J Paleolimnol (2010) 43:955–975
DOI 10.1007/s10933-009-9380-1

ORIGINAL PAPER

Controls on the contemporary distribution of lake thecamoebians (testate amoebae) within the Greater Toronto Area and their potential as water quality indicators

Helen M. Roe · R. Timothy Patterson · Graeme T. Swindles

Received: 27 January 2009 / Accepted: 3 September 2009 / Published online: 10 October 2009
© Springer Science+Business Media B.V. 2009

Abstract Thecamoebians were examined from 71 surface sediment samples collected from 21 lakes and ponds in the Greater Toronto Area to (1) elucidate the controls on faunal distribution in modern lake environments; and (2) to consider the utility of thecamoebians in quantitative studies of water quality change. This area was chosen because it includes a high density of kettle and other lakes which are threatened by urban development and where water quality has deteriorated locally as a result of contaminant inputs, particularly nutrients. Fifty-eight samples yielded statistically significant thecamoebian populations. The most diverse faunas (highest Shannon Diversity Index values) were recorded in lakes beyond the limits of urban development, although the faunas of all lakes showed signs of sub-optimal conditions. The assemblages were divided into five clusters using Q-mode cluster analysis, supported by Detrended Correspondence Analysis. Canonical Correspondence Analysis (CCA) was used to examine species-environment relationships and to explain the observed clusterings. Twenty-four measured environmental variables were considered, including water property attributes (e.g., pH, conductivity, dissolved oxygen), substrate characteristics, sediment-based phosphorus (Olsen P) and 11 environmentally available metals. The thecamoebian assemblages showed a strong association with phosphorus, reflecting the eutrophic status of many of the lakes, and locally to elevated conductivity measurements, which appear to reflect road salt inputs associated with winter de-icing operations. Substrate characteristics, total organic carbon and metal contaminants (particularly Cu and Mg) also influenced the faunas of some samples. A series of partial CCAs show that of the measured variables, sedimentary phosphorus has the largest influence on assemblage distribution, explaining 6.98% ($P < 0.002$) of the total variance. A transfer function was developed for sedimentary phosphorus (Olsen P) using 58 samples from 15 of the studied lakes. The best performing model was based on weighted averaging with inverse deshrinking (WA Inv, $r^2_{jack} = 0.33$, RMSEP = 102.65 ppm). This model was applied to a small modern thecamoebian dataset from a eutrophic lake in northern Ontario to predict phosphorus and performed satisfactorily. This preliminary study confirms that thecamoebians have considerable potential as quantitative water quality

Electronic supplementary material The online version of this article (doi:10.1007/s10933-009-9380-1) contains supplementary material, which is available to authorized users.

H. M. Roe (✉)
School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, Belfast BT7 1NN, UK
e-mail: h.roe@qub.ac.uk

R. T. Patterson
Ottawa-Carleton Geoscience Centre and Department of Earth Sciences, Carleton University, Ottawa, ON K1S 5B6, Canada

G. T. Swindles
Division of Archaeological, Geographical and Environmental Sciences (AGES), School of Life Sciences, University of Bradford, Bradford BD7 1DP, UK

indicators in urbanising regions, particularly in areas influenced by nutrient inputs and road salts.

Keywords Thecamoebians · Testate amoebae · Lakes · Eutrophication · Phosphorus · Transfer function

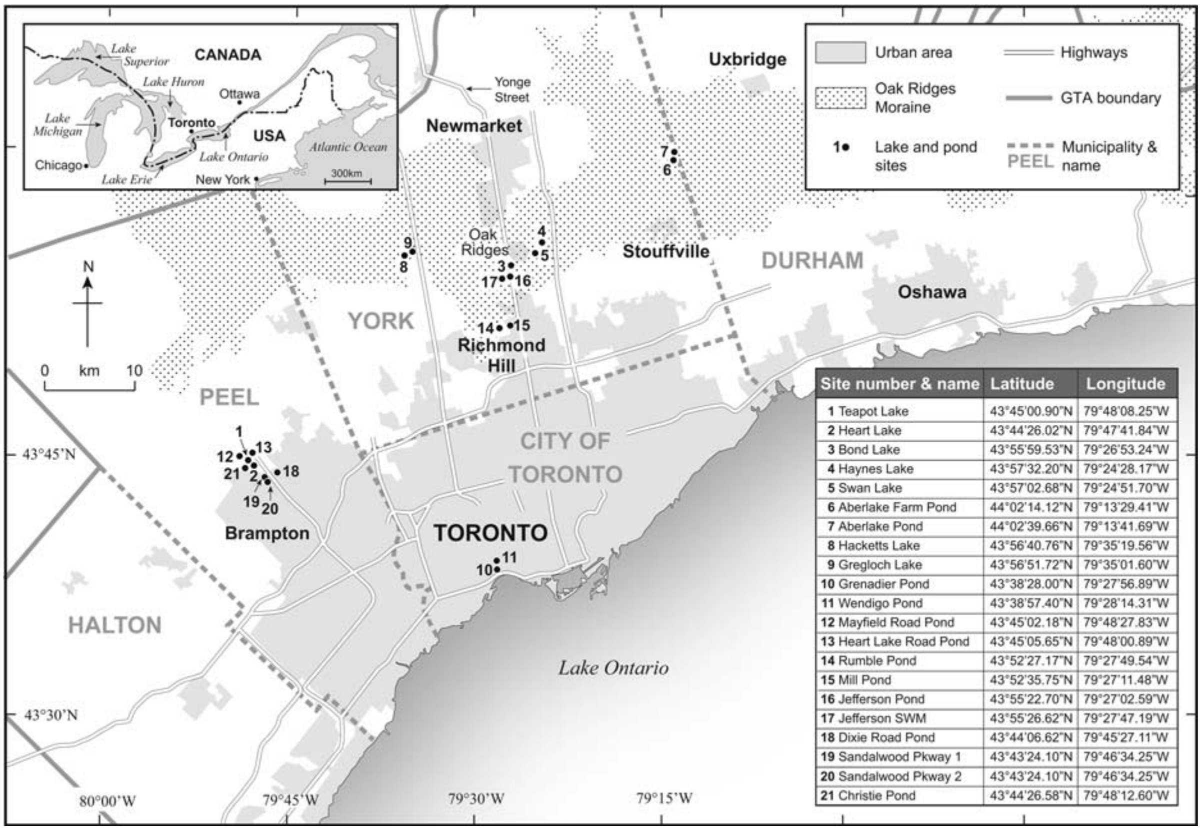
Introduction

Thecamoebians (testate amoebae), are a group of unicellular protozoans that occur in freshwater to brackish environments (Medioli and Scott 1983; Charman et al. 1998; Roe et al. 2002). They are useful for palaeoenvionmental reconstruction because they are sensitive to a wide variety of environmental variables (Charman et al. 2000; Patterson and Kumar 2002) and because their tests are generally resistant to dissolution (Medioli et al. 1990; Swindles and Roe 2007). In lakes, faunal assemblages can be correlated with many environmental parameters including substrate changes associated with forest fires, de-forestation and land clearance (Reinhardt et al. 2005), eutrophication (Schönborn 1990) water temperature change (Collins et al. 1990), salinity (Roe and Patterson 2006), pH (Escobar et al. 2008) and metal and organic pollutant contamination (Patterson and Kumar 2000a). Research on thecamoebians from peatlands is more developed than in lakes due to known, quantified relationships between modern peatland faunas and hydrological controls, particularly water table fluctuations (Booth and Jackson 2003; Swindles et al. 2007a, 2009). Training sets have also been developed for thecamoebians in saltmarsh environments as a basis for quantitative sea-level reconstruction (Gehrels et al. 2001). To date no training sets or transfer functions have been developed in any lake thecamoebian study.

In this paper we report on the findings of a preliminary investigation into the utility of thecamoebians as water-quality indicators in a series of contemporary lake environments in the Greater Toronto Area (GTA) and adjacent Oak Ridges Moraine (Fig. 1). This region was chosen because (1) it includes a high density of small kettle lakes, which are particularly sensitive to catchment disturbance; (2) it is characterised by several different types of land-use along an urban to rural (mainly agricultural) gradient; and (3) the area is currently under significant threat from development and there is an urgent need for baseline limnological data to aid future long-term watershed management and planning. Moreover, fossil thecamoebian datasets collected from kettle lakes in the region have indicated that thecamoebians appear to be sensitive to past episodes of anthropogenic disturbance, including land clearance and changes in the trophic status of the lakes associated with widespread fertiliser application in the mid-twentieth century (Patterson et al. 2002). The area thus represents an important testing area to examine the relationships between thecamoebian faunas, catchment disturbance and water quality in contemporary lake environments and to begin to quantify these relationships for application to fossil studies.

The GTA is a rapidly expanding urban area on the northern shore of Lake Ontario (Fig. 1). Due to its geography, most of the population growth within the region is occurring in the municipalities of Peel, York and Durham adjacent to and including the Oak Ridges Moraine, a predominantly rural region characterised by areas of agricultural land, woodland and small urban centres (Fig. 1). Significantly, this area serves as an important aquifer for the growing population and supports a number of endangered habitats, including kettle lakes. Both are under increasing threat from development (Williams et al. 1999). As urbanisation continues it is essential that regional planners have a firm understanding of pre-settlement (baseline) conditions so that ecologically informed and effective watershed management policies can be implemented to protect them.

As a precursor to assessing pre-urbanised conditions it is important to document the current environmental variability across lakes throughout the region, both within developed and non-developed areas. Whilst water quality datasets exist for some of the larger kettle lakes, the majority of the smaller lakes and ponds have remained uninvestigated. Microfossil-based proxy records of water quality change have an advantage over conventional monitoring approaches in that they provide a time-averaged signal that can be utilised to examine the longer-term response of lakes to disturbance. They can also provide a basis for comparison with pre-disturbance conditions recorded from core data. Ordination data and distributional training sets based on biological indicators can therefore be useful to environmental managers, enabling them to develop mitigative measures to



Site number & name	Latitude	Longitude
1 Teapot Lake	43°45'00.90"N	79°48'08.25"W
2 Heart Lake	43°44'26.02"N	79°47'41.84"W
3 Bond Lake	43°55'59.53"N	79°26'53.24"W
4 Haynes Lake	43°57'32.20"N	79°24'28.17"W
5 Swan Lake	43°57'02.68"N	79°24'51.70"W
6 Aberlake Farm Pond	44°02'14.12"N	79°13'29.41"W
7 Aberlake Pond	44°02'39.66"N	79°13'41.69"W
8 Hacketts Lake	43°56'40.76"N	79°35'19.56"W
9 Gregloch Lake	43°56'51.72"N	79°35'01.60"W
10 Grenadier Pond	43°38'28.00"N	79°27'56.89"W
11 Wendigo Pond	43°38'57.40"N	79°28'14.31"W
12 Mayfield Road Pond	43°45'02.18"N	79°48'27.83"W
13 Heart Lake Road Pond	43°45'05.65"N	79°48'00.89"W
14 Rumble Pond	43°52'27.17"N	79°27'49.54"W
15 Mill Pond	43°52'35.75"N	79°27'11.48"W
16 Jefferson Pond	43°55'22.70"N	79°27'02.59"W
17 Jefferson SWM	43°55'26.62"N	79°27'47.19"W
18 Dixie Road Pond	43°44'06.62"N	79°45'27.11"W
19 Sandalwood Pkway 1	43°43'24.10"N	79°46'34.25"W
20 Sandalwood Pkway 2	43°43'24.10"N	79°46'34.25"W
21 Christie Pond	43°44'26.58"N	79°48'12.60"W

Fig. 1 Location of sampling sites

medium- to long-term responses (decadal to century scale) to urban development in this hydrologically sensitive region.

