**3 | RESULTS**

Our main goal here was not to investigate how prevalence of black spot disease in fish communities is perceived across three scale levels (landscape, lake, site). First, we built landscape-level infection prevalence accumulation curves, then we compared the frequency distribution of the lake’s prevalence for each sampling method, and lastly, we identified predictors of site-scale prevalence patterns.

We recorded a total of 11 297 individual fishes belonging to 15 species for this study (Table S4). The minnow traps caught 1906 individuals from 10 species while seine nets caught 2427 individuals from 14 species (Table S5 and S6). 6964 individuals belonging to five taxonomic groups (four species and one family) were observed in the snorkeling transects (Table S7). The mean length of all fish captured through fishing methods was 5.59 ± 2.96 cm (N = 4333). The overall mean length of each lake, species and, species within each lake are presented in Table S8, S9 and S10 respectively.

All fish species sampled, except for *Ameiurus nebulosus* and *Esox masquinongy* are susceptible to black spot disease (see Table S11). However, our survey only found evidence of black spot infection in pumpkinseed sunfish (*Lepomis gibbosus*), rock bass (*Ambloplites rupestris*), yellow perch (*Perca flavescens*), smallmouth bass (*Micropterus dolomieu*), creek chub (*Semotilus atromaculatus*) and fathead minnow (*Pimephales promelas*). Details on infection prevalence of fish communities are presented for every sampling method in Table S12 (landscape-scale), Table S13 (lake-scale) and Table S14 (site-scale).

***3.1. Landscape-scale***

We used a resampling approach to compare how different sampling methods change the mean prevalence estimate through an increasing sampling effort across the landscape (Figure 7).

Each method curves overestimated the mean prevalence at low N values (number of samples) suggesting that infection prevalence is heterogeneously distributed in our landscape with presence of infection clusters or hotspots. The estimate of landscape prevalence differed between the sampling methods, varying between 21% and 36%. After 35 samples, the transect method generated the highest mean prevalence (36 %) followed by the methods combination (31 %), the minnow trap method (24 %) and the seine method (21 %) (see Table S15). However, the minnow trap curve did not stabilize after the 35 random samples suggesting that the prevalence value obtained is higher than the actual estimate (landscape infection prevalence measured by the minnow trap is 19%, see Table S12). However, relatively few random samples are necessary to estimate a landscape prevalence for the transect method (approximately 10 samples) while an accurate prevalence estimate only occurs after 30 samples for the seine method. Even if the method combination curve (in gray) stabilizes around 20 samples, it nevertheless displays some variation around the curve because of the variability among the different methods used.

***3.2. Lake-scale***

We compared the frequency distribution of the lake’s prevalence for each sampling method to see how the landscape prevalence is distributed among lakes and in their geographical context to identify spatial patterns (Figure 6).

