LOCAL ENVIRONMENT AND SAMPLING BIAS DRIVE PARASITE PREVALENCE ESTIMATES IN FISH COMMUNITIES

Juliane Vigneault[1,2] | Sandra Ann Binning[1,2] | Éric Harvey[1,2,3]

1 Département de sciences biologiques, Université de Montréal, QC, Canada
2 Groupe de recherche interuniversitaire en limnologie (GRIL), département de sciences biologiques, Université de Montréal, QC, Canada
[3] Département des sciences de l'environnement, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada.

Abstract

Parasite distribution patterns and drivers of infection vary in time and space, making understanding the underlying mechanisms highly complex. Comparative studies based on empirical data must therefore consider the factors of variation involved in estimating infection parameters in natural populations. Using a multi-scale approach, we explored the sources of variation in the estimation of infection prevalence, focusing on black spot disease in littoral freshwater fish communities sampled across 15 lakes in Québec, Canada. Our results show that infection prevalence is spatially heterogeneous across the landscape with evidence of infection hotspots and coldspots. Method-related sampling biases led to significant variations in prevalence estimates and spatial patterns of disease occurrence. Our results also indicated that low sampling efforts tend to overestimate the prevalence of infection in the landscape, and that the sampling effort required to estimate infection prevalence depends on the sampling method employed. Physico-chemical characteristics of the sites and local fish community structure were found to be the best predictors of infection at smaller spatial scale. Furthermore, our results suggest dilution effects due to obstruction and compatibility barriers limiting cercarial survival. Several relationships between infection prevalence and environmental predictors revealed non-linearity, suggesting complex interactions. Examining infection prevalence data at various spatial scales allowed to uncover many sources of variation revealing the importance of context-dependencies and scale-dependencies in host-parasite interactions.

Correspondence | Juliane Vigneault, Département des sciences biologiques, Université de Montréal, QC, Canada
Email : juliane.vigneault@umontreal.ca

1 | Introduction

Despite often being neglected in ecological studies (Chrétien et al., 2023; Morley, 2012; Scholz & