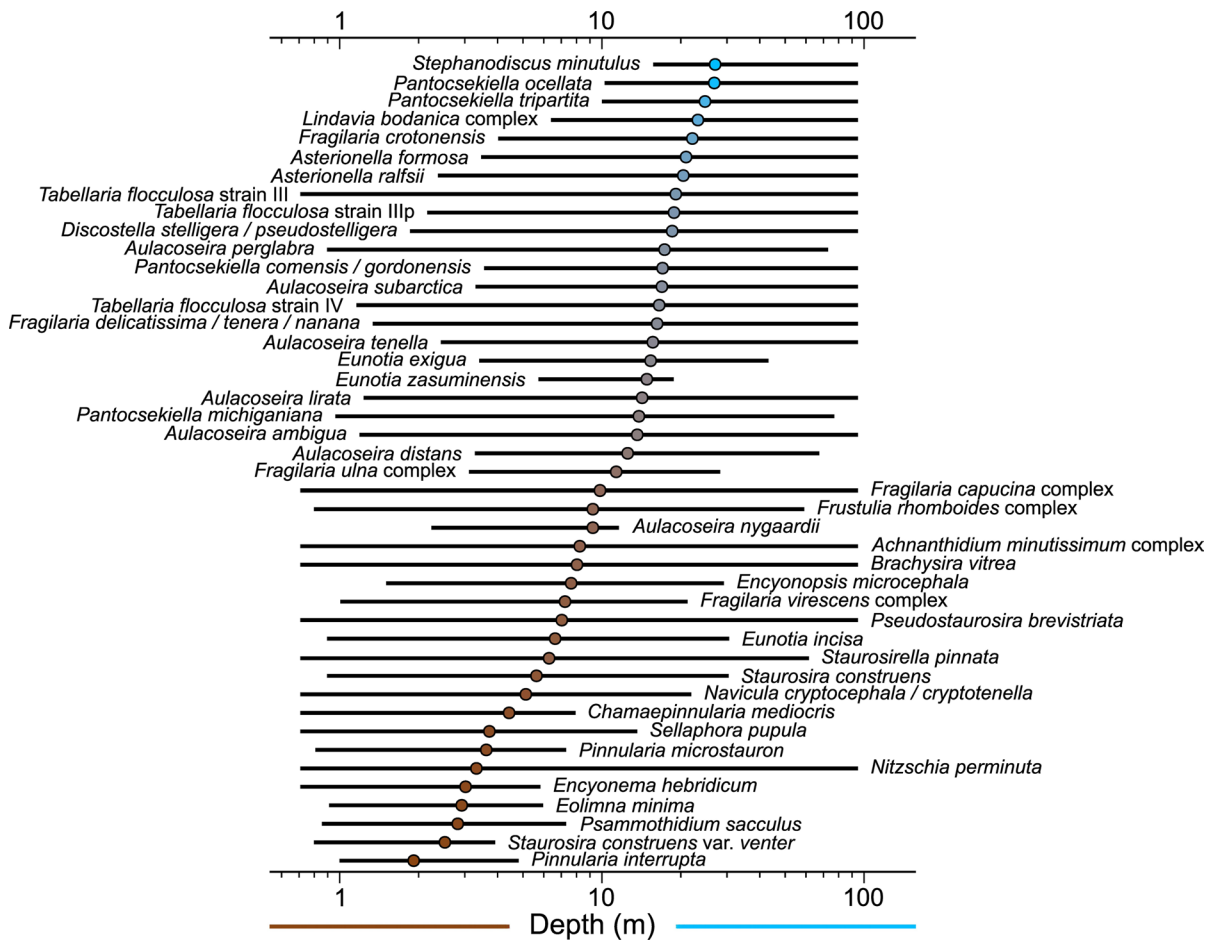


Fig. 4 Caterpillar plot of depth optima, illustrating taxa with ecologically significant responses (43/52 taxa). The dots represent optima and tails are the outer borders. Note that depth is shown on a \log_{10} scale

management concerns using smaller lake sets, such as lake-trout management in response to climate warming (Nelligan et al. 2016, 2020), increasing prevalence of cyanobacterial blooms in minimally impacted lakes (Favot et al. 2019; Favot 2021), and the effects of chloride on diatom assemblages (Valleau 2021; Valleau et al. 2022).

The high representation of study lakes in the Ontario Shield compared to the Hudson Bay Lowlands (Fig. 1; Table 2) is not surprising given the size of this ecozone in relation to the province and the remoteness of the Hudson Bay Lowlands. Given that Ontario spans three terrestrial ecozones that vary in bedrock geology, vegetation, climate, and catchment development, it is understandable that measurements from the 464 sites across this large province have a

broad range in lakewater pH and TP concentration. Differences in pH among the geologically diverse ecozones are largely a reflection of the differences in underlying bedrock and catchment characteristics, as well as the cumulative effects of regional smelting activity in many of the Ontario Shield lakes in the study (Dixit et al. 1991; Cheng et al. 2022). Total phosphorus concentrations are low in Ontario Shield lakes and were generally higher in the Mixedwood Plains and Hudson Bay Lowlands, largely reflecting differences in geological setting. While many of the study lakes from the Mixedwood Plains and Ontario Shield are relatively deep the lakes in the Hudson Bay Lowlands are distinctly shallow, which is typical of this sub-Arctic region. Due to the uneven distribution of sites among the three ecozones (i.e., the majority

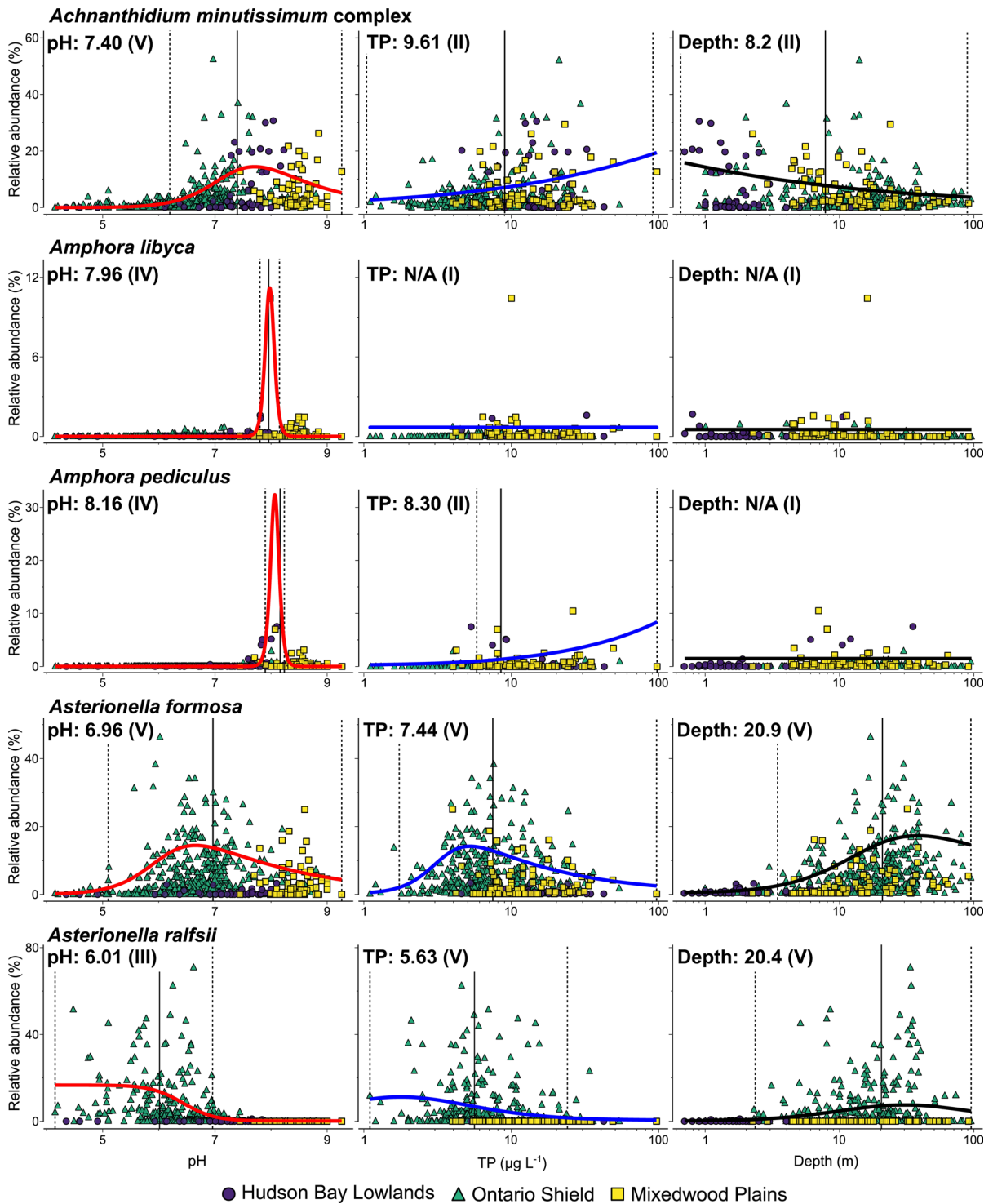


Fig. 5 Ecological response curves of the 52 common taxa in this dataset. Curves reflect the ideal model based on hierarchical logistic regression modelling. The optima and selected models are shown in the top left of each panel. The optimum

is shown with a solid vertical line and the outer borders are dashed vertical lines. Note that TP and depth are shown on a \log_{10} scale