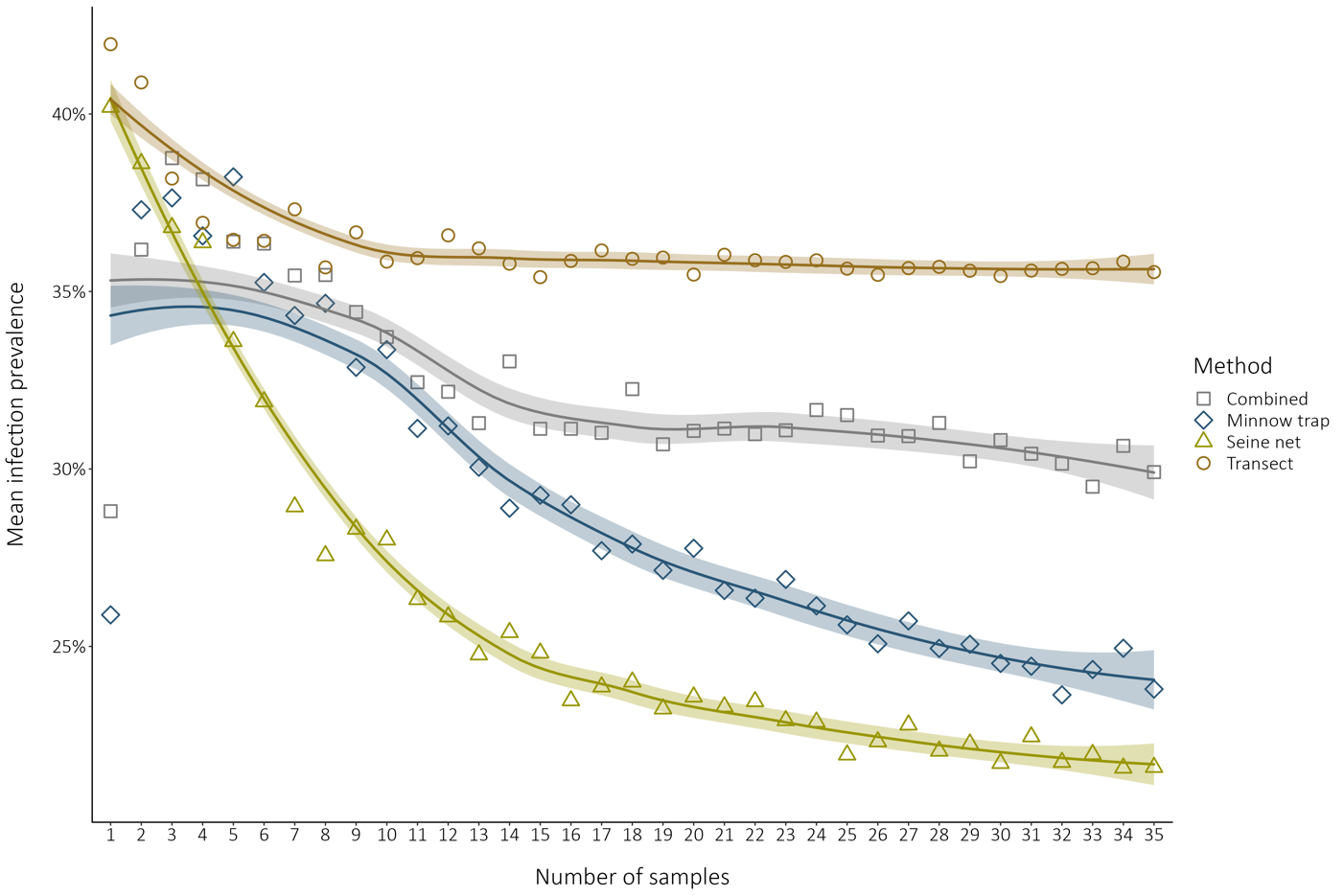
Following the landscape results, the lake fish community prevalence estimates are not constant across the landscape. Both frequency distributions and prevalence maps support a difference between methods in terms of sampling fish communities and/or behavior. The lake prevalence frequency distributions for the combined methods and for the transect method show a bimodal distribution, with the landscape composed of many low-prevalence and high-prevalence lakes. However, neither method yielded prevalence estimates over 80%. The distribution patterns for the two fishing methods (seine net and minnow trap) are less clear. These methods show a right-skew distribution representing more low-prevalence lakes. Accordingly, seine and minnow traps captured very similar prevalence estimates at the landscape-scale (20.4% and 19.2% respectively). Heavily infected and less infected lakes do not appear to be clustered in space at the regional scale (see maps in Figure 6). Moreover, close and connected lakes do not appear to follow a spatial infection gradient, suggesting that geographic attributes (e.g., position in the landscape, connectivity, distance to nearest lake, belonging watershed) are not important drivers of the local infection prevalence. Nevertheless, lake-scale geographic patterns were not statistically tested. Both frequency distributions and prevalence maps support a difference between methods in terms of sampling fish communities and/or behavior.