Kettle lake waters and sediments are particularly sensitive to surrounding land-use changes. This reflects their long water residence time due to limited surface inflows and outflows, their dependence on precipitation and hence atmospherically delivered inputs, and their close connection with groundwater. Due to slow water exchange rates, pollutants entering kettle lakes can become concentrated through time with harmful effects on aquatic biota (Diamond et al. 2002). Monitoring studies based on a limited number of kettle and other lakes in rapidly urbanising parts of the region (e.g., Bond and Wilcox lakes near Oak Ridges; Fig. 1) indicate that some lakes have already experienced a deterioration in water quality and loss of biodiversity as a result of contaminant inputs, particularly external phosphorus loading (Diamond et al. 2002). Bank-side vegetation communities are also changing locally as indigenous plants are being replaced by invasives (e.g., *Typha* species), which are better able to tolerate sub-optimal environments rich in contaminants. Excessive algal growth and eutrophication are also a problem in some urban lakes in the region (Bradford and Maude 2002).

Non-point source pollutants associated with residential developments and rapid infrastructure expansion are of particular concern to environmental managers within the GTA. These enter lakes via multiple pathways, including surface runoff and storm sewer networks, groundwater seepage and atmospheric inputs. Pollutants posing a particular threat to lakes and aquatic communities include nutrients (particularly phosphorus) from residential sources, including domestic waste, fertilisers and pesticides, and sewage inputs via combined sewer overflows (TRCA 2008). Other contaminants associated with urban development include trace metals, some of which are known to cause adverse affects on exposed biota (Skinner et al. 1999). Copper, lead and zinc are particularly common in urban runoff (Marsalek and

Shroeter 1988). These have many sources, including municipal waste, road traffic inputs and building materials (Callender and Rice 2000). Suspended solid inputs associated with development are also a concern because solid particles act as a transport vector for other contaminants such as phosphorus and heavy metals. Inputs of road salts into water bodies as a result of the extensive winter de-icing operations that occur in the region also pose a significant challenge to environmental managers (Williams et al. 1999).

This study has three specific objectives: (1) to examine the ecological preferences of lake thecamoebians along a urban–rural gradient in the GTA and adjacent ORM to elucidate the relationship between modern thecamoebian faunas, lake water quality and catchment disturbance; (2) to consider the utility of thecamoebian distributional datasets to future down-core studies; and (3) to provide baseline data for future long-term hydrological management and planning. To achieve these objectives we consider thecamoebian and sediment and water property data from a selection of natural (mainly kettle) and artificial (constructed) lakes within the study area and quantify species-environment relationships using multivariate statistical approaches.

Materials and methods

Sampling design

Two clusters of kettle lakes were selected for detailed thecamoebian distributional study along two development 'corridors' in the Richmond Hill-Oak Ridges and Brampton areas (Fig. 1). These peripheral zones of Toronto are characterised by rapid urban growth and include a mixture of well-established residential and urban core land and extensive new (<5 years) housing developments. Woodland and agricultural land occur at the margins. The sampling set included four lakes from urbanised sites within these areas, including Christie Pond (site 20) in Brampton which is surrounded by detached homes, Bond Lake in the rapidly expanding town of Oak Ridges, and nearby Jefferson Pond (site 16) which at the time of sampling was surrounded by an unfinished housing development (Fig. 1; Electronic Supplementary Material (ESM) 1c, d).

A second group of kettle lakes was located beyond the urban limits at the time of sampling yet within regions earmarked for future development. These included Teapot and Heart lakes (sites 1 and 2) in Brampton and Swan Lake (site 5) near Oak Ridges (Fig. 1). Teapot and Heart Lake lie within the 'Heart Lake Conservation Area' a 169-hectare area of 'green space' (mainly woodland) located inside the boundaries of the burgeoning city of Brampton. Three lakes from rural areas well beyond the urban limits were also sampled: Gregloch (site 9), Hacketts Lake (site 8) and Aberlake (site 7). Land use in these areas is mainly agricultural (arable or pasture) with some woodland. Haynes Lake (site 4), also in a rural setting east of Oak Ridges, was included as it is bounded by a road which is regularly treated with winter de-icing salts (Fig. 1; ESM 1a, b).

Further samples were collected from Grenadier Pond in a downtown Toronto, a 0.8-km-long natural lake adjacent to Lake Ontario which receives runoff from a large urban catchment and which has a history of contamination (ESM 1j). Samples were also collected from a small group of constructed lakes and ponds, including Mill Pond, a lined urban pond in the town of Richmond Hill, and six urban storm-water management (SWM) ponds, including Dixie Road Pond near Brampton (site 18), Rumble Pond in Richmond Hill (site 14) and Wendigo Pond in downtown Toronto (site 11) (Fig. 1). SWM ponds are being used increasingly in the region to improve water quality and to regulate storm runoff as part of a series of sustainable urban drainage initiatives. The ponds are designed to trap contaminants derived especially from residential developments or in lake catchments with a history of pollution, e.g., Grenadier Pond. It was anticipated that these ponds would also be characterised by elevated levels of contaminants, particularly heavy metals, and would again shed light on thecamoebian response to sediment pollution. In total 21 lakes and ponds were sampled, 11 natural kettle lakes, Grenadier Pond and nine artificial lakes or constructed (SWM) ponds. The sub-surface geology in the sampled areas comprises Quaternary glacigenic sediments, variously dominated by sands, silts and diamictons.

Field and laboratory methods

Seventy-one sediment/water interface samples were collected with an Ekman grab sampler suspended from a small boat. The upper 0.5 mm of sediment was retained for analysis. As thecamoebians are sensitive to changes in water depth (Patterson and Kumar 2002) multiple samples (>4 per site) were usually collected from the kettle lakes and from Grenadier Pond, whilst only 1–2 samples were collected from the generally shallower SWM ponds.

Water depths for each sampling station were determined using a Knudsen BP-320 sub-bottom profiler. Other field-based environmental variables measured included dissolved oxygen, redox potential, pH, conductivity and temperature, which were determined with a portable HydroLab water property probe. The percentage cover of *Typha* around the edge of each lake was estimated to examine possible relationships between thecamoebian distribution, lake contamination and *Typha* abundance.

In the laboratory, particle size analysis (% clay, silt and sand) was carried out using a Malvern Mastersizer-2000 particle size analyser, whilst total organic carbon (TOC) was measured following the method outlined by Hesse (1971). Sediment-based phosphorus was measured to consider the trophic status of the lakes. Phosphorus was examined using the Olsen's phosphorus (Olsen P) extraction method, which provides a measure of bio-available phosphorus (Zhou et al. 2001) and is a suitable extraction method for samples of neutral to alkaline pH. Phosphorus concentrations were determined using phosphomolybdate colorimetric technique (Watanabe and Olsen 1965). A number of environmentally available metals (Ca, Mg, Na, K, Fe, Mn, Zn, Cu, Pb, Cr, Ni) which are important by-products of urban run-off were also analysed from the sediments using a microwave digestion procedure US EPA 3051A (Link et al. 1998). Analyses employed a Perkin Elmer Analyst 200 atomic absorption spectrophotometer using a flame technique. A reference standard (WQB-1) from the National Water Research Institute, Canada was used as a quality control measure.

A 5-cc subsample of sediment was used for thecamoebian analysis. Samples were agitated for 1 h using a Burrell wrist shaker, screened with a 37-µm sieve to remove fine organic and mineral detritus. The >37-µm fraction samples were subdivided into aliquots for quantitative analysis using a wet splitter. Wet aliquots were examined under a dissecting binocular microscope until a statistically significant number of specimens were quantified (Patterson and Fishbein 1989). In most cases >150, and often >250 thecamoebians were counted per sample (ESM 2).

Identification of thecamoebians was undertaken with reference to keys, including Kumar and Dalby (1998). Lacustrine thecamoebian species can display a significant amount of environmentally controlled morphological instability (Medioli et al. 1987). The accepted practise by lacustrine researchers has been to designate informal infraspecic strain names for these ecophenotypes (Asioli et al. 1996; Patterson and Kumar 2002). Although these infraspecic level designations have no status under the International Zoological Code of Nomenclature, they are useful for delineating environmentally significant populations within lacustrine environments (Escobar et al. 2008) so these terms were employed during counting. Scanning electron micrograph images of common species and strains were obtained using a JEOL 6400 scanning electron microscope (Fig. 2).

Statistical methods

Twenty-three thecamoebian species and strains were identified in the 71 collected samples. As all thecamoebian species were found in statistically significant numbers in at least one sample (Patterson and Fishbein 1989), none were removed from ensuing multivariate data analysis. The 71 samples were also assessed to determine which were statistically significant using the approach of Fishbein and Patterson (1993). A total of 58 samples were deemed to have statistically significant populations. Thirteen discarded samples containing statistically insignificant populations were not included in subsequent multivariate analysis.

The Shannon Diversity Index (SDI) was used to examine the faunal diversity of the species found in each sample and thus provide an indication of the relative health of the lakes and ponds (Shannon 1948). The SDI is defined as:

$$\text{S.I.} = -\sum_{i=1}^{S} \left(\frac{X_i}{N_i}\right) \times \ln\left(\frac{X_i}{N_i}\right) \quad (1)$$

where X_i is the abundance of each taxon in a sample, N_i is the total abundance of the sample, and S is equal to the species richness of the sample. Environments are considered to be stable if the SDI falls between 2.5 and 3.5, in transition between 1.5 and 2.5, and stressed between 0.1 and 1.5 (Magurran 1988; Patterson and Kumar 2000a).