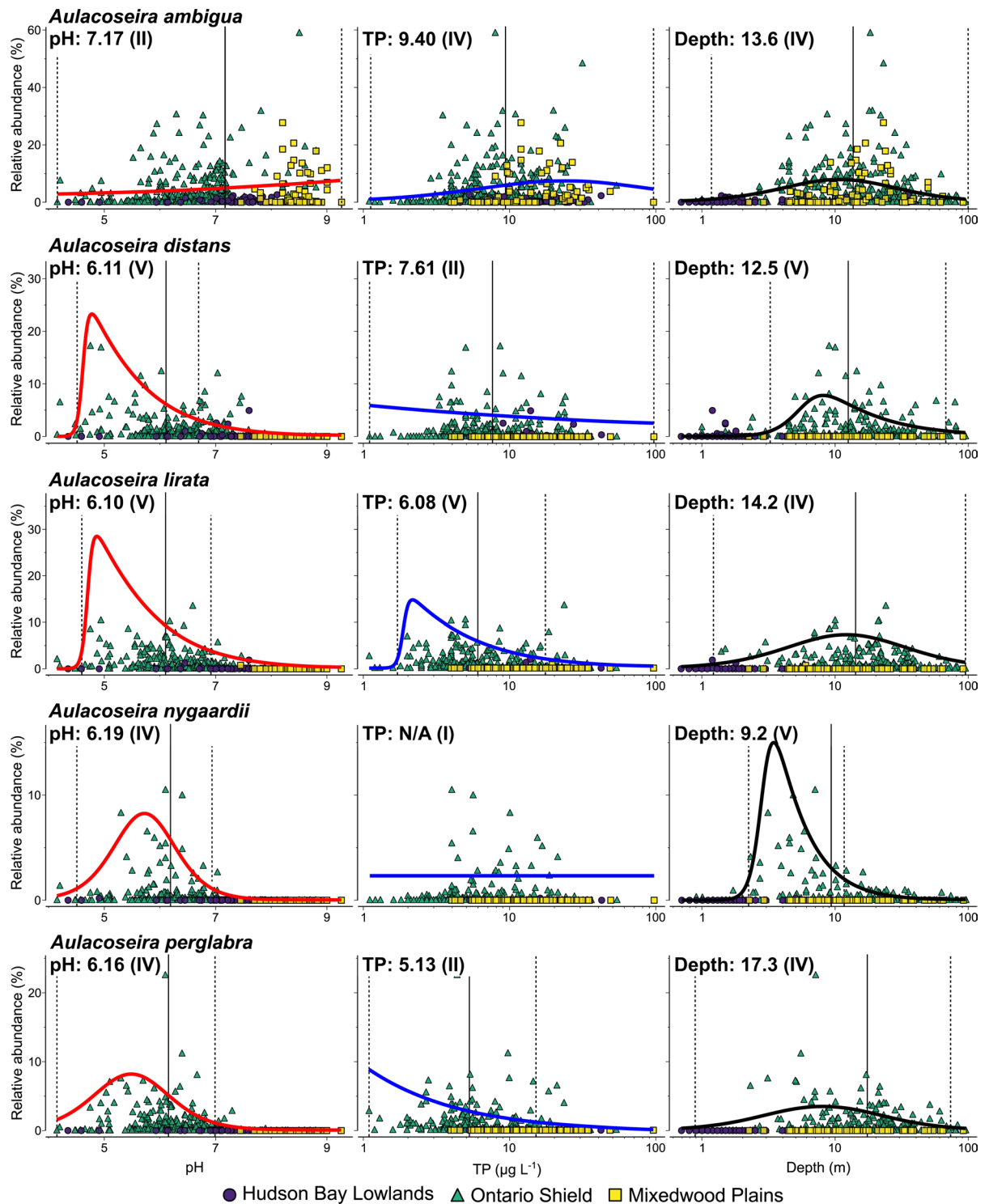


Fig. 5 (continued)

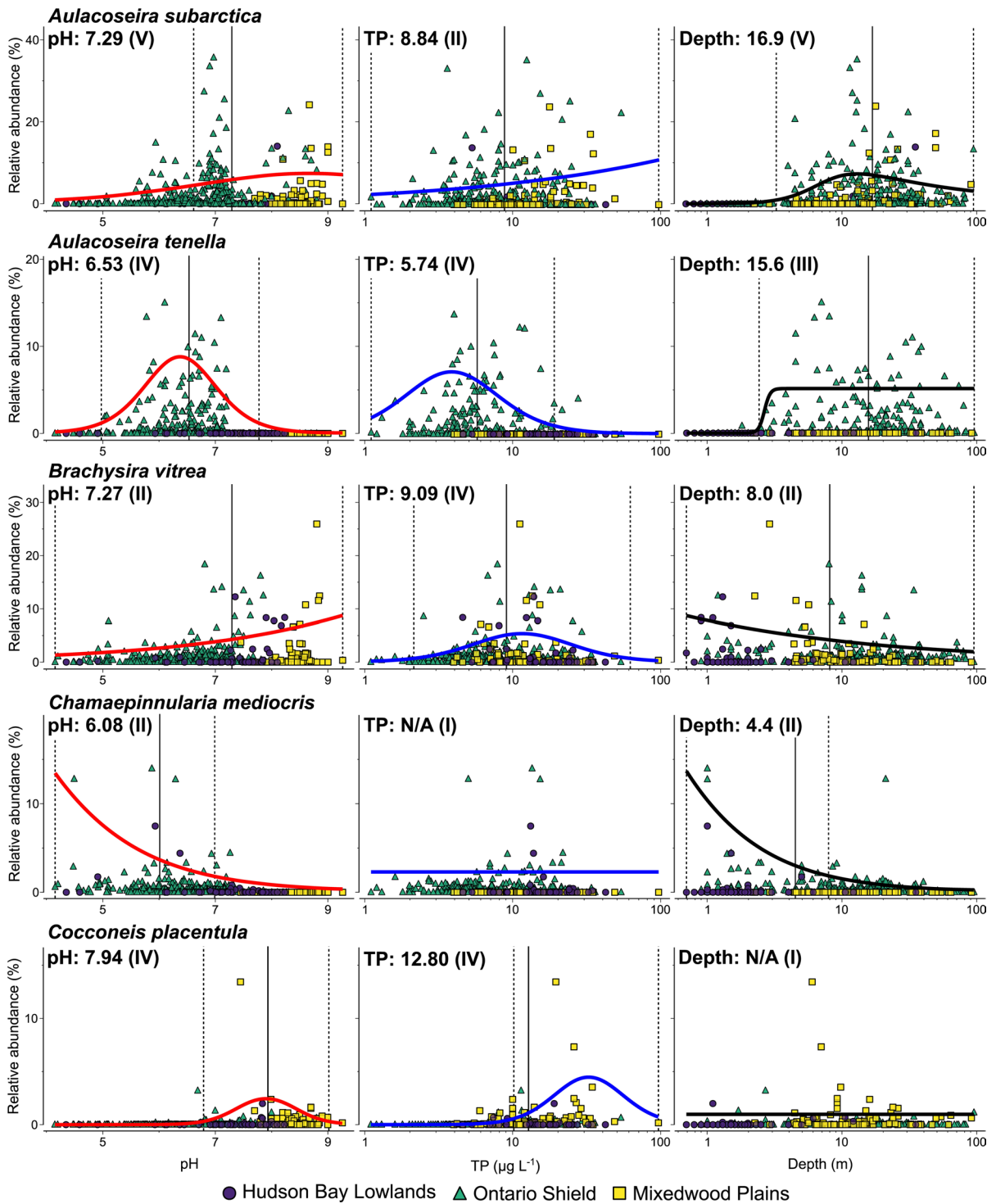


Fig. 5 (continued)

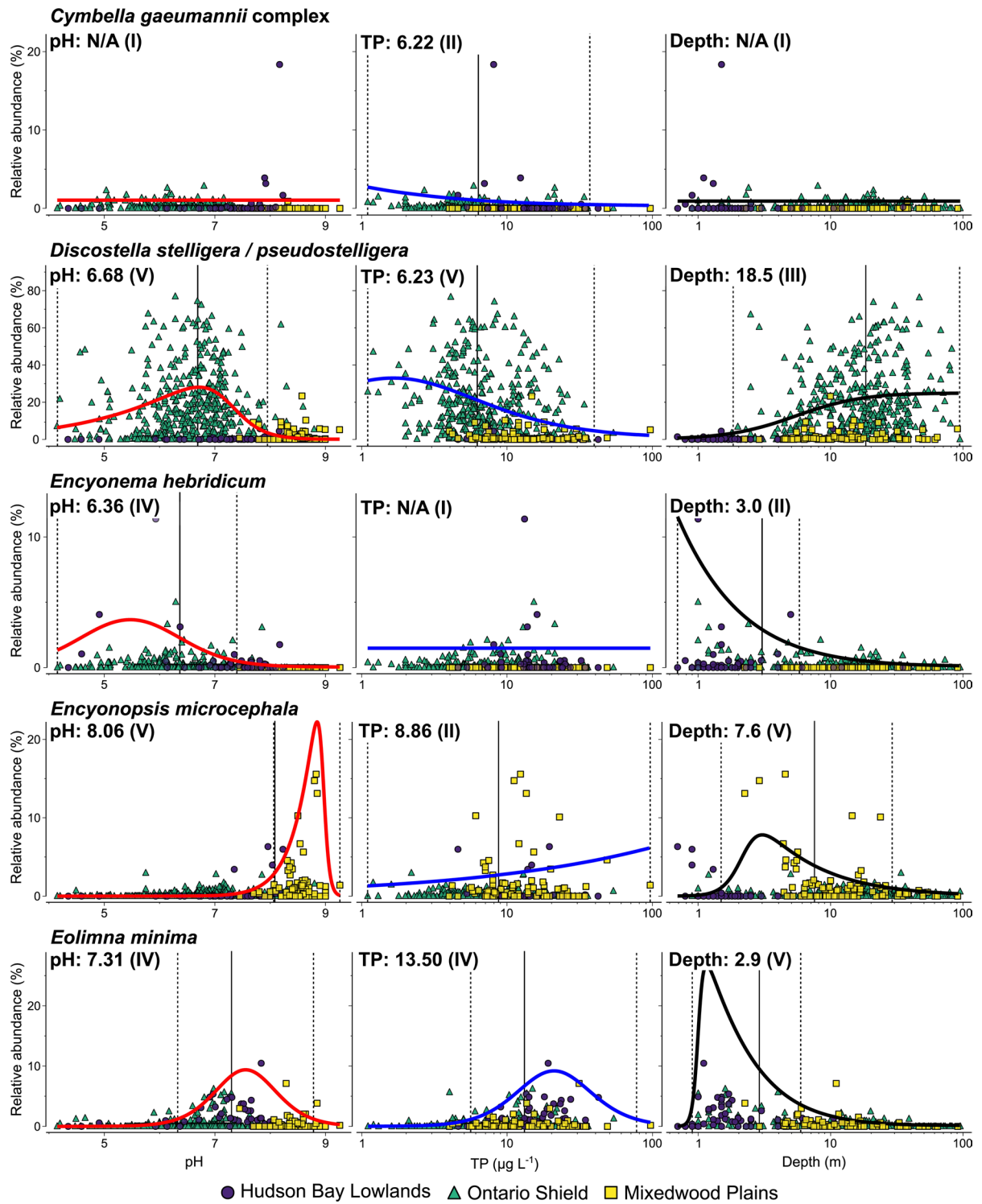


Fig. 5 (continued)

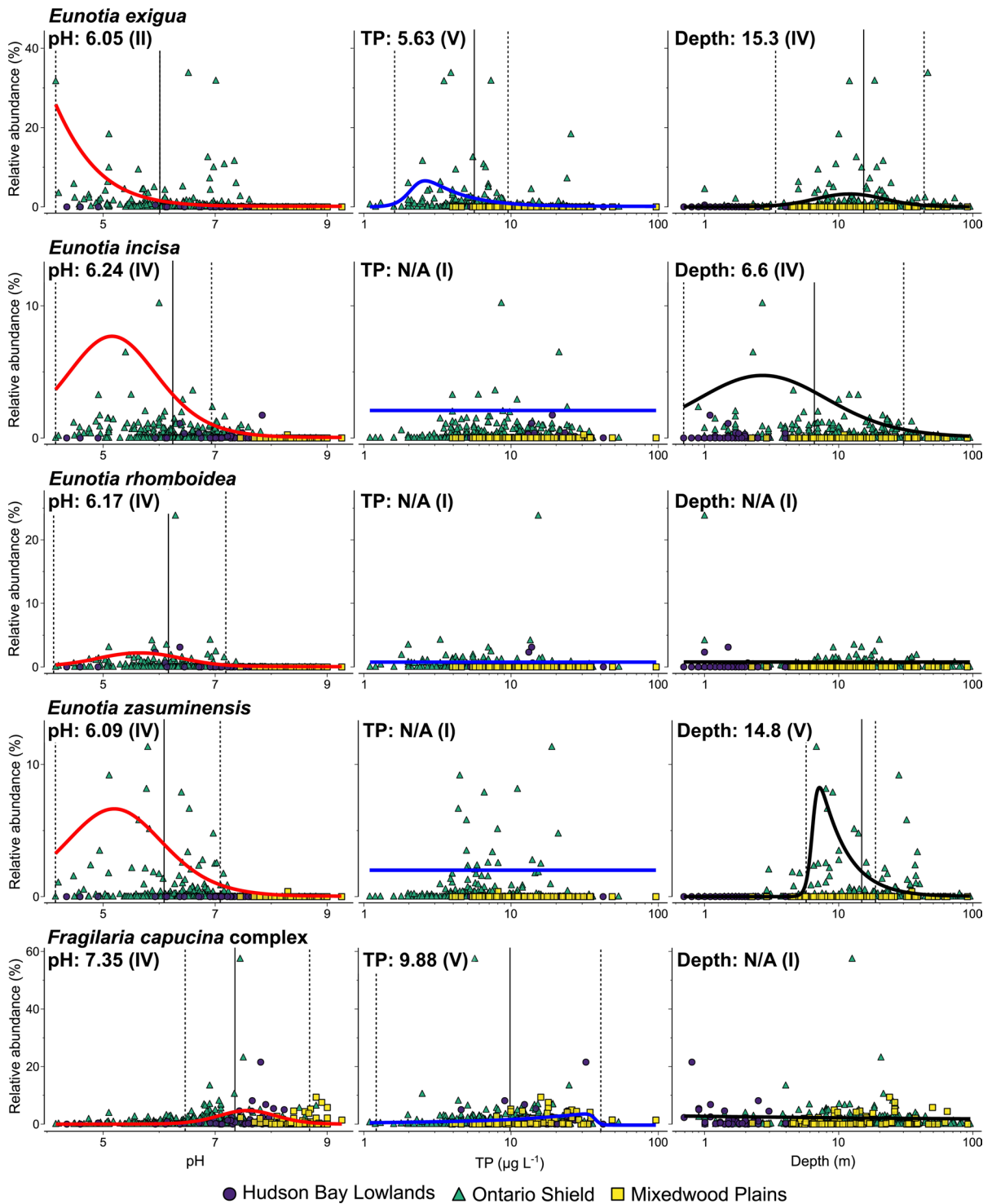


Fig. 5 (continued)

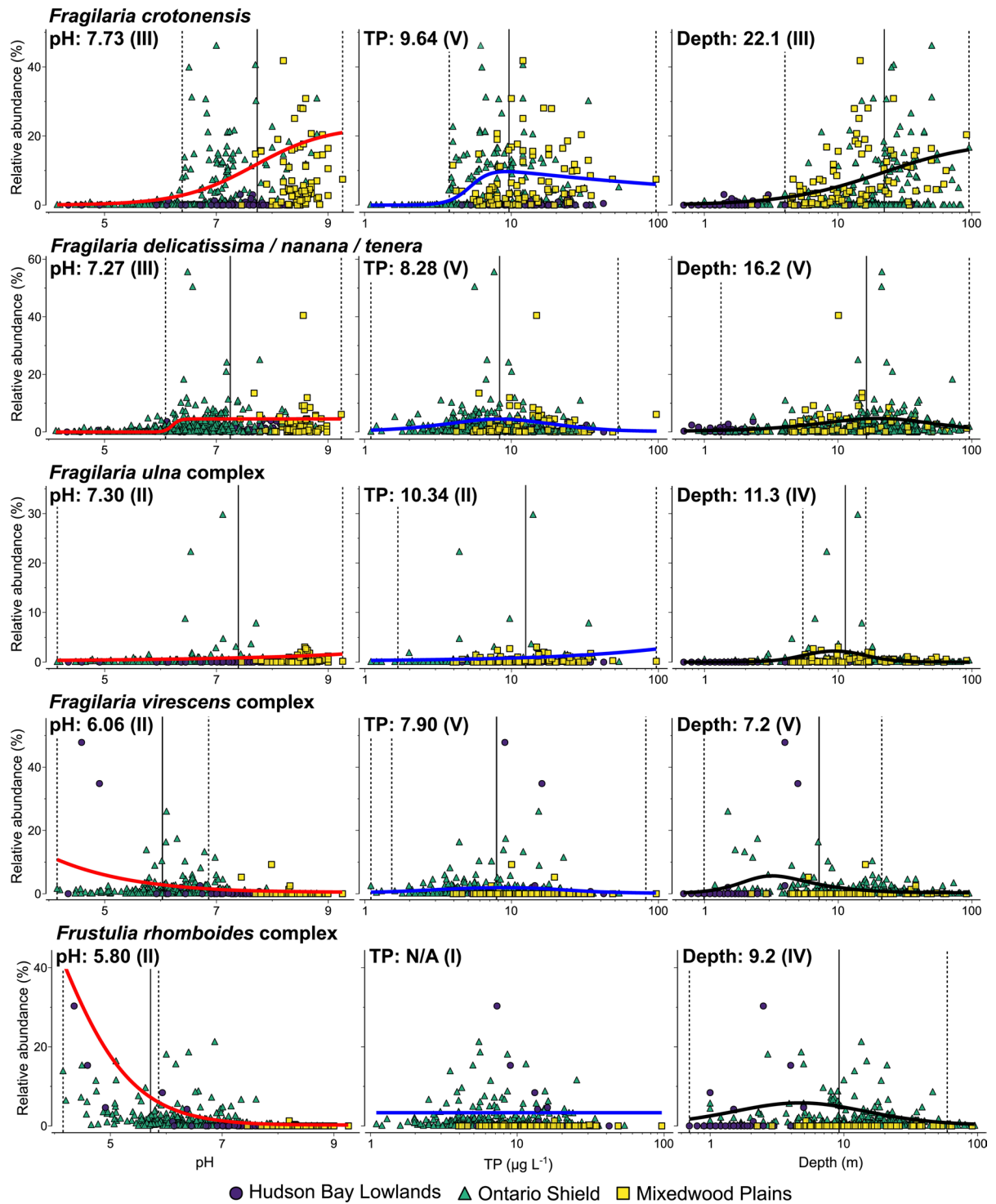


Fig. 5 (continued)