***3.2. Site-scale***

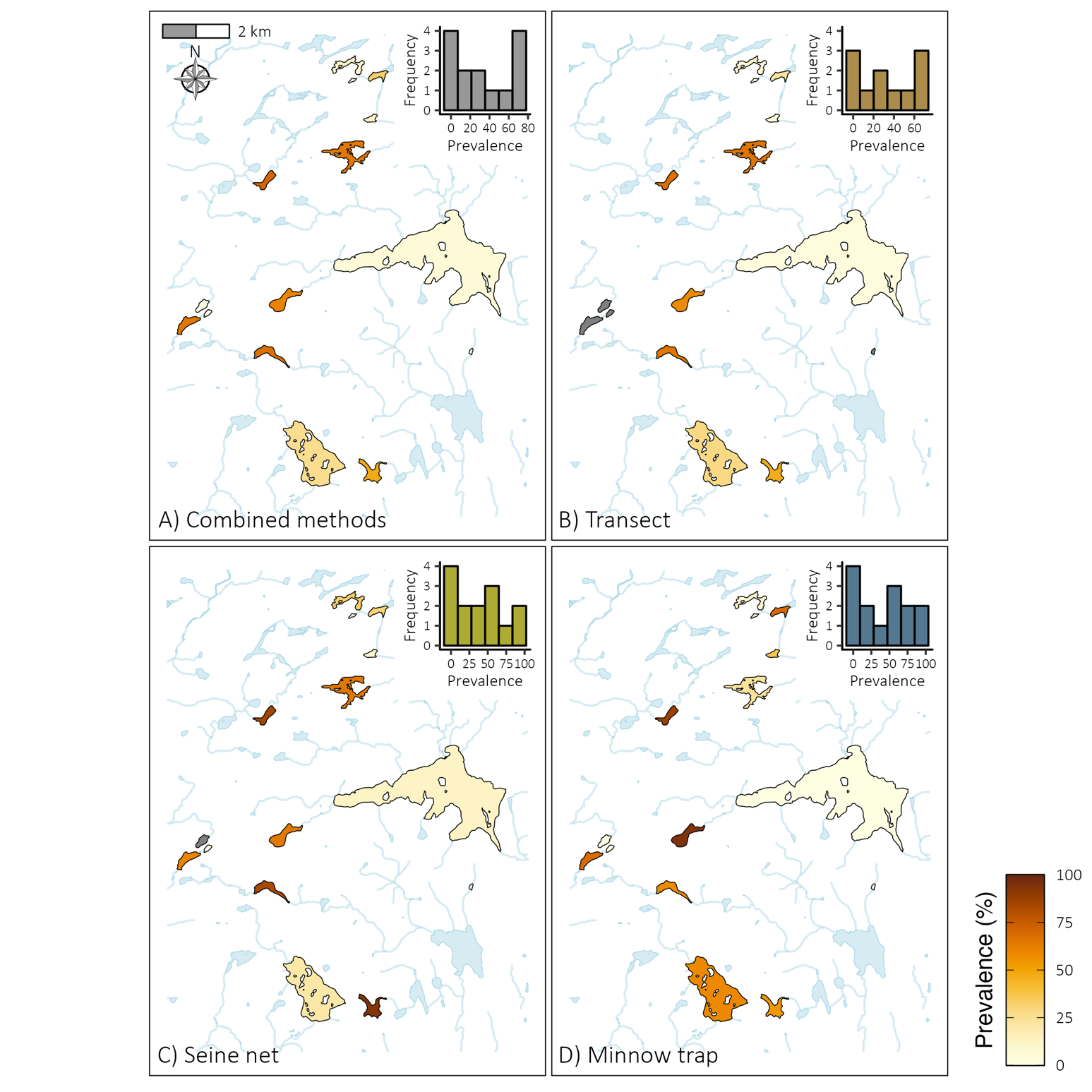
The relationships between the potential predictors and the site-scale prevalence (transect-level prevalence) were assessed with generalized additive mixed effects models (see Table S16 for models’ estimates). Environmental characteristics of transect sites used as model predictors are presented in Table S17.