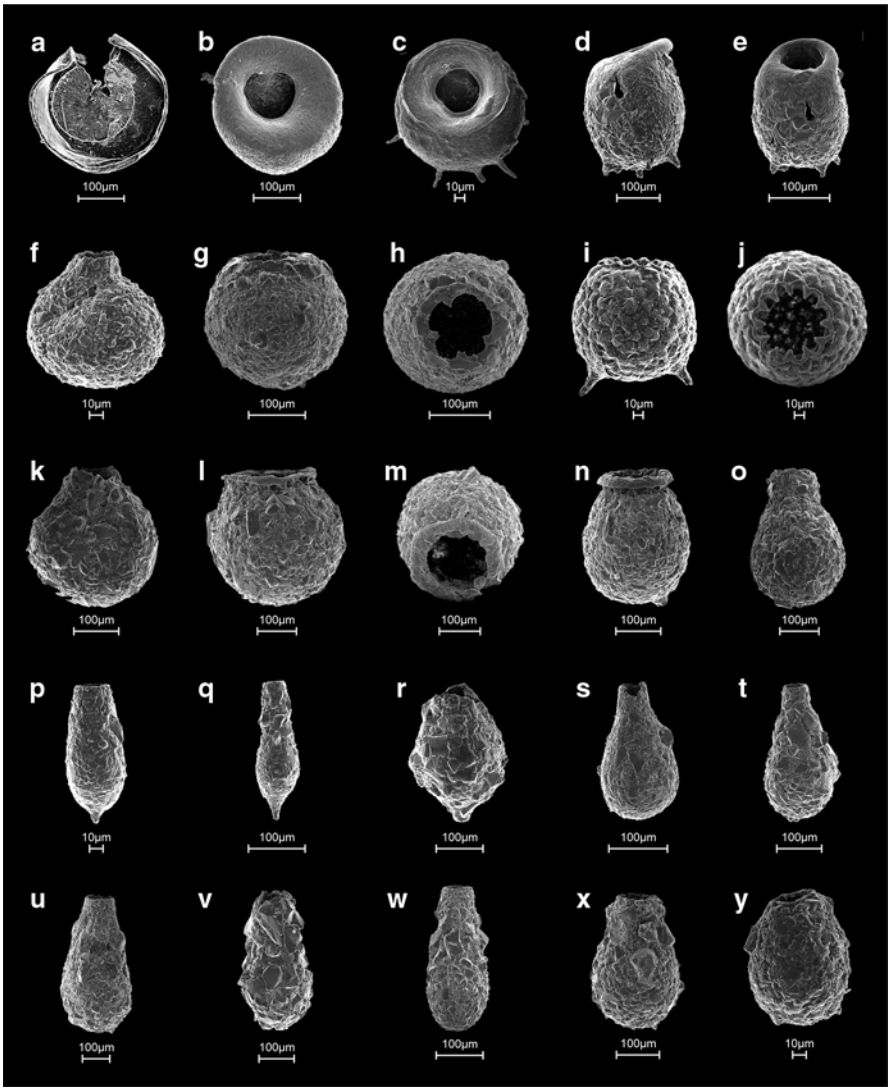


Fig. 2 **a**–**y** Scanning electron micrographs of selected thecamoebian tests from the study lakes. See ESM 6 for additional specimen information. **a** *Arcella vulgaris* Ehrenberg 1830 from Bond Lake. **b** *Centropyxis aculeata* (Ehrenberg 1832) strain "discoides" from Bond Lake. **c** *Centropyxis aculeata* (Ehrenberg 1832) strain "aculeata" from Teapot Lake. **d**, **e** *Centropyxis constricta* (Ehrenberg 1843) strain "constricta" from Swan Lake. **f** *Lesqueresia spiralis* (Ehrenberg, 1840) from Bond Lake. **g**, **h** *Cucurbitella tricuspis* (Carter 1856) from Bond Lake. **i**, **j** *Difflugia corona* Wallich, 1864 from Swan Lake. **k** *Pontigulasia compressa* (Carter 1864) from Bond Lake. **l** *Difflugia urceolata* Carter, 1864 strain "urceolata" from Swan Lake. **m**, **n** *Difflugia urceolata* Carter, 1864 strain "elongata" from Aberlake Farm Lake. **o** *Lagenodifflugia vas* Leidy, 1874 from Haynes Lake. **p** *Difflugia protaeiformis* Lamark 1816 strain "acuminata" from Swan Lake. **q** *Difflugia protaeiformis* Lamark 1816 strain "claviformis" from Bond Lake. **r** *Difflugia protaeiformis* Lamark 1816 strain "amphoralis" from Bond Lake. **s** *Difflugia oblonga* Ehrenberg, 1832 strain "oblonga" from Aberlake Farm Lake. **t** *Difflugia oblonga* Ehrenberg, 1832 strain "oblonga" from Bond Lake. **u** *Difflugia oblonga* Ehrenberg, 1832 strain "linearis" from Bond Lake. **v** *Difflugia oblonga* Ehrenberg, 1832 strain "bryophila" from Bond Lake. **w** *Difflugia oblonga* Ehrenberg, 1832 strain "lanceolata" from Bond Lake. **x** *Difflugia oblonga* Ehrenberg, 1832 strain "tenuis" from Bond Lake. **y** *Difflugia oblonga* Ehrenberg, 1832 strain "glans" from Bond Lake

R-mode cluster analysis was used to determine which species were most closely associated with others and thus best characterised a particular assemblage (Fishbein and Patterson 1993). Q-mode cluster analysis was used to group statistically similar populations using Ward's Minimum variance method, and recorded as squared-Euclidean distances (Fishbein and Patterson 1993). Q-mode and R-mode cluster analyses were carried out on the 23 thecamoebian species and strains in the 58 statistically significant samples and organised into a hierarchical diagram. Detrended Correspondence Analysis (DCA) was carried out to explore the inter-site characteristics of the testate amoebae communities. Pearson correlation analysis was used to determine the intercorrelations between environmental variables and hence assess the degree of redundancy in the dataset (Birks 1995).

Canonical Correspondence Analysis (CCA) was used to examine the relationships between testate amoebae taxa and the measured environmental variables using CANOCO version 4.5 and CANODRAW (ter Braak 2002; ter Braak and Šmilauer 2002). Rare species were down-weighted to reduce the problem of the chi-square measure giving rare species a large influence on the ordination (Legendre and Legendre 1998). The SDI values were also included in the CCA as a passive variable.

A series of partial CCAs (Borcard et al. 1992) were carried out to investigate the proportions of variance explained by the environmental variables. Monte–Carlo permutation tests (499 permutations under the full model) were used to determine the statistical significance of the ordination (Dale and Dale 2002).

As the multivariate statistical analyses revealed the importance of sedimentary phosphorus (P) as a major environmental control on thecamoebians in the study sites, a transfer function was developed for P based on weighted averaging regression with inverse deshrinking.

Results

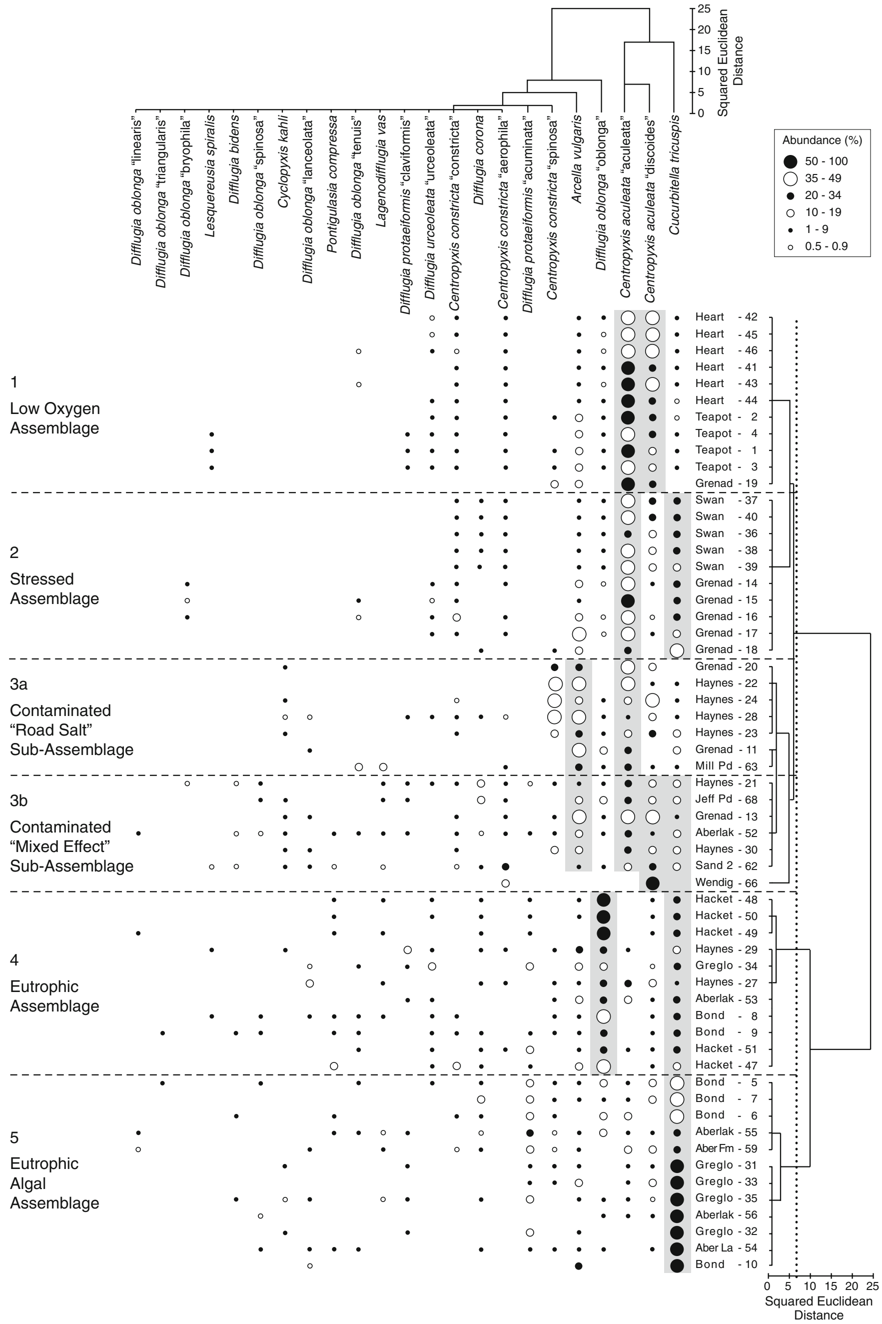
Zonation and CCA results

Interpretation of the Q-mode cluster analysis resulted in the recognition of five distinct thecameobian assemblages: (1) 'Low oxygen assemblage'; (2) 'Stressed assemblage'; (3) 'Contaminated assemblage'; (4) 'Eutrophic assemblage'; and (5) 'Eutrophic algal assemblage' (Fig. 3). The 'Contaminated assemblage' was further subdivided into a 'Contaminated Road Salt' (3a) and 'Mixed effect' (3b) sub-assemblages (Fig. 3). Although 23 species of thecamoebians were included in the analysed dataset, R-mode cluster analysis indicated that only seven species and strains significantly influence assemblage composition: *Cucurbitella tricuspis* Carter, 1856, *Centropyxis aculeata* (Ehrenberg 1832) strain "discoides", *Centropyxis aculeata* (Ehrenberg 1832) strain "aculeata", *Difflugia oblonga* Ehrenberg, 1832 strain "oblonga", *Arcella vulgaris* Ehrenberg 1832, *Centropyxis constricta* (Ehrenberg 1843) "spinosa" and *Difflugia protaeiformis* Lamark 1816 strain "acuminata" (Fig. 3). The CCA results (Figs. 4, 5) and DCA (ESM 3) are discussed in the context of the five identified faunal assemblages. To further test the validity of the CCA results, a Redundancy Analysis (RDA) analysis with Hellinger-transformed species data (Legendre and Gallagher 2001) was carried out, which provided very similar results (ESM 4). The CCA plots, which show the 58 samples with statistically significant thecamoebian populations, also include nine samples with statistically insignificant counts (<50 specimens per sample), which are plotted passively on the sample-environment CCA bi-plot (Fig. 4). An additional four samples (Grenadier Pond 12, Haynes Lake 26, Mayfield Road Pond 70, Aberlake Farm Pond 57) were barren and were excluded from the analyses. CCA axes one (Eigenvalue = 0.418) and two (Eigenvalue = 0.170) explain 34.3% of the variance in the species data and 53.3% of the species-environment relationship (ESM 2), whilst the measured environmental variables together explain 64.3% of the variation in the testate amoebae data. The Monte–Carlo permutation tests show that the first and all canonical axes are significant at $P < 0.002$, Axis 2 is significant at $P < 0.0160$ and Axis 3 is significant at $P < 0.1360$.

Assemblage 1—Low oxygen assemblage

This assemblage is dominated by *C. aculeata* "aculeata" (43.1–54.8%) with lesser proportions of *C. aculeata* "discoides" (15.6–43.2%) and *Arcella vulgaris* (2.7–14.8%) (Fig. 3). The samples within this assemblage were all from Heart Lake and nearby Teapot Lake (Fig. 1), both located in the Heart Lake Conservation Area north of Brampton. The results of the DCA analysis show that low oxygen assemblage (LOA) samples have a high degree of similarity as they are tightly grouped at both the inter- and intra-lake level (ESM 3). CCA analysis of the sample distribution (Fig. 4) indicates that this assemblage is

Fig. 3 R-mode vs Q-mode cluster diagram for the 58 samples with statistically significant thecamoebian counts. Five faunal assemblages (1–5) are indicated, including two sub-assemblages (**a**, **b**) for Assemblage 3. The *dashed line* discriminates clusters of samples with correlation coefficients greater than the selected level of significance. Aberlak = Aberlake Pond; Aber Farm = Aberlake Farm Pond; Greglo = Gregloch Lake; Grenad = Grenadier Pond; Hacket = Hacketts Lake; Jeff Pond = Jefferson Pond; Sand 2 = Sandalwood Parkway SWM pond 2



negatively correlated with levels of dissolved oxygen, with *C. aculeata* "aculeata", the dominant species of the assemblage, also displaying a negative correlation with dissolved oxygen levels (Fig. 5). The LOA tends to characterise inhospitable cold water and low oxygen environments near or below the thermocline in these lakes. The SDI values obtained for LOA samples confirm the generally inhospitable nature of the environments characterised by this assemblage; low within Heart Lake (SDI = 1.13–1.27) with more moderate values obtained for Teapot Lake (SDI = 1.44–1.83).