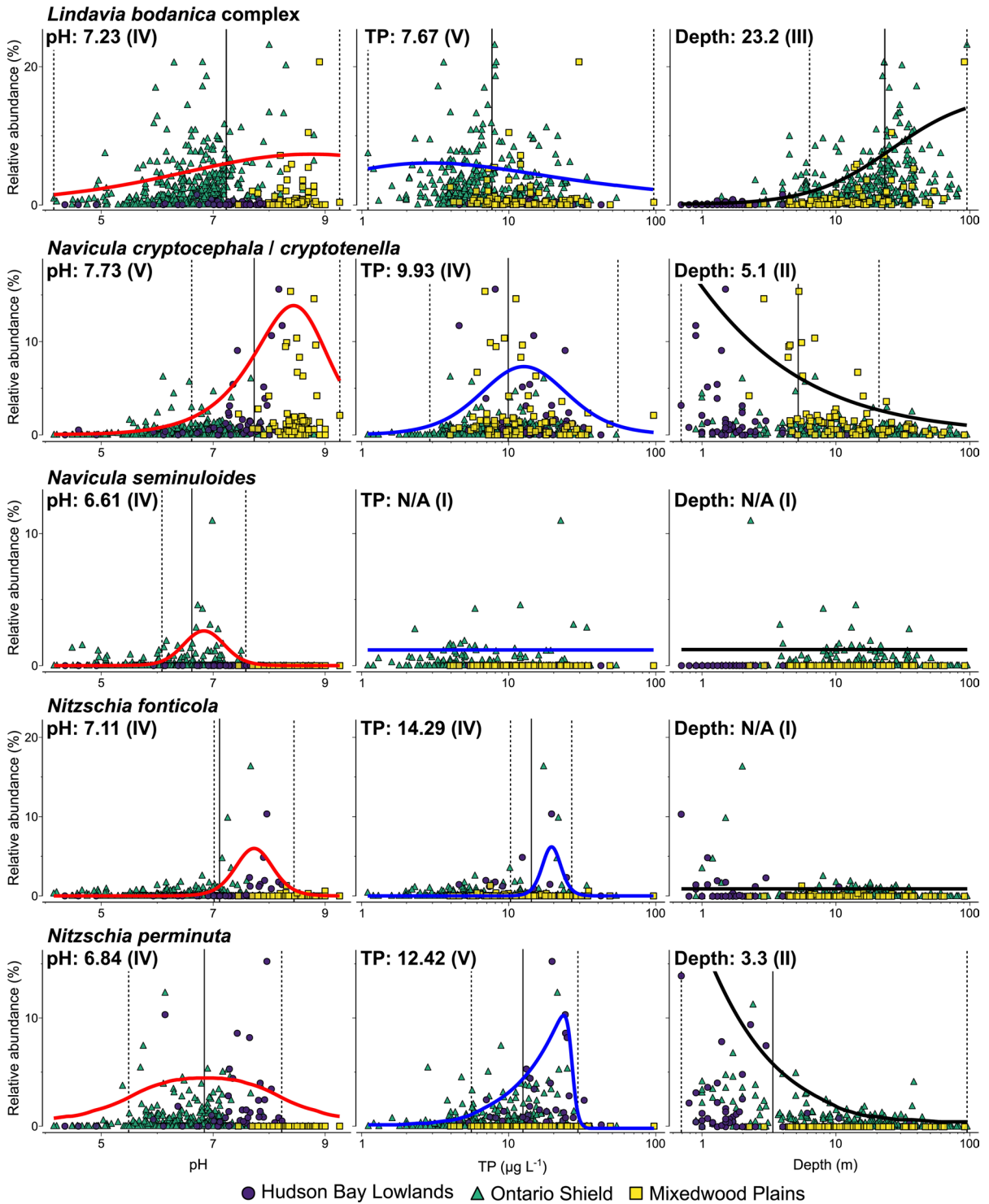


Fig. 5 (continued)

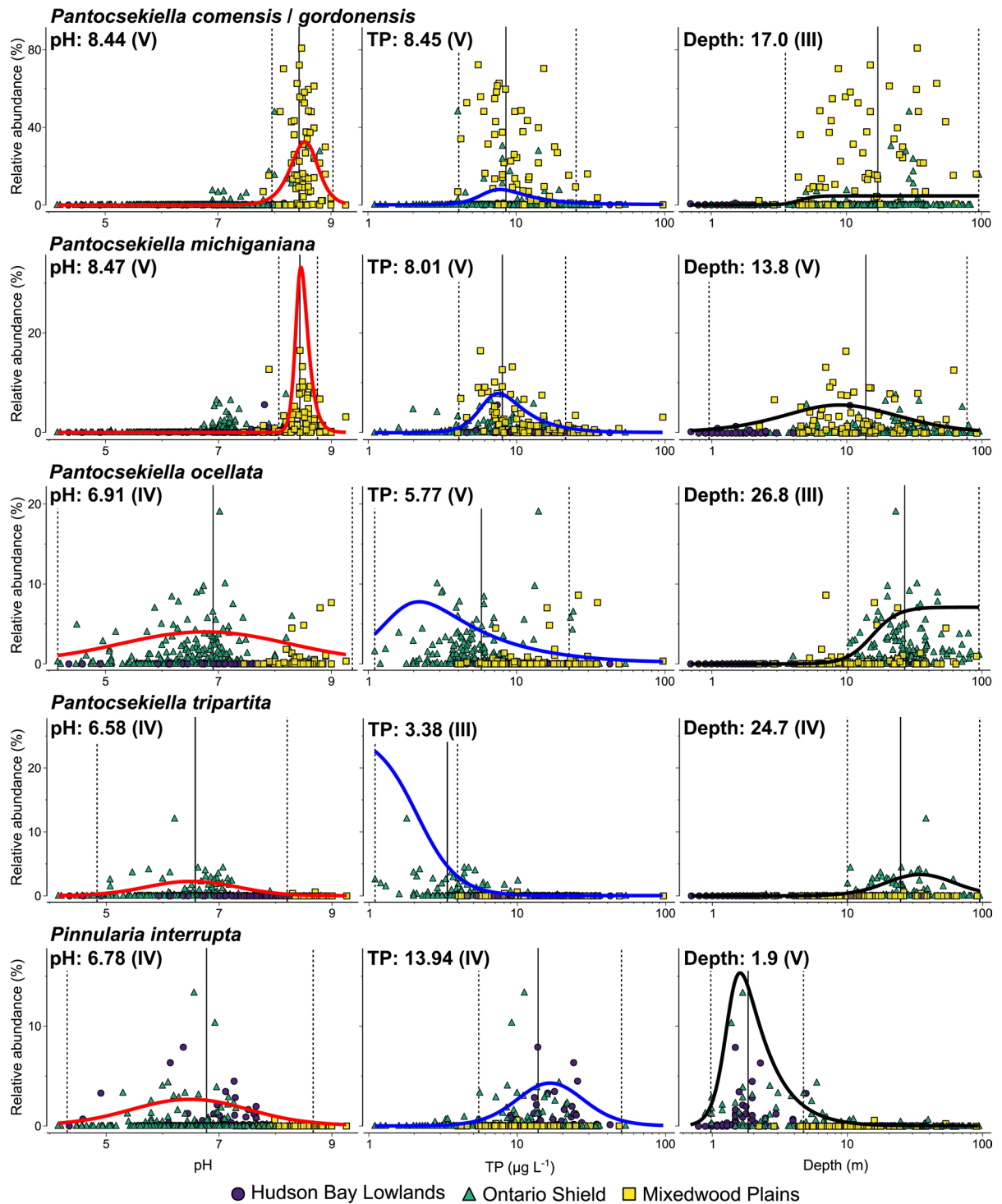


Fig. 5 (continued)

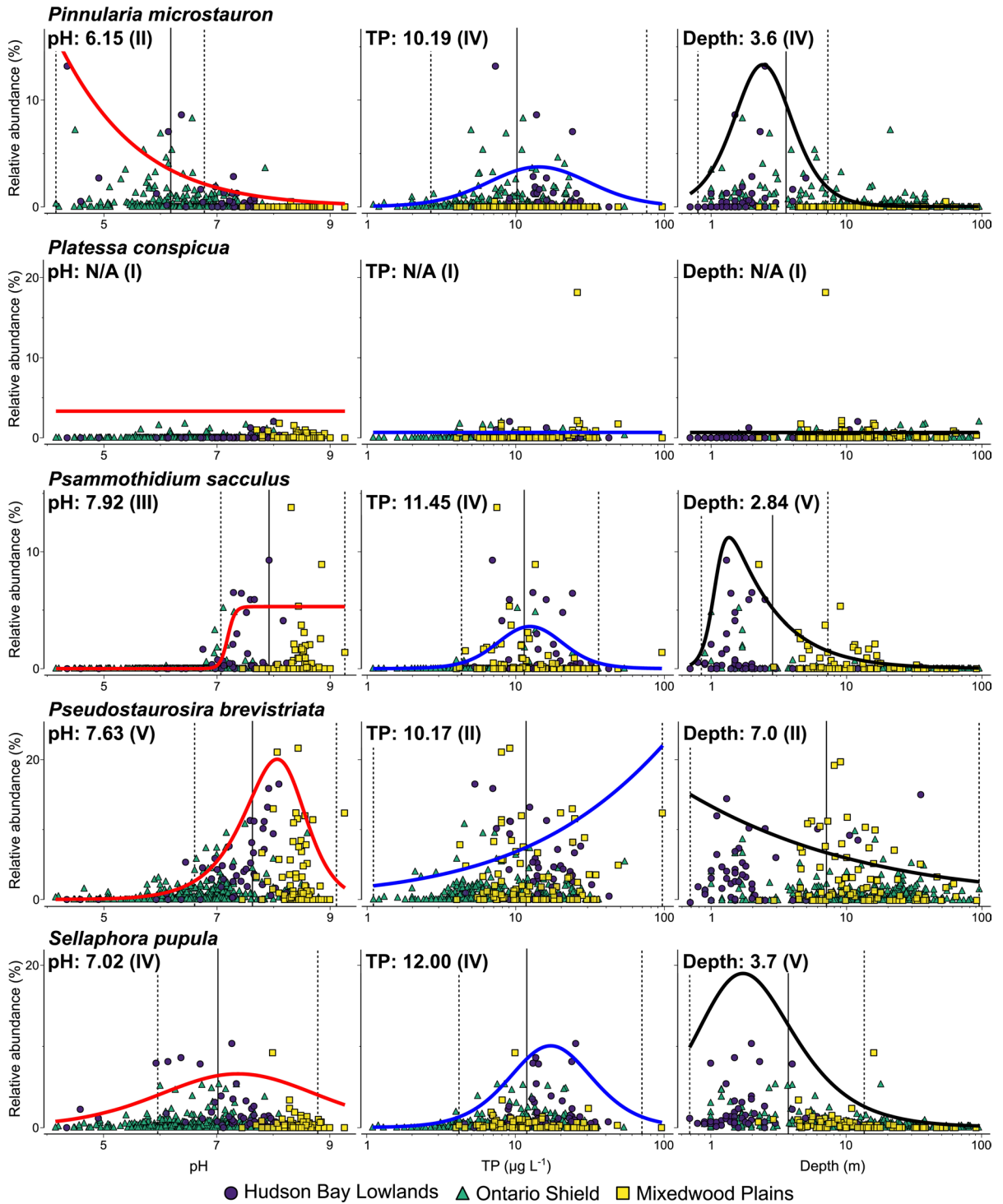


Fig. 5 (continued)

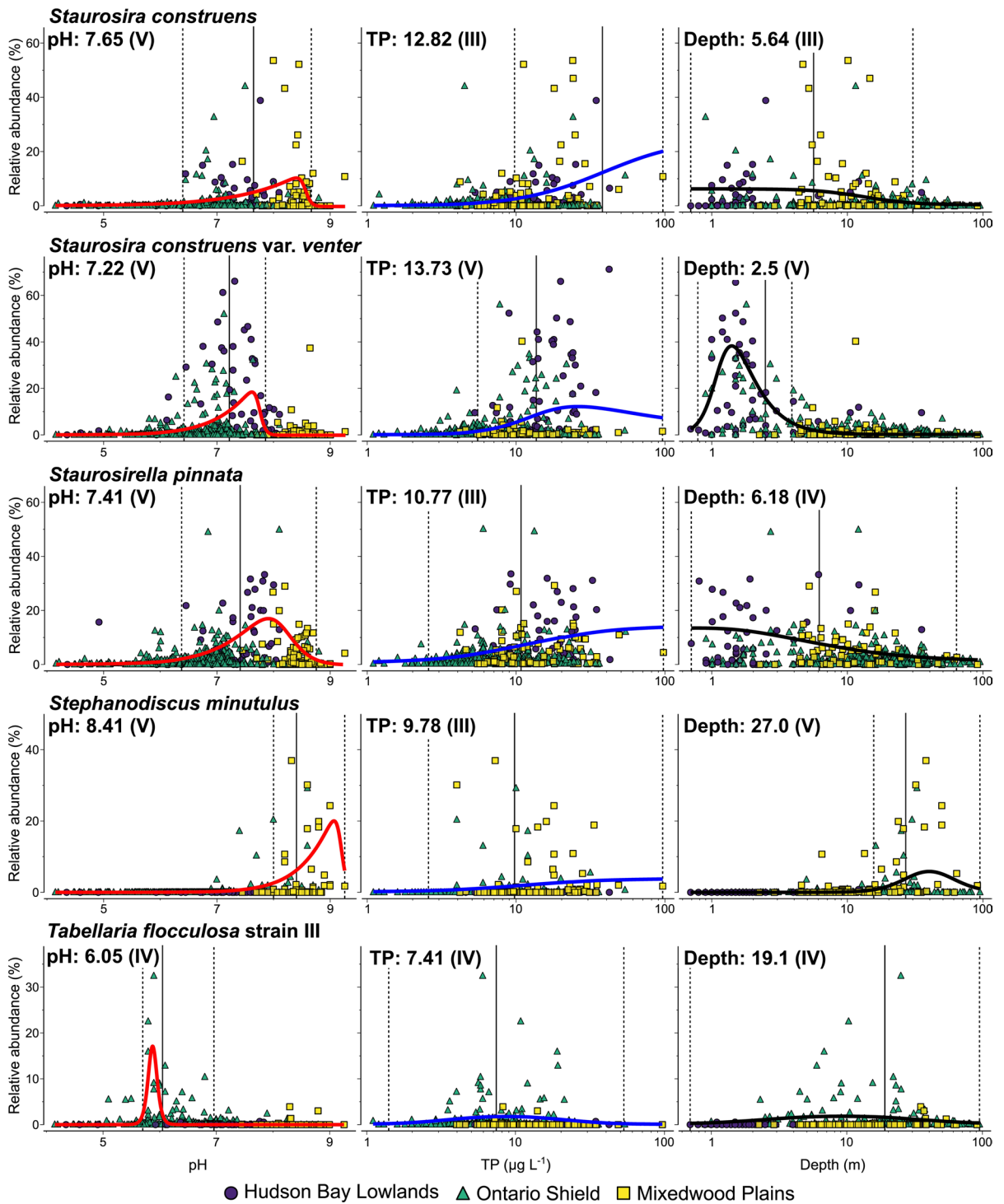


Fig. 5 (continued)

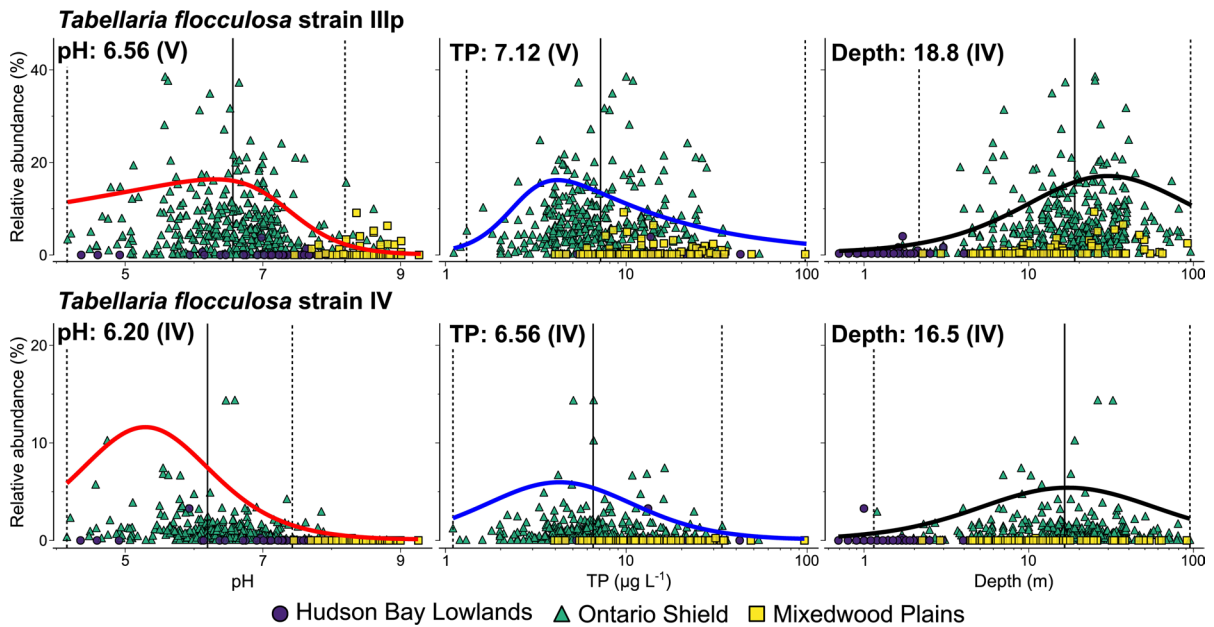


Fig. 5 (continued)

of sites are in the Ontario Shield), optima of diatom taxa were not calculated for each ecozone separately. Many previously published individual calibration sets, however, provide region-specific optima for the most common diatom taxa (Hall and Smol 1996; Reavie and Smol 2001; Werner and Smol 2005; Hadley et al. 2013).