The partial effects of environmental variables on prevalence are shown for all significant models in Figure 8. The turbidity model has the best fit (D2 = 88,71%) and is mostly non-linear. The relation evidence a plateau for high-prevalence values, indicating that a prevalence saturation is reached for turbidity values above 2 NTU. That said, this model must be carefully interpreted as we sampled only a few high-turbidity sites. Water temperature has an increasingly proportional relationship with infection prevalence. The prevalence estimates also increased proportionally with pH, so that prevalence increases in more alkaline lakes. The conductivity and prevalence have a non-linear, unimodal relationship peaking around 80 (μS/cm). The relationship must also be carefully interpreted because of some gaps in the conductivity values (between 61.3 and 129.2 μS/cm). Prevalence values increases almost proportionally with dissolved oxygen concentration in the water. However, we must take into consideration that the variation interval is very large for low-concentration oxygen values because of only three values below 7 mg/L were recorded. The relationship between TN:TP ratio and prevalence is unimodal but highly non-linear. For the macrophyte coverage, we found a decreasing relationship between the amount of macrophyte cover and the prevalence of infection, meaning that low macrophyte cover correlates with high prevalence of site-scale communities. A parabolic curve is observed in the Area:Perimeter model although, high ratio values are more uncommon increasing the variation interval. The perimeter model shows a highly non-linear presenting two distinctive peaks and a large confidence interval at higher perimeter values. Total fish abundance and non-host abundance (here abundance of cyprinids) both showed a negative relationship with prevalence suggesting that sites with more fishes and/or non-host individuals have lower prevalence of infection. The relationship between the prevalence and the species diversity index shows a decreasing trend. At the site-scale, fish communities are slightly to moderately diverse (Simpson’s diversity index between 0 and 0.64) indicating the dominance of some species. Since cyprinids were not differentiated, it is hard to unravel who is the most dominant species based on the transect method data.