Centropyxids, particularly the strains *C. aculeata* "discoides" and *C. aculeata* "aculeata" are capable

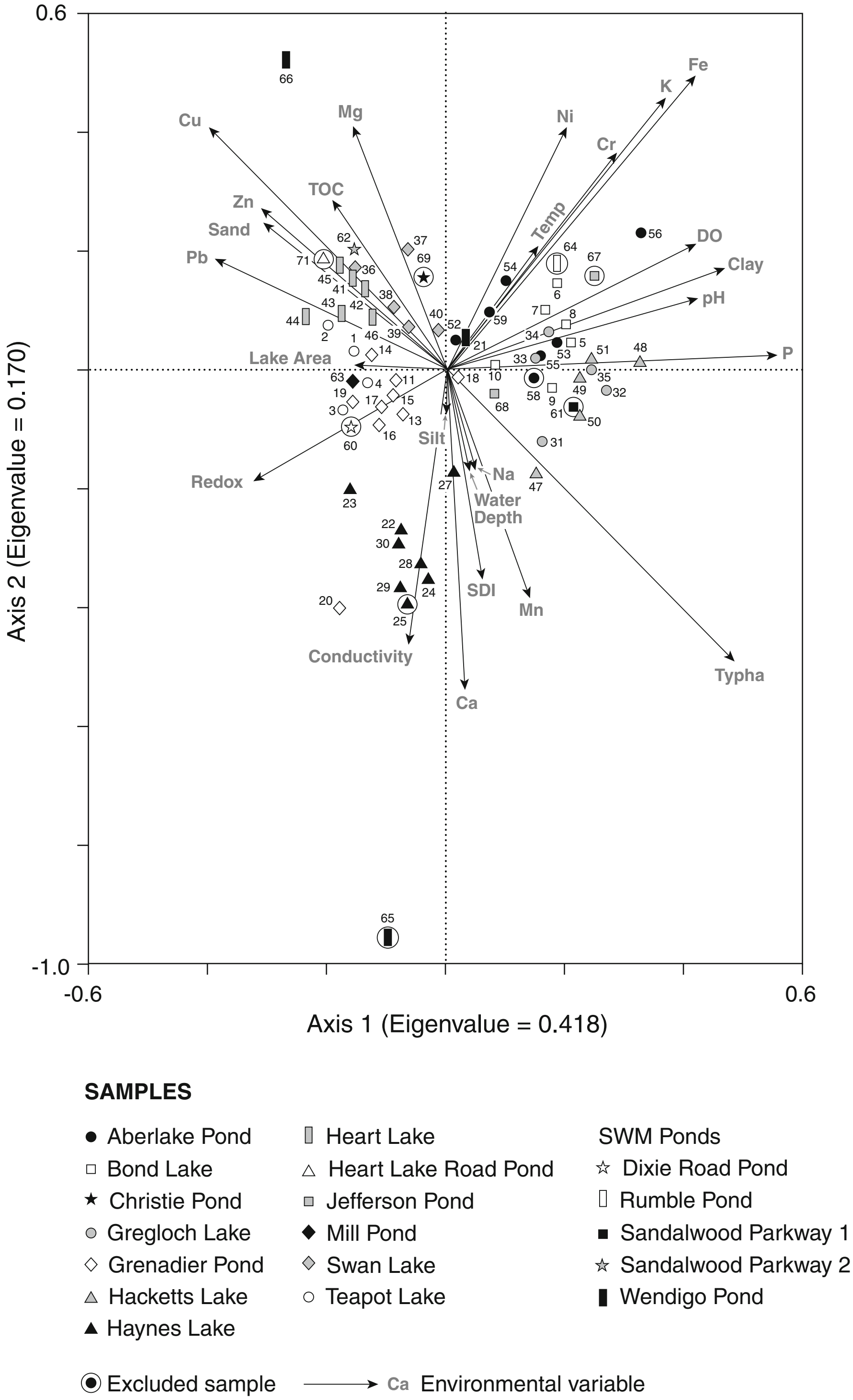


Fig. 4 Canonical Correspondence Analysis (CCA) sample-environment bi-plot for the 67 samples that yielded thecamoebians. Samples 25, 50, 58, 60, 64, 65, 67, 69 and 71 failed to yield statistically significant faunal populations so were excluded from the analyses and are plotted passively on the diagram. TOC = Total organic carbon, DO = dissolved oxygen, *Typha* = % edge of lake occupied by *Typha*; Redox = redox potential; Temp = temperature

of withstanding a variety of hostile lacustrine conditions better than most thecamoebian species. These include low salinities (<5 ppt) (Scott and Medioli 1980), oligotrophic conditions, low nutrient conditions (Schönborn 1984), sites contaminated by metals and other pollutants (Reinhardt et al. 1998; Patterson and Kumar 2000b) as well as the cold and low oxygen conditions that characterise the LOA

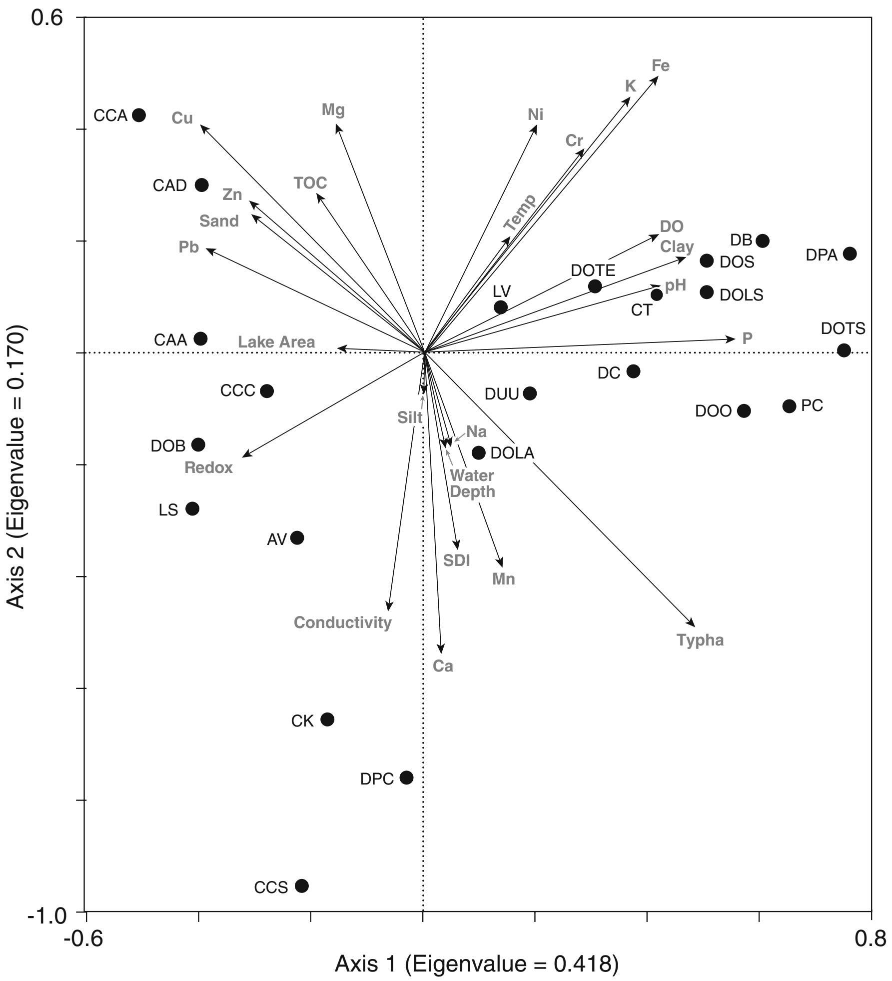


Fig. 5 CCA species-environment bi-plot. AV = *Arcella vulgaris*; CAA = *Centropyxis aculeata* "aculeata"; CAD = *Centropyxis aculeata* "discoides"; CCA = *Centropyxis constricta* "aerophila"; CCC = *Centropyxis constricta* "constricta"; CCS = *Centropyxis constricta* "spinosa"; CK = *Cyclopyxis kahli*; CT = *Cucurbitella tricuspis*; DB = *Difflugia bidens*; DC = *Difflugia corona*; DOB = *Difflugia oblonga* "bryophila"; DOLA = *Difflugia oblonga* "lanceolata"; DOO = *Difflugia oblonga* "oblonga"; DOLS = *Difflugia oblonga* "linearis"; DOS = *Difflugia oblonga* "spinosa"; DOTE = *Difflugia oblonga* "tenuis"; DOTS = *Difflugia oblonga* "triangularis"; DPA = *Difflugia protaeiformis* "acuminata"; DPC = *Difflugia protaeiformis* "claviformis"; DUU = *Difflugia urceolata* "urceolata"; LS = *Lesquereusia spiralis*; LV = *Lagenodifflugia vas*; PC = *Pontigulasia compressa*

(Patterson and Kumar 2002). Interestingly, the CCA analyses (Fig. 4) show that some of the LOA samples, particularly those from Heart Lake, are also associated with elevated levels of copper, lead and zinc, which may additionally explain the centropyxid-dominated assemblages. These metals are common airborne pollutants in urban environments and are closely associated with vehicle traffic (Schroeder et al. 1987). Copper and zinc are also components of runoff in residential areas, where they may be sourced from roofing materials (Boller and Steiner 2002). Although Heart Lake lies within the Heart Lake Conservation Area and is surrounded by a fringe of woodland, the hinterland to the east and south of the lake is predominantly residential in character and is bisected by freeways, which may have acted as a source for these metals.

Two outliers within the LOA, Heart Lake samples 45 and 46, were characterised by near normal oxygen levels (8.24–9.20 mg/l). These samples were both distinguished by sandy substrates (35.49–83.88%), which negatively influence faunal diversity as they are typically low in nutrients and are inhospitable to thecamoebians (Patterson and Kumar 2002; Roe and Patterson 2006). Grenadier Pond Sample 19 also anomalously clustered with the LOA. Although this sample clusters near the LOA samples from Teapot and Heart lakes on the DCA plot, it is more closely associated with the more broadly dispersed grouping of samples from Grenadier Pond (ESM 3). The low diversity (SDI = 1.21) centropyxid-dominated fauna in sample 19 is in this case more likely a result of stressed environmental conditions in Grenadier Pond related to the presence of a contaminated substrate.

Assemblage 2—Stressed assemblage

This assemblage is dominated by *C. aculeata* "aculeata" (43.1–54.8%) with lower abundances of *Cucurbitella tricuspis* (15.6–43.2%) and *A. vulgaris* (2.7–14.8%). Samples from this assemblage were exclusively from Swan Lake and Grenadier Pond. DCA analysis indicates that all stressed assemblage (SA) samples group very closely (ESM 3). The samples from Swan Lake are correlated with total organic carbon (TOC) with those from Grenadier Pond less so (Fig. 4). The SDI values for samples from this assemblage are mixed with values obtained from Grenadier Pond (SDI = 1.14–1.64) being indicative of stressed conditions while those obtained from Swan Lake (SDI = 1.51–1.68) are transitional.