Significant ecological responses to pH were observed for all but two of the most common diatom taxonomic units (the *Cymbella gaeumani* complex; *Platessa conspicua*), and model IV (symmetrical and unimodal) was the most common response type. The strong relationship between diatoms and pH has been observed in several datasets from Canada and around the world and robust pH transfer functions have been developed to track acidification and subsequent pH recovery (Birks et al. 1990; Dixit et al. 2002). Species such as *Asterionella ralfsii*, the *Frustulia rhomboides* complex and *Eunotia exigua* recorded the lowest pH optima in our dataset and have been identified as key indicators of lake acidification in previous investigations (Camburn and Charles 2000; Dixit et al. 2002). Planktonic taxa such as *A. ralfsii*, *Tabellaria flocculosa* strain III, *Eunotia zasuminensis*, and several filamentous *Aulocoseira* taxa (*A. distans*, *A. lirata*, *A. perglabra*) have low pH optima and were more common in deeper lakes (> 10 m; Table 3). Meanwhile,

benthic taxa such as the *Frustulia ormosas* complex, *Eunotia incisa*, and the *Fragilaria virescens* complex also have relatively low pH optima, but had lower depth optima (< 10 m; Table 3). Although acidobiontic *Eunotia exigua* is a benthic taxon, it had a depth optimum of 15.3 m. This benthic taxon may thrive in the littoral zones of deep, clear, acidic lakes, highlighting that maximum lake depth is a coarse measure that cannot account for the complexities in lake morphology or the light environment, which can influence the relative availability of littoral habitat.

Significant ecological responses to TP concentrations were observed for 43 of the 52 most common taxonomic units in this dataset. A strong understanding of how diatom assemblages respond to changes in lake-water-TP concentrations and the development of transfer functions have helped recognize eutrophication trajectories in lakes (Ramstack et al. 2004; Cumming et al. 2015). In our Ontario dataset, the majority of lakes were considered oligotrophic to mesotrophic in TP concentrations, and therefore “low” and “high” TP optima are described within this context. Generally, epiphytic and benthic taxa (*Cocconeis placentula*, *Eolimna minima*, *Nitzschia fonticola*, *Nitzschia perminuta*, *Pinnularia interrupta*, *Sellaphora pupula*, *Staurosira construens*, and *Staurosira construens* var. *venter*) had the highest TP optima in our Ontario lakes

(TP optima range = 12.00–14.29 $\mu\text{g L}^{-1}$). Relatively high optima for TP concentrations were also recorded for some planktonic taxa, such as *Fragilaria crotonensis* (TP optimum = 9.64 $\mu\text{g L}^{-1}$), the *Fragilaria ulna* complex (TP optimum = 10.34 $\mu\text{g L}^{-1}$), and *Stephanodiscus minutulus* (TP optimum = 9.78 $\mu\text{g L}^{-1}$).

Although our study lakes are distributed across large environmental gradients (Fig. 1; Table 2), the majority are circumneutral (Fig. S1), oligotrophic (Fig. S2), and reach a maximum depth of between 10 and 25 m (Fig. S4). Therefore, the weighted-average (WA) optima described in this study may suffer from the ‘edge effect’ for taxa indicative of eutrophic or deep conditions, where optima are poorly estimated due to the truncation of their ecological response curves at the extremes of the gradient (Simpson and Hall 2012). Given that the majority of the lakes in our study are skewed towards low TP concentration and relatively shallow maximum depth, WA optima for taxa that more commonly thrive in eutrophic or very deep conditions may be underestimated. Another factor that may affect our WA optima is high seasonal variability in surface-water-TP concentration. For example, in oligotrophic Precambrian Shield lakes, TP concentrations may decline in the summer months when many lakes in the study were sampled (Clark et al. 2010). The taxa most likely affected by edge effects are represented by Model II responses, where the taxa optima are outside of the measured range in our dataset. These optima should be viewed with an understanding of the distribution, and the knowledge that there may not be sufficient data to accurately estimate the optima.

Notably, taxa that are often considered strong indicators of eutrophication were not common in our dataset (i.e., were not included in the group of 52 common taxa). For example, indicators of eutrophic waters, such as *Stephanodiscus hantzschii* and *Cyclostephanos invisitatus*, were only observed in 16 lakes (~3% of the lake set) and 9 lakes (~2%), respectively, and therefore did not meet the criteria required for undertaking detailed analysis. Although the TP concentration gradient of the Ontario dataset is large (1.1–96.8 $\mu\text{g L}^{-1}$; Table 2), the majority of lakes are oligotrophic (60% of sampled sites have TP < 10 $\mu\text{g L}^{-1}$) and therefore taxa that indicate highly eutrophic conditions (TP > 24 $\mu\text{g L}^{-1}$) were not commonly observed (Fig. 1).

In general, the most common diatoms had broad ecological responses to lakewater pH, TP concentration, and maximum depth (Fig. 5). This is not surprising given that these are three commonly and easily measured variables that encompass other limnological factors important for diatom growth, thereby potentially confounding measurements of diatom distributions. For example, TP concentration and depth are associated with light penetration, and TP levels and pH may be associated with dissolved organic carbon (DOC) concentrations. Dissolved organic carbon is directly linked to the depth of light penetration, which can be a key variable driving variation in diatom communities (Gushulak et al. 2017). Other factors that diatoms respond to and that were not measured in our study include the length of the growing season, thermal stratification and turbulent mixing patterns, which are increasingly important with accelerated climate warming (Rühland et al. 2015). For instance, the relative abundances of two of the most common planktonic diatom taxa in the dataset (*D. stelligera/pseudostelligera* and *A. formosa*) have been associated with climate-mediated changes to lake thermal properties in many Ontario lakes (Enache et al. 2011; Rühland et al. 2013; Hadley et al. 2013; Barrow et al. 2014; Sivarajah et al. 2016, 2018).

In the 546-sample dataset, the most common distribution of ecological response curves was Model IV (~31%), followed by Model V (~30%), Model II (~16%), Model I (~14%) and Model III (~9%). This distribution generally agrees with earlier work, which noted that symmetrical response curves are the most common, followed by monotonic responses, skewed and null responses, and finally plateau responses (Oksanen and Minchin 2002; reviewed in Birks et al. 2012). A key difference in our study is the larger number of Model I (null) responses and Model V (skewed, unimodal) responses. We propose two potential reasons for this discrepancy. First, the loss of some taxonomic resolution during the necessary harmonization step could have led to null or skewed response curves (e.g., the *Fragilaria capucina* complex; Fig. S6). Our conservative harmonization approach was required because the data were collected over three decades by multiple researchers. Second, high variability in taxa relative abundance may have resulted in null response curves as a strong response curve could not be fit to the data. This occurred when taxa were found in low relative abundance across the majority of lakes

and were rarely observed at high abundance (e.g., *Amphora libyca*, *Cocconeis placentula*; Fig. S6), or when taxa were found in high abundance but in relatively few lakes (e.g., *Eunotia zasuminensis*, *Nitzschia fonticola*; Fig. S6). These issues can be addressed by a pre-established study design assessing lake conditions, such as the ongoing Natural Sciences and Engineering Research Council of Canada (NSERC) Lake Pulse network in which even proportions of lakes were selected that considered trophic status, human stressors, and other factors (Huot et al. 2019).

The development of large biological datasets can help to explore a variety of questions and serve as a starting point for future research. In this study, we quantified environmental optima for the most common diatoms reported in Ontario lakes. Future studies could explore how environmental optima for individual species differ between each geographic region by comparing the results from this study to previously published work from Ontario. Similarly, this large dataset could be used to develop an Ontario-wide diatom-based transfer function to reconstruct past changes in TP concentrations and pH. The strength and performance of the new model could be assessed against existing models and measured data from monitoring programs to determine if a transfer function with more than 500 sites can reconstruct environmental variables more accurately. The availability of a large dataset also enables the use of innovative techniques that require large datasets to reconstruct limnological variables. For example, the moving window approach selects a subset of assemblages (40, 60 ... 200) from the large calibration dataset (400+ sites) that are similar to the fossil sample to reconstruct environmental variables (Hübener et al. 2008, 2009). The selection of a subset of diatom taxa helps to increase the likelihood of finding good modern analogues for the fossil assemblages, which can greatly improve the reconstruction of environmental variables.

Conclusions

By combining many surface sediment samples collected from Ontario lakes that were analyzed in the same lab using a similar methodology, we identified the overall patterns of diatom species response curves to three key limnological variables. Such analyses

can help refine diatom autecology and advance the understanding of how common taxa respond to environmental variables that are important for lake management. By modelling ecological response curves for Ontario's 52 commonly encountered taxa, we contribute to the application of diatoms as bioindicators for environmental assessment. The information from this dataset can also be useful for inferring past limnological conditions (particularly TP and pH) and/or biological recovery trajectories at sites that have been impacted by multiple environmental stressors. These data may be particularly valuable in understudied regions with similar limnological environments and diatom communities where calibration sets are not available. Moreover, lakes can undergo dramatic changes through time, thereby requiring large lake sets spanning large environmental gradients to increase the likelihood of finding modern analogues for downcore-diatom assemblages. The taxon-specific optima for lakewater pH and TP concentrations can help to qualitatively assess past trends in these key limnological variables, which can serve as important management and biomonitoring tools.

Acknowledgements We are grateful to Mihaela Enache, Christine Greenaway, and Sushil Dixit who contributed data to this synthesis but kindly declined to co-author. This research was funded by the Natural Sciences and Engineering Research Council of Canada (PGSD3-546925-2020) awarded to MPD and an NSERC Discovery Grant awarded to JPS.

Author contributions MPD and BS compiled the data, completed statistical analyses, and prepared figures. MPD, BS, KMR, AMP, and JPS wrote the original manuscript. KMR, JLB, YC, EJF, KRH, RIH, KEH, CN, EDR, REV, PW, and CW contributed significant data for the model. All authors contributed text and approved the final manuscript.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Ács E, Ari E, Duleba M, Dreßler M, Genkal SI, Jakó E, Rimet F, Ector L, Kiss KT (2016) *Pantocsekiella*, a new centric diatom genus based on morphological and genetic studies. *Fottea* 16:56–78
- Alexsson EE, Reavie ED, Van de Vijver B, Wetzel CE, Ector L, Wellard Kelly HA, Aliff NA, Estep LR (2022) Revision of the needle-shaped *Fragilaria* species (Fragilariaceae,

- Bacillariophyta) in the Laurentian Great Lakes (United States of America, Canada). *J Great Lakes Res* 48:999–1020
- Anderson DM, Glibert PM, Burkholder JM (2002) Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. *Estuaries* 25:704–726
- Barrow JL, Jezierski A, Rühland KM, Hadley KR, Smol JP (2014) Diatoms indicate that calcium decline, not acidification, explains recent cladoceran assemblage changes in south-central Ontario softwater lakes. *J Paleolimnol* 52:61–75
- Battarbee RW (1990) The causes of lake acidification, with special reference to the role of acid deposition. *Phil Trans R Soc B* 327:339–347
- Battarbee RW, Jones VJ, Flower RJ, Cameron NG, Bennion H, Carvalho L, Juggins S (2001a) Diatoms. In: Smol JP, Birks HJB, Last WM (eds) *Tracking environmental change using lake sediments*, vol 3. Terrestrial, Algal, and Siliceous Indicators. Kluwer Academic Publishers, Dordrecht, pp 155–202
- Battarbee RW, Charles DF, Bigler C, Cumming BF, Renberg I (2010) Diatoms as indicators of surface-water acidity. In: Smol JP, Stoermer EF (eds) *The diatoms: applications for the environmental and earth sciences*, 2nd edn. Cambridge University Press, Cambridge, pp 98–121
- Battarbee RW, Juggins S, Gasse F, Anderson NJ, Bennion H, Cameron NG, Ryves DB, Pailles C, Chalié, F, Telford R (2001b) European Diatom Database (EDDI). An information system for palaeoenvironmental reconstruction. (ECRC Research Report No. 81). University College London Environmental Change Research Centre, London
- Birks HJB, Line JM, Juggins S, Stevenson AC, ter Braak CJF (1990) Diatoms and pH reconstruction. *Phil Trans R Soc B* 327:263–278
- Birks HJB, Juggins S, Lotter A, Smol JP (2012) *Tracking environmental change using lake sediments*. Vol 5. Data handling and numerical techniques. Springer, Dordrecht, p 745
- Camburn KE, Charles DF (2000) Diatoms of low-alkalinity lakes in the northeastern United States. *Academy of Natural Sciences of Philadelphia, Special Publication* 18, p 152
- Chen G, Dalton C, Leira M, Taylor D (2008) Diatom-based total phosphorus (TP) and pH transfer functions for the Irish Ecoregion. *J Paleolimnol* 40:143–163
- Cheng Y, Michelutti N, Paterson AM, Meyer-Jacob C, Smol JP (2022) Developing diatom-based inference models to assess lake ecosystem change along a gradient of metal smelting impacts: Sudbury lakes revisited. *J Phycol* 58:530–542
- Clark BJ, Paterson AM, Jezierski A, Kelsey S (2010) Assessing variability in total phosphorus measurements in Ontario lakes. *Lake Reserv Manag* 26:63–72
- Cumming BF, Smol JP, Kingston JC, Charles DF, Birks HJB, Camburn KE, Dixit SS, Uutala AJ, Selle AR (1992) How much acidification has occurred in Adirondack Region Lakes (New York, USA) since preindustrial times? *Can J Fish Aquat Sci* 49:128–141
- Cumming BF, Laird KR, Gregory-Eaves I, Simpson KG, Sokal MA, Nordin RN, Walker IR (2015) Tracking past changes in lake-water phosphorus with a 251-lake calibration dataset in British Columbia: tool development and application in a multiproxy assessment of eutrophication and recovery in Osoyoos Lake, a transboundary lake in Western North America. *Front Ecol Evol* 3:84
- Dixit SS, Dixit AS, Smol JP (1989) Relationship between chrysophyte assemblages and environmental variables in seventy-two Sudbury lakes as examined by canonical correspondence analysis (CCA). *Can J Fish Aquat Sci* 46:1667–1676
- Dixit SS, Dixit AS, Smol JP (1991) Multivariable environmental inferences based on diatom assemblages from Sudbury (Canada) lakes. *Freshw Biol* 26:251–266
- Dixit SS, Smol JP, Kingston JC, Charles DF (1992) Diatoms: powerful indicators of environmental change. *Environ Sci Technol* 26:22–33
- Dixit SS, Dixit AS, Smol JP (2002) Diatom and chrysophyte functions and inferences of post-industrial acidification and recent recovery trends in Killarney lakes (Ontario, Canada). *J Paleolimnol* 27:79–96
- Enache MD, Paterson AM, Cumming BF (2011) Changes in diatom assemblages since pre-industrial times in 40 reference lakes from the Experimental Lakes Area (northwestern Ontario, Canada). *J Paleolimnol* 46:1–15
- Environment Canada (1994a) *Manual of analytical methods. Volume 1: Major ions and nutrients*. National Laboratory for Environmental Testing, Canadian Centre for Inland Waters, Burlington, ON, Canada
- Environment Canada (1994b) *Manual of analytical methods. Volume 2: Trace metals*. National Laboratory for Environmental Testing, Canadian Centre for Inland Waters, Burlington, ON, Canada
- Favot EJ, Rühland KM, DeSellas AM, Ingram R, Paterson AM, Smol JP (2019) Climate variability promotes unprecedented cyanobacterial blooms in a remote, oligotrophic Ontario lake: evidence from paleolimnology. *J Paleolimnol* 62:31–52
- Favot EJ (2021) *Investigating long-term environmental trends in central Ontario lakes impacted by cyanobacterial blooms*. Doctoral dissertation, Queen's University.
- Fritz SC, Juggins S, Battarbee RW (1993) Diatom assemblages and ionic characterization of lakes of the Northern Great Plains, North America: a tool for reconstructing past salinity and climate fluctuations. *Can J Fish Aquat Sci* 50:1844–1856
- Glew JR (1988) A portable extruding device for close interval sectioning of unconsolidated sediment samples. *J Paleolimnol* 1:235–239
- Glew JR (1989) A new trigger mechanism for sediment subsampler. *J Paleolimnol* 2:241–243
- Glew JR (1991) Miniature gravity corer for recovering short sediment cores. *J Paleolimnol* 5:285–287
- Greenaway CM, Paterson AM, Keller W, Smol JP (2012) Dramatic diatom species assemblage responses in lakes recovering from acidification and metal contamination near Wawa, Ontario, Canada: a paleolimnological perspective. *Can J Fish Aquat Sci* 69:656–669
- Guiry MD, Guiry GM (2022) *AlgaeBase*. World-wide electronic publication, National University of Ireland, Galway. <https://www.algaebase.org>
- Gushulak CAC, Laird KR, Bennet JR, Cumming BF (2017) Water depth is a strong driver of intra-lake diatom distributions in a small boreal lake. *J Paleolimnol* 56:231–241