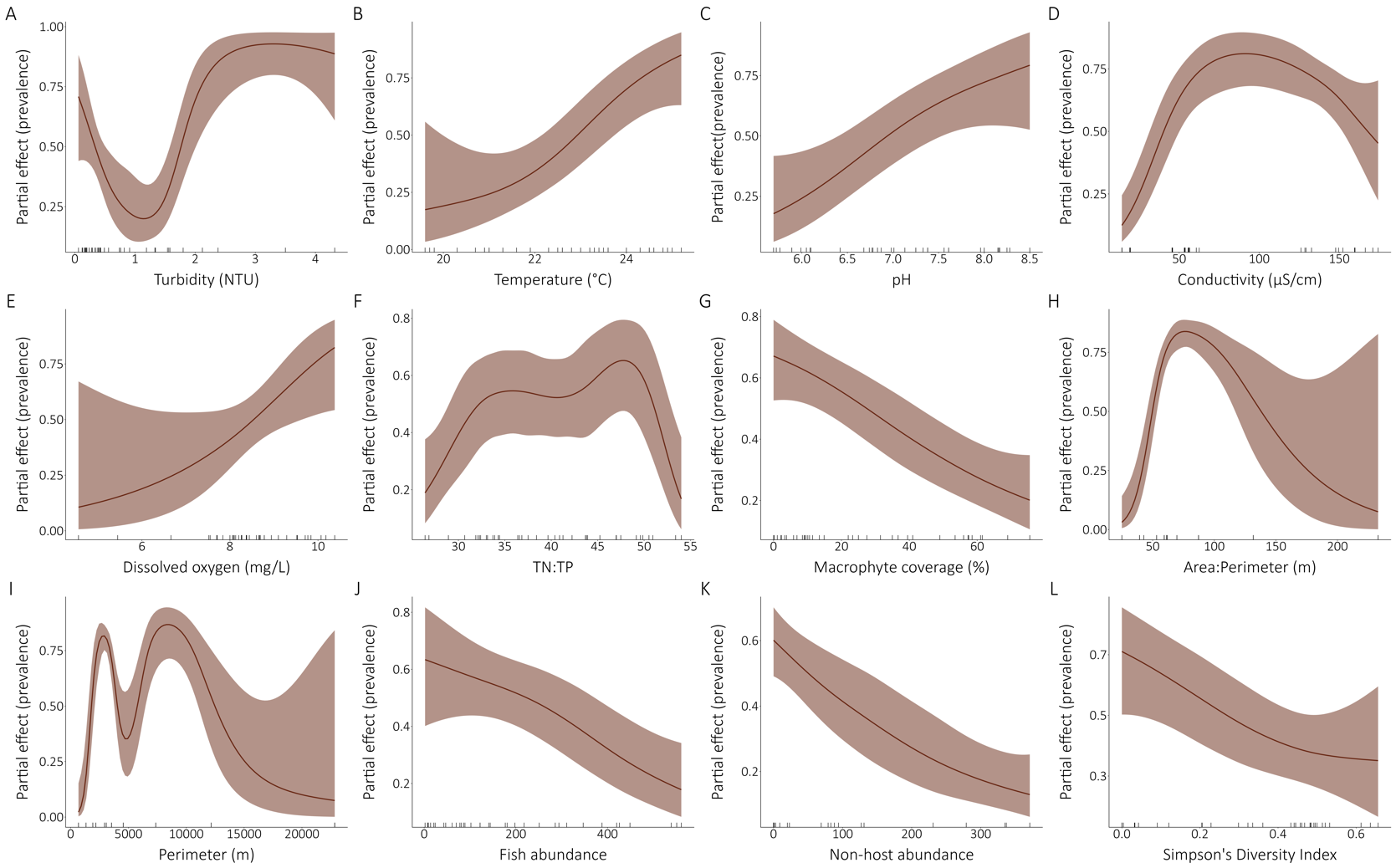
The models for TN, TP, TOC, type de substrate (silt, sand, rock and boulder), mean transect depth, number of trunks, lake surface area, lake maximum depth, lake mean depth, water residence time, drainage area, lake elevation, distance to the nearest lake and species richness were not significant (see Table S18). We did not find any relationships between black spot prevalence and any spatial attributes. Water physico-chemistry and community metrics were good predictors of infection prevalence of fish communities in our sites.



**Figure 7.** Method comparison of landscape prevalence estimates through an increasing random sampling effort. The data points indicate the mean prevalence for a given number of samples (N) randomly selected. The shadings indicate the 95% confidence interval on the loess smooth. The pool data used for the resampling with the minnow traps, the seine nets, the transects and all methods combined respectively contained 225, 75, 39 and 339 samples.

******

**Figure 6.** Method comparison of the lakes’ local community prevalence estimates across the landscape. Frequency distributions are given in the top right corner of each map. Lakes in grey were not sampled except for lake Tracy that was excluded because of insufficient data (one fish sampled through all methods). **(A)** All the sampling methods combined (weighted mean regional prevalence is 29.55 %). **(B)** The transect method (weighted mean regional prevalence is 35.55 %). **(C)** The seine nets (weighted mean regional prevalence is 20.44 %). **(D)** The minnow traps (weighted mean regional prevalence is 19.20 %).

******

**Figure 8.** Relationships between the site-scale community infection prevalence and selected environmental drivers. All the models presented are univariate binomial generalized additive mixed models with the lake as a random effect on the intercept. The partial effects of the environmental variables on the prevalence are presented for the significant models. The ticks on the x-axes indicate a data point. **(A)** Turbidity. **(B)** Temperature. **(C)** pH. **(D)** Conductivity. **(E)** Dissolved oxygen. **(F)** TN:TP ratio. **(G)** Macrophyte coverage. **(H)** Area:Perimeter ratio. **(I)** Perimeter. **(J)** Fish abundance. **(K)** Non-host abundance. **(L)** Simpson’s Diversity Index. Area:Perimeter ration and Perimeter are lake attributes while other parameter are site-scale measurements.