The dominance of *C. aculeata* "aculeata" and *A. vulgaris* provides evidence of stressed environmental conditions within the SA. *C. tricuspis* is most abundant in eutrophic lakes and ponds characterised by conspicuous floating algal mats (Medioli et al. 1987) but has also been found in highly eutrophic lakes devoid of such mats (Medioli and Scott 1983). The close correlation between the distribution of *C. tricuspis* and phosphorus levels, a significant contributor to lake eutrophication on the species-environment CCA plot (Fig. 5), provides confirmation of this relationship. The presence of significant proportions of *C. triscuspis* in the SA thus provides faunal evidence of eutrophication despite the samples from the SA being negatively correlated with phosphorus loading (Fig. 4). It is important to note that the negative correlation between Swan Lake and Grenadier Pond, and phosphorus as plotted in Fig. 4 is relative as all lakes examined in the study are characterised by high Olsen P levels (ESM 2). Grenadier Pond acts as the final filter for water from the downtown Toronto region before it enters Lake Ontario (Fig. 1; ESM 1j). Considering the levels of contaminants that flow into the lake, it is to be expected that the lake ecosystem would be stressed and would support centropyxid-dominated faunas. For example, this lake was characterised by the highest conductivity readings (1,305 mS/cm) of all the studied lakes, presumably reflecting high inputs of dissolved salts and electrolytes from a variety of urban sources. The categorization of Swan Lake, which is associated with much lower conductivity values (<81 mS/cm), with Grenadier Pond is more enigmatic. Swan Lake is surrounded by agricultural land (ESM 1g) where runoff containing fertiliser residues would be expected to enhance productivity (Watchorn et al. 2008), resulting in an associated population increases of indicator taxa such as *C. tricuspis* (Patterson et al. 2002). Although there is an obvious link between land clearance and development of a stressed thecamoebian ecosystem, there are no anomalous water property data or lake sediment geochemistry results, aside from high TOC levels (50.6–65.7%), that may be invoked to precisely explain the observed relationship. Only one sample, sample 37, yielded unusually values of lead, copper and zinc (at 259, 252 and 50 mg/kg^{-1}, respectively), which might have promoted stressed conditions. These metals could have been derived from fertiliser runoff from the surrounding agricultural land, although it is unclear why only one sample was affected. Further, more detailed distributional research is therefore required.

Assemblage 3a—Contaminated 'road salt' sub-assemblage

This assemblage is co-dominated by *A. vulgaris* (15.0–47.1%) and *C. aculeata* "aculeata" (8.5–39.0%), although other centropyxids, principally *C. constricta* "spinosa" are also abundant in some samples. Samples corresponding to this sub-assemblage are primarily from the eastern margin of Haynes Lake the northern and southernmost parts of Grenadier Pond and Mill Pond a constructed urban pond in Richmond Hill (Fig. 1). The DCA analysis results show that all contaminated 'road salt' sub-assemblage (CRSS) samples are relatively closely correlated with each other (ESM 3). SDI values fall within the stressed to transitional range with lower values obtained from Grenadier Pond (SDI = 1.29–1.37) and more moderate values for Haynes Lake and Mill Pond (SDI = 1.37–2.09). Examination of the CCA sample-environment bi-plot (Fig. 4) indicates a strong correlation between CRSS samples and conductivity, with a weaker correlation with redox potential. On the species-environment bi-plot (Fig. 5) *A. vulgaris* is seen to have a particularly strong positive correlation to both conductivity and redox. As discussed in relation to the 'Stressed Assemblage', centropyxids have a tolerance for brackish conditions and all sample localities corresponding to CRSS are subject to road salt contamination due to winter highway de-icing operations. The influence of salt contamination is particularly noticeable at Haynes Lake where a road skirting the entire eastern lake margin (ESM 1a, b) is often partially submerged in winter resulting in the immediate introduction of salt to the lake whenever de-icing operations occur. Two sample stations in Grenadier Pond, 11 and 20, are also subject to periodic salt contamination. Station 20 is located adjacent to a major freeway that runs along the southern end of the lake that receives road salt during de-icing operations (ESM 1j). Similarly, station 11 at the extreme northern end of Grenadier Pond probably receives a considerable influx of contaminated salty water following any precipitation event, as urban storm drains empty into the lake at this point. Sample 63 from Mill Pond was also collected from shallow water immediately adjacent to a road where de-icing salt is introduced.

Assemblage 3b—Contaminated 'mixed effect' sub-assemblage

This sub-assemblage is similar to the CRSS in that both *A. vulgaris* (1.2–24.5%) and *C. aculeata* "aculeata" (15.3–34.1%) are important assemblage components. However, two additional taxa, *C. aculeata* "discoides" (7.8–21.9%) and most significantly *C. tricuspis* (6.1–13.8%) are also present in moderate abundancies. Contaminated 'mixed effect' sub-assemblage (CMES) samples plot relatively closely with each other on the DCA plot (ESM 3), but the lack of overlap between CRSS and CMES samples indicates that division into sub-assemblages is justified (Fig. 4). The SDI values observed for the CMES are higher than for the CRSS, all-falling into the transitional range (SDI = 1.76–2.26), indicating a generally healthy ecosystem. Examination of the CCA sample-environment bi-plot indicates a mixture of influences (Fig. 4), with the presence of *C. tricuspis*, an indicator of eutrophication, providing the unifying factor. Samples from Haynes Lake and Grenadier Pond have a distribution very similar to CRSS samples from those lakes, with the relationship largely controlled by salinity levels. Both CMES sample stations in Haynes Lake (21 and 30) are located adjacent to each other at the northeast end of the lake, very near the road (ESM 1a, b). Water and sediment property data collected from nearby CRSS samples stations are almost indistinguishable to those obtained for CMES samples. Since *C. tricuspis* is seasonally planktic (Medioli et al. 1987) it is possible that lake currents have transported specimens of this species from the western part of the lake to stations 21 and 30 (ESM 1a). CMES sample 13 from Grenadier Pond was collected from just south of the CRSS and just north of SA sample stations in the lake. The fauna characterising sample 13 is thus transitional, being influenced by both the road salt that appears to control the CRSS and to a lesser extent the more eutrophic but still stressed SA (Fig. 4). The mixed nature of influences on this sub-assemblage are further emphasised by examination of CMES sample 62 (Sandalwood Parkway SWM Pond 2), which is strongly influenced by TOC levels, and sample 68 (Jefferson Pond), which is most closely correlated with phosphorus levels and the abundance of *Typha* around the lake (Fig. 4).

Sample 66 collected from Wendigo Pond, a SWM pond to the north of Grenadier Pond plots as an outlier to both the CRSS and CMES, within the Contaminated Assemblage (Fig. 3). This sample, and another sample (65) from the pond also occur as outliers on the CCA sample environment bi-plot (Fig. 4). The thecamoebian assemblage from Wendigo Pond is overwhelmingly dominated by *C. aculeata* "discoides" (88.2%), indicative of extremely ecologically stressed conditions. Wendigo Pond is designed to capture contaminants washing from downtown Toronto streets and storm sewers before the water passes into Grenadier Pond (ESM 1j). The polluted nature of these sediments is indicated by elevated levels of metals (ESM 2).

Assemblage 4—Eutrophic assemblage

This assemblage is dominated by *D. oblonga* "oblonga" (13.9–69.3%) and to a lesser extent by *C. tricuspis* (3.6–27.8%). The associated samples came from five different lakes including Hacketts Aberlake, Bond, Gregloch and Haynes (Fig. 1). Examination of the DCA plot indicates that eutrophic assemblage (EA) samples from Hacketts, Gregloch and Bond lakes are closely correlated with each other while those from Haynes and Aberlake group slightly less well (ESM 3). Samples from all EA lakes, with the exception of those from Haynes, correlate closely with sedimentary phosphorus, which is strongly correlated with Axis 1 on the sample-environment bi-plot (Fig. 4). Phosphorus is often regarded as the main cause of eutrophication in lakes, particularly those subjected to point source pollution from sewage or agriculture, with the trophic state of lakes corresponding well to phosphorus levels in water (Kerekes et al. 2004). Assemblages dominated by *D. oblonga* "oblonga" are typically quite diverse with high SDI values and large populations. As is the case in this study, *Difflugia* thrive under eutrophic conditions and rely on abundant sources of organics to permit maintenance of a habitat with a high carrying capacity (Patterson and Kumar 2000b). This is generally true in this study as SDI values in the EA, with the notable exception of samples from Hacketts Lake (SDI = 0.96–1.82), which are mostly in the transitional range. The source of phosphorus in some of the study lakes in obvious. Both Gregloch and Aberlake are immediately adjacent to farms where faecal matter is introduced to the lakes from horses grazing near the water's edge. In the case of Aberlake, a horse barn located upslope of the lake may be a point source of phosphorus. In the case of Hacketts and Haynes lakes there are cottages along the shoreline, where even minor sewage seepage might result in eutrophication.

The recognition of the EA in Bond Lake is noteworthy, as the construction of a development corridor north from Richmond Hill across the ORM (Fig. 1) has generated considerable controversy (Diamond et al. 2002). Bond Lake lies within the development corridor and is adjacent to large new housing developments (ESM 1i). For this reason the lake was chosen as a representative lake to study the effects of urban residential development by the Ontario Ministry of the Environment (MOE) (Diamond et al. 2002). Although the lake is eutrophic at present, the degree to which modern suburban development is contributing to phosphorus loading is not known. Analysis of cores from the lake spanning the interval prior to settlement is required to assess baseline conditions.

As discussed in the context of the CRSS and CMES, Haynes Lake is subject to considerable salt contamination so that even samples collected at a distance from the road (e.g., samples 27 and 29) are still influenced by salt inputs (ESM 1a). As a result, the EA samples from the western part of Haynes Lake bear some resemblance to the CRSS (Fig. 3).

Assemblage 5—Eutrophic algal assemblage

This assemblage is generally similar to the EA, differing in that eutrophic algal assemblage (EAA) samples are overwhelmingly dominated by *C. tricuspis* (31.0–78.7%). As expected where one species is so dominant, SDI values are quite low with only a few samples being characterised by more diverse faunas (SDI = 0.56–1.87). The distribution of samples closely mirrors that of the EA and many are from the same lakes. Lakes characterised by EA samples include Bond, Aberlake, Aberlake Farm Pond and Gregloch. EAA samples are closely correlated with each other on the DCA plot, but are quite distinct from the EA distribution (ESM 3). As observed with the EA samples, the EAA correlates closely with phosphorus levels, the primary contributor to lake eutrophication. EAA samples from Aberlake (54 and 59) also plot closely with iron, nickel and chromium on the sample-environment bi-plot (Fig. 4). These metals are constituents of fertilisers and animal waste (Tracy and Baker 2005) and may have been introduced from the surrounding farmland.