- Hadley KR, Paterson AM, Hall RI, Smol JP (2013) Effects of multiple stressors on lakes in south-central Ontario: 15 years of change in lakewater chemistry and sedimentary diatom assemblages. *Aquat Sci* 75:349–360
- Hall RI, Smol JP (1996) Paleolimnological assessment of long-term water-quality changes in south-central Ontario lakes affected by cottage development and acidification. *Can J Fish Aquat Sci* 53:1–17
- Hargan KE, Nelligan C, Jeziorski A, Rühland PAM, Keller W, Smol JP (2016) Tracking the long-term responses of diatoms and cladocerans to climate warming and human influences across lakes of the Ring of Fire in the Far North of Ontario, Canada. *J Paleolimnol* 56:153–172
- Heegaard E (2002) The outer border and central border for species—environmental relationships estimated by non-parametric generalised additive models. *Ecol Modell* 157:131–139
- Hill MO (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology* 54:427–432
- Hübener T, Dreßler M, Schwarz A, Langner K, Adler S (2008) Dynamic adjustment of training sets ('moving-window' reconstruction) by using transfer functions in paleolimnology—a new approach. *J Paleolimnol* 40:79–95
- Hübener T, Adler S, Werner P, Schult M, Erlenkeuser H, Meyer H, Bahnwart M (2009) A multi-proxy paleolimnological reconstruction of trophic state reference conditions for stratified carbonate-rich lakes in Northern Germany. In: Buczkó K, Korponai J, Padišák J, Starratt SW (eds) *Paleolimnological proxies as tools of environmental reconstruction in fresh water*. Springer, Dordrecht, pp 303–327
- Huisman J, Olf H, Fresco LFM (1993) A hierarchical set of models for species response analysis. *J Veg Sci* 4:37–46
- Huot Y, Brown CA, Potvin G, Antoniadis D, Baulch HM, Beisner BE, Bélanger S, Brazeau S, Cabana H, Cardille JA, del Giorgio PA, Gregory-Eaves I, Fortin M-J, Lang AS, Laurion I, Maranger R, Prairie YT, Rusak JA, Segura PA, Siron R, Smol JP, Vinebrooke RD, Walsh DA (2019) The NSERC Canadian Lake Pulse Network: a national assessment of lake health providing science for water management in a changing climate. *Sci Total Environ* 695:133668
- Janhurs S (ed) (1998) 1995 and 1996 performance report: general chemistry and microbiology analyses section. Ontario Ministry of Environment report, p 306
- Jansen F, Oksanen J (2013) How to model species responses along ecological gradients—Huisman–Olf–Fresco models revisited. *J Veg Sci* 24:1108–1117
- Juggins S, Birks HJB (2012) Quantitative environmental reconstructions from biological data. In: Birks HJB, Lotter AF, Juggins S, Smol JP (eds) *Tracking environmental change using lake sediments: data handling and numerical techniques*. Springer, pp 431–494
- Julius ML, Theriot EC (2010) The diatoms: a primer. In: Smol JP, Stoermer EF (eds) *The diatoms: applications for the environmental and earth sciences*, 2nd edn. Cambridge University Press, Cambridge, pp 8–22
- Keller W (2009) Limnology in northeastern Ontario: from acidification to multiple stressors. *Can J Fish Aquat Sci* 66:1189–1198
- Koppen JD (1975) A morphological and taxonomic consideration of *Tabellaria* (Bacillariophyceae) from the north-central United States. *J Phycol* 11:236–244
- Krammer K (1997) Die cymbelloiden Diatomeen. Eine Monographie der weltweit bekannten Taxa. Teil 2. *Encyonema* Part., *Encyonopsis* und *Cymbellopsis*. *Bibl Diatomol* 37:1–469
- Krammer K, Lange-Bertalot H (1986–1991). Bacillariophyceae. In: Ettl H, Gerloff D, Heynig D, Mollenhauer D (eds) *Süßwasserflora von Mitteleuropa*. Gustav Fischer Verlag, Stuttgart, pp 876, 596, 576, 437
- Lecointe C, Coste M, Prygiel J (1993) "Omnidia": software for taxonomy, calculation of diatom indices and inventories management. *Hydrobiologia* 269(270):509–513
- Mann DG, Vanormelingen P (2013) An inordinate fondness? The number, distributions, and origins of diatom species. *J Eukaryot Microbiol* 60:414–420
- Nelligan C, Jeziorski A, Rühland KM, Paterson AM, Smol JP (2016) Managing lake trout lakes in a warming world: a paleolimnological assessment of nutrients and lake production at three Ontario sites. *Lake Reserv Manag* 32:315–328
- Nelligan C, Jeziorski A, Rühland KM, Paterson AM, Meyer-Jacob C, Smol JP (2020) A multibasin comparison of historical water quality trends in Lake Manitou, Ontario, a provincially significant lake trout lake. *Lake Reserv Manag* 36:243–259
- Ontario Ministry of Environment (OME) (1979) Determination of the susceptibility to acidification of poorly buffered surface water. Technical Report, p 22
- Ontario Ministry of Environment (OME) (1981) Outlines of analytical methods. Technical report
- Ontario Ministry of the Environment (OME) (1983) Handbook of analytical methods for environmental samples, vols 1 and 2. Laboratory Services Branch, Ontario Ministry of the Environment and Energy, Sudbury
- Oksanen J, Minchin PR (2002) Continuum theory revisited: what shape are species responses along ecological gradients? *Ecol Modell* 157:119–129
- Patrick R, Reimer CW (1966) The diatoms of the united states. vol 1. Monographs of the Academy of Natural Sciences of Philadelphia, pp 1–688
- R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ramstack JM, Fritz SC, Engstrom DR, Heiskary SA (2003) The application of a diatom-based transfer function to evaluate regional water-quality trends in Minnesota since 1970. *J Paleolimnol* 29:79–94
- Ramstack JM, Fritz SC, Engstrom DR (2004) Twentieth century water quality trends in Minnesota lakes compared with presettlement variability. *Can J Fish Aquat Sci* 61:561–576
- Reavie ED, Kireta AR (2015) Centric, Araphid and Eunotioid Diatoms of the Coastal Laurentian Great Lakes. In: *Bibliotheca Diatomologica* Band. Vol 62. J. Cramer. Berlin, p 184
- Reavie ED, Juggins S (2011) Exploration of sample size and diatom-based indicator performance in three North American phosphorus training sets. *Aquat Ecol* 45:529–538

- Reavie ED, Smol JP (2001) Diatom-environmental relationships in 64 alkaline southeastern Ontario (Canada) lakes: a diatom-based model for water quality reconstructions. *J Paleolimnol* 25:25–42
- Round FE, Crawford RM, Mann DG (1990) *The diatoms: biology and morphology of the Genera*. Cambridge University Press, p 747
- Rühland KM, Smol JP (2002) Freshwater diatoms from the Canadian Arctic treeline and development of paleolimnological inference models. *J Phycol* 38:249–264
- Rühland KM, Smol JP, Pienitz R (2003) Ecology and spatial distributions of surface-sediment diatoms from 77 lakes in the subarctic Canadian treeline region. *Can J Bot* 81:57–73
- Rühland KM, Paterson AM, Keller W, Michelutti N, Smol JP (2013) Global warming triggers the loss of a key Arctic refugium. *Proc R Soc B* 280:20131887
- Rühland KM, Hargan KE, Jezierski A, Paterson AM, Keller W, Smol JP (2014) A multi-trophic exploratory survey of recent environmental changes using lake sediments in the Hudson Bay Lowlands, Ontario, Canada. *Arct Antarct Alp Res* 46:139–158
- Rühland KM, Paterson AM, Smol JP (2015) Lake diatom responses to warming: reviewing the evidence. *J Paleolimnol* 54:1–35
- Schindler DW (1988) Effects of acid rain on freshwater ecosystems. *Science* 239:149–157
- Schindler DW, Hecky RE, Stainton MP, Parker BR, Paterson MJ, Beaty KG, Lyng M, Kasian SEM (2008) Eutrophication of lakes cannot be controlled by reducing nitrogen input: results of a 37-year whole-ecosystem experiment. *Proc Natl Acad Sci USA* 105:11254–11258
- Schultz K, Dreßler M (2022) Transfer of *Cyclotella michiganiana* Skvortzov to the genus *Pantocsekiella* K.T. Kiss & Ács (Stephanodiscaceae, Bacillariophyta). *Not Algarum* 250:1–3
- Simpson GL, Hall RI (2012) Human impacts: applications of Numerical methods to evaluate surface-water acidification and eutrophication. In: *Tracking Environmental Change Using Lake Sediments*, pp 579–614
- Simpson GL, Oksanen J (2020) analogue: analogue matching and modern analogue technique transfer function models. (R package version 0.17–5) <https://cran.r-project.org/package=analogue>
- Sivarajah B, Rühland KM, Labaj AL, Paterson AM, Smol JP (2016) Why is the relative abundance of *Asterionella formosa* increasing in a Boreal Shield lake as nutrient levels decline? *J Paleolimnol* 55:357–367
- Sivarajah B, Paterson AM, Rühland KM, Köster D, Karst-Riddoch T, Smol JP (2018) Diatom responses to 20th century shoreline development and climate warming in three embayments of Georgian Bay, Lake Huron. *J Great Lakes Res* 44:1339–1350
- Siver PA, Kling H (1997) Morphological observations of *Aulacoseira* using scanning electron microscopy. *Can J Bot* 75:1807–1835
- Smol JP (2008) *Pollution of lakes and rivers: a paleoenvironmental perspective*. 2nd edition. Wiley-Blackwell Publishing, Oxford, p 383
- Smol JP, Stoermer EF (eds) (2010) *The diatoms: applications for the environmental and earth sciences*. Cambridge University Press, p 686
- Spaulding SA, Potapova MG, Bishop IW, Lee SS, Gasperak TS, Javanoska E, Furey PC, Edlund MB (2022) Diatoms.org: supporting taxonomists, connecting communities. *Diatom Res* 36:291–304
- ter Braak CJF, Looman CWN (1986) Weighted averaging, logistic regression and the Gaussian response model. *Veg-etatio* 65:3–11
- Tremblay R, Pienitz R, Legendre P (2014) Reconstructing phosphorus levels using models based on the modern diatom assemblages of 55 lakes in southern Quebec. *Can J Fish Aquat Sci* 71:887–914
- Valleau RE, Rühland KM, Paterson AM, Smol JP (2022) Using diatoms to track road-salt seepage into small, shallow, softwater Ontario lakes. *Can J Fish Aquat Sci* 79:9
- Valleau RE (2021) *Assessing the effects of road salt application (NaCl) on Cladocera and diatoms across a gradient of chloride concentrations*. PhD dissertation, Queen's University
- Van Staden TL, Van Meter KJ, Basu NB, Parsons CT, Akbarzadeh Z, Van Cappellen P (2022) Agricultural phosphorus surplus trajectories for Ontario, Canada (1961–2016), and erosional export risk. *Sci Tot Environ* 818:151717
- Werner P, Smol JP (2005) Diatom-environmental relationships and nutrient transfer functions from contrasting shallow and deep limestone lakes in Ontario, Canada. *Hydrobiologia* 533:145–173
- Wetzel RG (2001) *Limnology: lake and river ecosystems*, 3rd edn. Academic Press, San Diego, p 1006
- Wickham H (2016) *ggplot2: elegant graphics for data analysis*. Springer, New York, p 213
- Wilkins C (2021) *Have watershed disturbances over the past ~200 years affected diatom composition in Algonquin Provincial Parks? A paleolimnological assessment*. Master's dissertation, Queen's University
- Wilson SE, Cumming BF, Smol JP (1994) Diatom-salinity relationships in 111 lakes from the Interior Plateau of British Columbia, Canada: the development of diatom-based models for paleosalinity reconstructions. *J Paleolimnol* 12:197–221
- Wilson SE, Cumming BF, Smol JP (1996) Assessing the reliability of salinity inference models from diatom assemblage: an examination of a 219-lake data set from western North America. *Can J Fish Aquat Sci* 53:1580–1594

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.