Partial canonical correspondence analysis

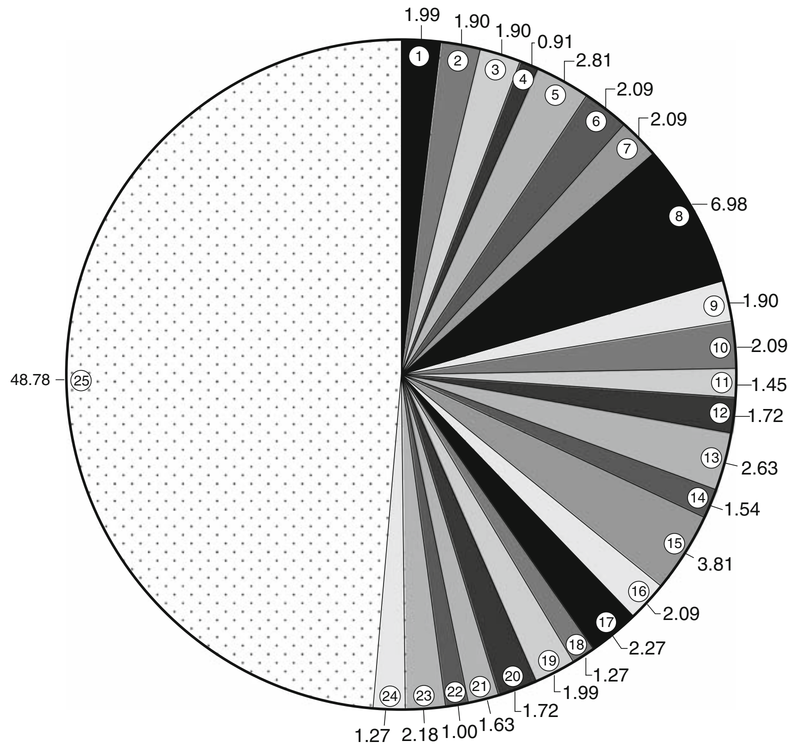
The partial Canonical Correspondence analysis (pCCA) quantifies the proportion of the variance in the thecamoebian dataset that can be attributed to the measured environmental variables (Fig. 6) and confirms that several factors are influencing faunal distribution in the studied lakes. Not surprisingly, the most significant control on thecamoebian distribution is sedimentary phosphorus (Olsen P), which explains 6.98% ($P < 0.002$) of the total variance. This is a significant result which highlights the sensitivity of lake thecamoebians to eutrophication. Conductivity, which we hypothesise has an important influence on the composition of Assemblage 3, explains 2.81% ($P < 0.096$), whilst dissolved oxygen and TOC explain 2.09% ($P < 0.260$) and 1.90% ($P < 0.356$) respectively. Some of the metals, notably Mg (at 3.81%; $P < 0.016$), K (2.27%; $P < 0.214$), Cr (2.18%; $P < 0.284$) and Na (2.09%; $P < 0.268$) also explain similar proportions of the total variance. In the context of this study these metals could have a number of sources. Magnesium, for example, may be sourced from magnesium-rich groundwater percolating through the coarse textured glacial deposits, which are prevalent in the region, or from de-icing salts. Mg and K are also both constituents of fertilisers (Tracy and Baker 2005), whilst Mg is also a common urban contaminant due to its widespread use in ferrous alloys, electrical and industrial products (Morrison and Rauch 2007). The highest values of Mg (>9,000 mg/kg^{-1}) were recorded in the SWM ponds (Wendigo and Dixie Road) and Grenadier Pond, all in urban settings, suggesting that urban inputs of Mg may be influencing faunal composition on a local scale. Pearson correlation analysis confirms that inter-correlations between some of the measured variables are very strong. For example, there is a strong inverse correlation between water depth and temperature ($r = -0.918$; $P < 0.01$) and between water depth and dissolved oxygen ($r = -0.610$; $P < 0.01$). These are expected results, particularly as some of the lakes (e.g., Heart, Teapot, Bond) display a pronounced thermocline. The strong correlation between conductivity and Na ($r = 0.761$; $P < 0.01$) and conductivity and Ca ($r = 0.654$; $P < 0.01$) is also as expected.

Development of a transfer function

The significant taxa-environment result for sedimentary phosphorus (Olsen P) indicated in the CCA and pCCA suggests that a transfer function can be developed for phosphorus based on the GTA lake dataset. To determine whether unimodal or linear-based regression models would be appropriate for this, Detrended Canonical Correspondence Analysis (DCCA) was used to establish the gradient length of the data (Birks 1995; Swindles et al. 2007b, 2009). Unimodal models were selected, as the gradient length was greater than 2σ units (Birks 1995). Transfer function models were developed using weighted averaging (WA), tolerance down-weighted weighted averaging (WA-Tol), weighted averaging partial least squares (WA-PLS) regression and Maximum Likelihood (ML) using the C2 software package (Juggins 2003). The model performance was assessed using the root mean square error of prediction (RMSEP) and the coefficient of determination (r^2) calculated as apparent and leave-one-out cross-validated ('jack-knifed') values (Table 1). The performance of the transfer function models was improved through removal of samples with high residual values.

The best performing model with the lowest RMSEP value for sedimentary phosphorus is WA with inverse deshrinking ($RMSEP_{jack} = 192.59$, $r^2_{jack} = 0.24$) (Table 1). Other, more complex models offer no improvement in model performance (Table 1). Analysis of observed and model estimated variables shows that there are a number of outlier samples with high residual values (Fig. 7). Residual samples at the top and bottom 15% of the total range in the predicted residual data were thus removed to improve model performance, leaving a total of 49 samples in the improved model. The improved transfer function for sedimentary phosphorus is still based on WA with inverse deshrinking and has an $RMSEP_{jack}$ of 102 ppm and $r^2_{jack} = 0.33$ (Table 1).

The tolerance and optima values for the 23 thecamoebian taxa recorded in the dataset are shown in Fig. 8. Taxa showing a particularly wide and high tolerance to P include *Difflugia oblonga* "linearis", *Pontigulasia compressa* Carter 1864 and *Difflugia oblonga* "oblonga", whilst *Centropyxis constricta* "constricta", *Centropyxis constricta* "aerophila" and *Difflugia oblonga* "tenuis" appear to be intolerant to P levels in excess of ~200–250 ppm (Fig. 8). Care must be exercised in interpreting some of these ranges as for several species (e.g., *P. compressa* and *Difflugia bidens* Penard 1902) the number of species occurrences was low (Fig. 3). Not surprisingly, the phosphorus tolerance ranges show some degree of consistency with the ecological inferences drawn from the Q-R mode cluster analyses (Fig. 3) and the CCA species-environment bi-plot (Fig. 4) particularly the observation that *D. oblonga* "oblonga", so



Variable	%	p<
1 Lake area	1.99	0.312
2 Water depth	1.90	0.354
3 Temperature	1.90	0.346
4 pH	0.91	0.916
5 Conductivity	2.81	0.096
6 Redox	2.09	0.234
7 Dissolved oxygen	2.09	0.260
8 P	6.98	0.002
9 T.O.C.	1.90	0.356
10 Clay	2.09	0.294
11 Silt	1.45	0.634
12 Sand	1.72	0.460
13 *Typha*	2.63	0.120
14 Ca	1.54	0.528
15 Mg	3.81	0.016
16 Na	2.09	0.268
17 K	2.27	0.214
18 Fe	1.27	0.744
19 Mn	1.99	0.320
20 Zn	1.72	0.448
21 Cu	1.63	0.518
22 Pb	1.00	0.850
23 Cr	2.18	0.284
24 Ni	1.27	0.740
25 Intercorrelations	48.78	n/a

Fig. 6 Partial canonical correspondence analyses (pCCA) results showing the percentage variance in the thecamoebian dataset explained by the measured environmental variables and intercorrelations. *P* values for each variable are shown

dominant in Assemblage 4, is a significant indicator taxon for eutrophication.

The usefulness of calibration models between fauna and environmental variables is best measured by their predictive ability. To test the utility of the transfer function, the WA with inverse deshrinking model was applied to a small contemporary thecamoebian dataset (three samples) collected from the pH 6.7–6.9, northern basin of James Lake in NE Ontario (Patterson and Kumar 2000a, b). The James Lake phosphorus analyses were carried out on pore waters derived from sediment–water interface samples by colorometric analysis of an acid digestion (R. E. A Boudreau, unpublished data). Whilst this method differs from the Olsen P extraction used in the present study, this pore-water extraction method was chosen because it similarly provides biologically available phosphorus (Wetzel 2001) and are thus likely to be broadly compatible. In applying the model sample-specific prediction errors were generated by 1,000 bootstrap cycles.

The results of the application of the WA-Inv transfer function to the James Lake thecomoebian dataset are presented in ESM 5. The results show that the observed sedimentary P measurements lie within the boostrap error range of the model-inferred P value. This preliminary result highlights the potential of this transfer function for prediction of sedimentary phosphorus in eutrophic lakes in this region and for future application in down-core thecamoebian studies.

Discussion

This study set out to elucidate the controls on contemporary lake thecamoebian distribution within the GTA and to examine the relationships between land use, faunal assemblages and lake sediment and water property data. Kettle lakes were primarily selected because they act as a sink for pollutants and are a particularly threatened habitat within the GTA. The resulting faunas are generally of low to moderate diversity (SDI values 1.00–2.00) and are dominated by centropyxids, which are particularly tolerant of sub-optimal environmental conditions (Patterson and Kumar 2000b).

Table 1 Thecamoebian-sedimentary phosphorus transfer function performance statistics

Total dataset (all samples: $n = 67$)	WA Inv	WA Cla	WA-Tol (Inv)	WA-Tol (Cla)	WAPLS 1	WAPLS 2	ML
Phosphorus (ppm)							
RMSE	*168.07*	261.14	169.75	267.58	168.07	157.58	204.58
r^2	*0.41*	0.41	0.40	0.40	0.41	0.49	0.51
Average bias	*0.00*	0.00	0.00	0.00	0.00	0.00	−38.65
Maximum bias	*613.54*	554.62	618.35	792.12	613.54	512.82	397.96
Jack r^2	*0.24*	0.29	0.15	0.19	0.24	0.26	0.34
Jack average bias	*−7.86*	−21.61	−6.07	−18.37	−7.86	−3.24	−26.66
Jack maximum bias	*715.12*	625.97	745.95	746.10	715.12	626.66	433.11
RMSEP	*192.59*	278.15	207.08	303.77	192.59	194.41	217.74

Dataset after removal of residuals ($n = 49$)	WA Inv	WA Cla	WA-Tol (Inv)	WA-Tol (Cla)
Phosphorus (ppm)				
RMSE	*79.98*	104.18	79.05	102.15
r^2	*0.59*	0.59	0.60	0.60
Average bias	*0.00*	0.00	0.00	0.00
Maximum bias	*238.22*	78.58	228.33	68.77
Jack r^2	*0.33*	0.40	0.27	0.33
Jack average bias	*−6.52*	−12.51	−6.57	−12.39
Jack maximum bias	*354.56*	180.64	344.89	171.50
RMSEP	*102.65*	115.22	107.74	120.53

WA Inv weighted averaging with inverse deshrinking, *WA Cla* weighted averaging with classical deshrinking, *WA-Tol Inv* weighted averaging-tolerance downweighted with inverse deshrinking, *WA-tol Cla* weighted averaging-tolerance downweighted with classical deshrinking, *WAPLS* weighted averaging partial least squares (with component number), *ML* maximum likelihood

The performance statistics are shown as apparent and jackknifed (Jack) measures

RMSEP root mean square error of prediction. Values for the best performing model (*WA Inv*) are italicised

Canonical Correspondence Analysis (CCA) and pCCA analysis have confirmed that thecamoebians are responding to a number of environmental variables in the studied lakes, the most important of which appear to be related to the trophic status of the lakes as measured by levels of bioavailable phosphorus (Olsen P) in the surficial sediments. The observed Olsen P levels are particularly high for lake sediments (a mean value of 276 ppm was recorded for the 71 studied samples) and confirm that all the lakes are eutrophic (Zhou et al. 2001). The presence of *C. tricuspis* in nearly all samples, a species commonly associated with floating algae, supports this interpretation.

In terms of land use, the study has shown that the 'healthiest' thecamoebian faunas are generally associated with lakes that lie beyond the current limits of or at the margins of urban development, with the highest SDI values (>2) being recorded in some of the samples from Hacketts, Haynes and Bond Lake (Fig. 1). Several lakes show some degree of intra-lake faunal variability that can be linked with local changes in dissolved oxygen (with the most diverse faunas being recovered from above the thermocline), TOC or substrate changes, or localised inputs of contaminants (particularly salts) from around the lake margins. The influence of salt inputs was particularly evident at Haynes Lake and Grenadier Pond where samples collected in close proximity to roads treated with winter de-icing salts included greater numbers of brackish species, e.g., *Centropyxis aculeata* strains. This finding underlines the potential utility of thecamoebians as bio-indicators of salt contamination in this region where road-salt has been identified as a major contributor to groundwater contamination (Williams et al. 1999). Further work is now required

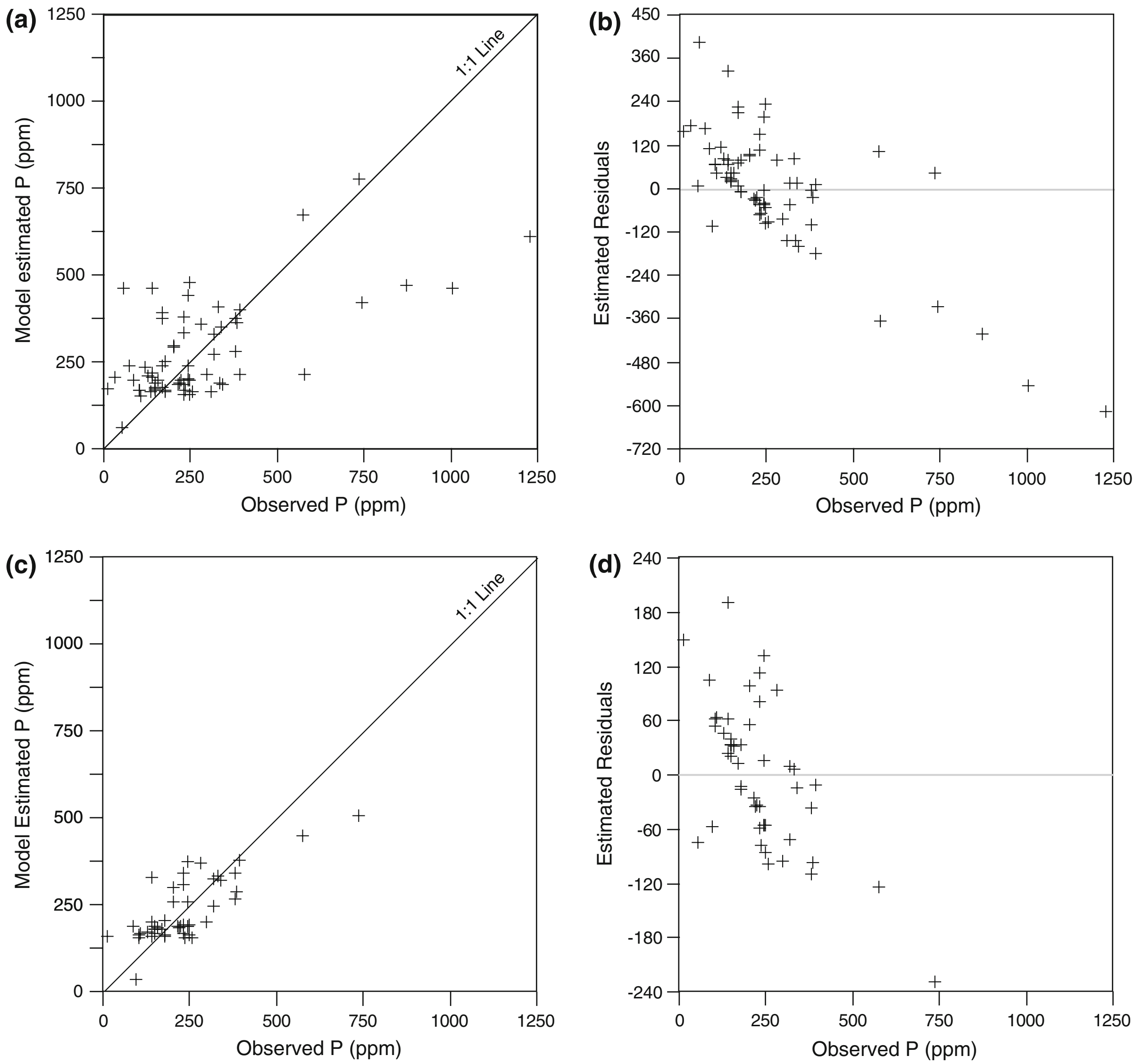


Fig. 7 Observed vs model estimated sedimentary phosphorus; **a** original data; **b** original data residuals; **c** dataset after removal of outliers (improved model); **d** residuals of improved model

to examine the spatial and temporal response of thecamoebians to salt inputs in other regional lakes, particularly in aquifer-sensitive areas. Fossil thecamoebian datasets collected from selected lakes would also provide insights into baseline conditions prior to salt application and thus aid monitoring studies.

There is limited evidence to confirm that the presence of a vegetation 'buffer' around some of the lakes may afford the lakes some degree of protection from urban contaminants as suggested by numerous aquatic biodiversity and environmental management studies (Norman 1996). The results from Teapot Lake (lake 1), Heart Lake Road Pond (lake 13) and Christie Pond (lake 20) exemplify this (Fig. 1). These small kettle lakes lie within 2 km of each other, yet the former is protected by a >100 m fringe of woodland, whilst Heart Lake Road Pond is less protected and lies closer to a highway (ESM 1f). Christie Pond is surrounded by residential land with a 20 m buffer of shrubs (ESM 1h). Four samples from Teapot Lake yielded moderately diverse faunas (SDI = 1.44–

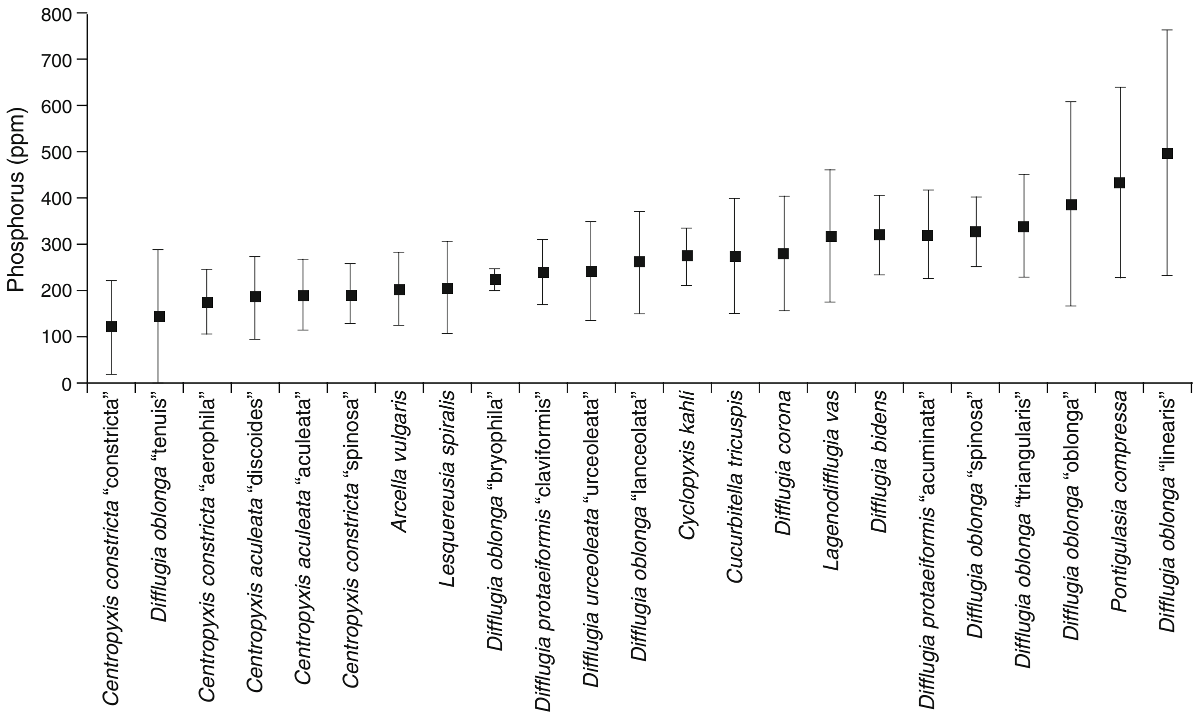


Fig. 8 Phosphorus tolerance and optima statistics for the 23 thecamoebian species encountered in the study

1.83), whilst single samples from the latter two lakes failed to yield statistically significant assemblages and were characterised by lower diversities (SDI = 0.59 and 1.28, respectively). This preliminary result suggests that recent provincial legislation, the Oak Ridges Moraine Protection Act, 2001, which stipulates that a 30 m 'Minimum Vegetation Zone' be maintained around kettle lakes in future development areas (ORMCP Technical Paper 12 2002) may be insufficient for protecting ecological diversity in these hydrologically sensitive lakes. Further evaluation of the influence of vegetation buffers on thecamoebian distribution is required to test this more rigorously.

For other lakes the relationship between land use, environmental variables and thecamoebian communities is less apparent from this preliminary study. This outcome is not surprising given that many of the measured contaminants, particularly phosphorus and some of the heavy metals, occur widely as non-point source pollutants within the region and are not confined to urban areas. Sediment-based Olsen P levels were found to be surprisingly high even in lakes not expected to be significantly influenced by phosphorus inputs from recognised sources; for example, storm drains carrying domestic runoff, or agricultural inputs (livestock or fertiliser). Thus, whilst most of the studied lakes show some degree of eutrophication at present, the degree to which modern urban development is contributing to phosphorus loading is not fully clear. Application of the P-based transfer function developed in this study to fossil thecamoebian datasets from the area may provide a stronger basis for discriminating between P inputs from urban and other (e.g., agricultural) sources, particularly when combined with historical records of land use change. Core-based studies would again enable baseline conditions prior to anthropogenic disturbance to be determined which would in turn allow water quality restoration initiatives to be more fully evaluated.

Given the strong relationship between thecamoebian communities and Olsen P, future work might also consider the response of thecamoebians to other geochemical lake trophic status indicators, for example total nitrogen (TN), nitrates and ammonium which were not measured in this study. Measurements of total phosphorus (TP) might also be collected and compared with the Olsen P data to allow the training set to be developed and applied to oligiotrophic lakes.

Samples from the small set of SWM management ponds included in the study were generally characterised by impoverished thecamoebian faunas or were barren. In the case of Wendigo Pond, this appears to reflect elevated levels of Mg, Zn and Pb in the surface sediments, although other trace metals (e.g., arsenic, vanadium or selenium), or organic pollutants (e.g., PAHs), which were not measured in the study, might also be influential. This result warrants further investigation. The sparse nature of some of the SWM pond faunas may also reflect high-suspended sediment inputs into the lakes as a result of recent housing developments. The recently constructed Jefferson SWM pond (ESM 1c) and Dixie Road SWM pond (Fig. 1) were both characterised by low TOC levels and were observed to include high levels of suspended solids, although the latter were not measured.

Conclusions

This study has provided new insights into the controls on the distribution of lake thecamoebians in the rapidly urbanising GTA and has resulted in the generation of a training set, which, with further development, will be applicable in future studies based on fossil thecamoebian datasets from the region and elsewhere. Of 71 surface sediment samples analysed from 11 small lakes, a large urban lake and nine constructed ponds, 58 samples yielded statistically significant thecamoebian populations dominated by seven key species and strains. Five faunal assemblages identified via Q-mode cluster analysis have been linked with a number of driving variables, including sedimentary phosphorus inputs from both urban and agricultural sources and the application of road salt from winter de-icing operations. All the studied lakes showed evidence of eutrophication and/or stressed conditions.

Ordination (CCA) analysis has confirmed that 24 measured environmental variables explain 64% of the variance in the thecamoebian dataset. pCCA analysis has further confirmed that sedimentary phosphorus (Olsen P) has the largest influence on assemblage variance, explaining 6.98% ($P < 0.002$). This result underlines the sensitivity of thecamoebians to eutrophication. Other factors influencing faunal distribution are conductivity, which explains 2.81% ($P < 0.096$) of the total variance, Mg 3.81% ($P < 0.016$) and TOC 1.90% ($P < 0.356$).

Whilst lakes beyond the limits of urban development generally produced the healthiest thecamoebian faunas (=highest SDI values), the relationship between land use and thecamoebians was found to be complex, reflecting the widespread nature of some of the measured contaminants and the unexpectedly high sediment-based P levels in both urbanised and rural areas of the region. Several lakes also showed intra-lake assemblage variability, which can be linked with local inputs of salts around the margins, or the presence of a well-developed thermocline. Such variability must be taken into account when considering fossil datasets from cores.

The transfer function developed for sedimentary phosphorus (Olsen P) was based on a training set of 58 samples from 15 of the studied lakes. The best performing model was developed via weighted averaging with inverse deshrinking ($RMSEP_{jack} = 102.65$ ppm, $r^2_{jack} = 0.33$). Further work is now required to refine this model for future application to fossil thecamoebian datasets from the region and elsewhere to better elucidate temporal and spatial changes in phosphorus inputs and to evaluate sources.

The preliminary data presented here suggests that thecamoebians hold considerable potential as water quality indicators in urban and agricultural lake environments, particularly for examining (1) changes in lake trophic status driven by fluctuations in phosphorus; and (2) salt contamination. Additional distributional work is required across both of these environmental gradients to further develop training sets for future application to down-core studies.

Acknowledgments This research was funded by a grant to HMR from the The Foundation for Canadian Studies in the United Kingdom and the Dept. of Foreign Affairs and International Trade (Sustained Studies in Contemporary Canadian Issues Program 2003–05) as well as a Canadian Natural Sciences and Engineering Research Council Discovery Grant and funding from the Toronto and Region Conservation Authority to RTP. We thank Miriam Diamond (University of Toronto), Bernie Smith (Queen's University Belfast), Stephen Maude (Ontario Ministry of the Environment) and Don Ford (Toronto and Region Conservation Authority) for valuable discussions on aspects of lake pollution, conservation and management and John McAlister (Queen's University, Belfast) and Bob Boudreau for help with the geochemical analyses and interpretations. We also thank Davin Carter for assistance with fieldwork, Julia Simpson and Yoma Megarry for help with laboratory procedures and Gill Alexander for cartographic support.

References

Asioli A, Medioli FS, Patterson RT (1996) Thecamoebians as a tool for reconstruction of paleoenvironments in some Italian lakes in the foothills of the southern Alps (Orta, Varese and Candia). J Foram Res 26:248–263

Birks HJB (1995) Quantitative palaeoenvironmental reconstruction. In: Maddy D, Brew JS (eds) Statistical modelling of quaternary science data. Quat Res Assoc, London, pp 161–251

Boller MA, Steiner M (2002) Diffuse emission and control of copper in urban surface runoff. Water Sci Technol 46(6–7):173–181

Booth RK, Jackson ST (2003) A high-resolution record of Late Holocene moisture variability from a Michigan raised bog. Holocene 13:865–878

Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. Ecology 73:1045–1055

Bradford A, Maude S (2002) Achieving ecological integrity on the Oak Ridges Moraine: towards ecologically relevant water management. Unpublished report, University of Guelph, 5 p

Callender E, Rice KC (2000) The urban environmental gradient: anthropogenic influences on the spatial and temporal distributions of lead and zinc in sediments. Environ Sci Technol 34(2):232–238

Charman DJ, Roe HM, Gehrels WR (1998) The use of testate amoebae in studies of sea-level change: a case study from the Taf Estuary, south Wales, UK. Holocene 8:209–218

Charman DJ, Hendon D, Woodland W (2000) The identification of peatland testate amoebae. Quaternary Research Association Technical Guide No. 9, London, 147 pp

Collins ES, McCarthy FM, Medioli FS, Scott DB, Honig CA (1990) Biogeographic distribution of modern thecamoebians in a transect along the eastern North American coast. In: Hemleben C, Kaminski MA, Kuhnt W, Scott DB (eds) Paleoecology, biostratigraphy, paleoceanography and taxonomy of agglutinated foraminifera. NATO Advanced Study Institute Series, Series C, Mathematical and Physical Sciences, vol 327, pp 783–791

Dale B, Dale AL (2002) Application of ecologically based statistical treatments to micropalaeontology. In: Haslett SK (ed) Quaternary environmental micropalaeontology. Arnold, London

Diamond M, Dougan J, Helferty N, Hodge E, Niblett P, Rose M, Rowe S (2002) Natural heritage systems in urbanizing settings: sustainable practices for the Oak Ridges Moraine. Report prepared for Save the Rouge Valley System and City of Toronto, 109 pp

Escobar J, Brenner M, Whitmore TJ, Kenny WF, Curtis JH (2008) Ecology of testate amoebae (thecamoebians) in subtropical Florida lakes. J Paleolim 40(2):715–731

Fishbein E, Patterson RT (1993) Error-weighted maximum likelihood (EWML): a new statistically based method to cluster quantitative micropaleontological data. J Paleontol 67:475–485

Gehrels WR, Roe HM, Charman DJ (2001) Foraminifera, testate amoebae and diatoms as sea level indicators in UK saltmarshes: a quantitative multi-proxy approach. J Quat Sci 16(3):201–220

Hesse PR (1971) A textbook of soil chemical analysis. John Murray, London 520 pp

Juggins S (2003) C2 user guide. Software for ecological and palaeoecological data analysis and visualisation. University of Newcastle, Newcastle Upon Tyne

Kerekes JJ, Blouin AC, Beauchamp ST (2004) Trophic response to phosphorus in acidic and non-acidic lakes in Nova Scotia, Canada. Hydrobiologia 191:105–110

Kumar A, Dalby AP (1998) Identification key for Holocene lacustrine arcellacean (thecamoebian) taxa. Paleontologia Electronica 1:34 p. On-line: http://palaeo-electronica.org

Legendre P, Gallagher E (2001) Ecologically meaningful transformations for ordination of species data. Oecologia 129:271–280

Legendre P, Legendre L (1998) Numerical ecology: developments in environmental modeling 20. Elsevier Science B.V., Amsterdam 853 pp

Link DD, Walter PJ, Kingston HM (1998) Development and validation of the new EPA microwave-assisted leach method 3051A. Environ Sci Technol 32:3628–3632

Magurran AE (1988) Ecological diversity and its measurement. Princeton University Press, Princeton 179 pp

Marsalek J, Shroeter H (1988) Annual loadings of toxic contaminants in urban runoff from the Canadian Great Lakes basin. Water Qual Res J Can 23(3):360–378

Medioli FS, Scott DB (1983) Holocene Arcellacea (Thecamoebians) from eastern Canada. Cushman foundation for foraminiferal research special publication, vol 21, 63 pp

Medioli FS, Scott DB, Abbott BH (1987) A case study of protozoan interclonal variability: taxonomic implications. J Foram Res 17:28–47

Medioli FS, Scott DB, Collins ES, Wall JH (1990) Fossil thecamoebians: present status and prospects for the future. In: Hemleben C, Kaminski MA, Kuhnt W, Scott DB (eds) Paleoecology, biostratigraphy, paleoceanography and taxonomy of agglutinated foraminifera. NATO Advanced Study Institute Series, Series C, Mathematical and Physical Sciences, vol 327, pp 813–840

Morrison GM, Rauch S (2007) (eds) Highway and urban environment. Proceedings of the 8th urban environment symposium series, vol 12, alliance for global sustainability bookseries, Springer, The Netherlands, pp 1571–4780

Norman AJ (1996) The use of vegetative buffer strips to protect wetlands in southern Ontario. In: Mulamoottil G, Warner BG, McBean EA (eds) Wetlands. Environmental gradients, boundaries and buffers. CRC Press, New York, pp 263–275

ORMCP—Oak Ridges Moraine Conservation Plan (2002) Technical paper series. 12. Hydrological evaluations for hydrologically sensitive features. Ontario Ministry of Municipal Affairs and Housing

Patterson RT, Fishbein E (1989) Re-examination of the statistical methods used to determine the number of point counts needed for micropaleontological quantitative research. J Paleontol 63:245–248

Patterson RT, Kumar A (2000a) Use of arcellacea to gauge levels of pollution and remediation of industrially polluted lakes. In: Martin RE (ed) Environmental micropaleontology, topics in geobiology, vol 15. Kluwer Academic/Plenum Publication, New York, pp 257–278

Patterson RT, Kumar A (2000b) Assessment of arcellacea (thecamoebian) assemblages, species and strains as contaminant indicators in variably contaminated James Lake, north eastern Ontario. J Foram Res 30:310–320

Patterson RT, Kumar A (2002) A review of current testate rhizopod (thecamoebian) research in Canada. Palaeogeogr Palaeoclim Palaeoecol 180:225–251

Patterson RT, Dalby A, Kumar A, Henderson LA (2002) Arcellaceans as indicators of land use change: settlement history of the Swan Lake area, Ontario as a case study. J Paleolimnol 28:297–316

Reinhardt ED, Dalby A, Kumar A, Patterson RT (1998) Utility of arcellacean morphotypic variants as pollution indicators in mine tailing contaminated lakes near Cobalt, Ontario, Canada. Micropaleontology 44:1–18

Reinhardt EG, Little M, Donato S, Findlay D, Krueger A, Clark C, Boyce J (2005) Arcellacean (thecamoebians) evidence of land-use change and eutrophication in Frenchman's Bay, Pickering, Ontario. Environ Geol 47:729–739

Roe HM, Patterson RT (2006) Distribution of thecamoebians (testate amoebae) in small lakes and ponds, Barbados, West Indies. J Foram Res 36:116–134

Roe HM, Charman DJ, Gehrels WR (2002) Fossil testate amoebae in coastal deposits in the UK: implications for studies of sea-level change. J Quat Sci 17(5–6):411–429

Schönborn W (1984) Studies on remains of Testacea in cores of the Great Woryty Lake (N. E. Poland). Limnologica 16:185–190

Schönborn W (1990) Shell polymorphism and habitat structure in testacea (Rhizopoda). J Protozool 37:62A

Schroeder HA, Dobson M, Kane DM (1987) Toxic trace elements associated with airborne particulate matter: a review. J Air Pollut Control Assoc 37:1267–1285

Scott DB, Medioli FS (1980) Post-glacial emergence curves in the Maritimes determined from marine sediments in raised basins. Proceedings of coastlines 1980, published by National Science and Engineering Research Council, pp 428–449

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423

Skinner L, de Peyster A, Schiff K (1999) Development effects of urban storm water in Medaka (*Oryzias latipes*) and Inland Silverside (*Mednidia beryllina*). Arch Environ Contam Toxicol 37:227–235

Swindles GT, Roe HM (2007) Examining the dissolution characteristics of testate amoebae (Protozoa: Rhizopoda) in low pH conditions: implications for peatland palaeoclimate studies. Palaeogeogr Paleoclim Paleoecol 252:486–496

Swindles GT, Plunkett G, Roe HM (2007a) A delayed climatic response to solar forcing at 2800 cal. BP: multi-proxy evidence from three Irish peatlands. Holocene 17:177–182

Swindles GT, Plunkett G, Roe HM (2007b) A multi-proxy climate record from a raised bog in County Fermanagh, Northern Ireland: a critical examination of the link between bog surface wetness and solar variability. J Quat Sci 22:667–679

Swindles GT, Charman DJ, Roe HM, Sansum PA (2009) Environmental controls on peatland testate amoebae (Protozoa: Rhizopoda) in the North of Ireland: implications for paleoclimate studies. J Paleolimnol 42:123–140

ter Braak CJF (2002) CANOCO for Windows version 4.51. Biometris-Plant Research International, Wageningen

ter Braak CJF, Šmilauer P (2002) CANOCO reference manual and CanoDraw for Windows user's guide: software for canonical community ordination (version 4.5). Microcomputer Power, Ithaca, p 500

Tracy D, Baker B (2005) Heavy metals in fertilizers used in organic production. Unpublished report, Organic Materials Review Institute, USA, 24 pp

TRCA (Toronto and Region Conservation Authority) (2008) Humber River state of the watershed report—surface water quality, 38 p

Watanabe FS, Olsen SR (1965) Test of an ascorbic acid method for determining phosphorus in water and $NaHCO_3$ extracts of soil. Soil Sci Soc Am Proc 29:677–678

Watchorn MA, Hamilton PB, Anderson T, Roe HM, Patterson RT (2008) Diatoms and pollen as indicators of water quality and land use change: a case study from the Oak Ridges Moraine, Southern Ontario. J Paleolimnol 39:491–509

Wetzel RG (2001) Limnology: lake and river ecosystems. Academic Press, San Diego, 1006 p

Williams DD, Williams NE, Cao Y (1999) Road salt contamination of groundwater in a major metropolitan area and development of a biological index to monitor its impact. Water Res 34(1):127–138

Zhou Q, Gibson CE, Zhu Y (2001) Evaluation of phosphorus bioavailability in sediments of three contrasting lakes in China and the UK. Chemosphere 42:221–225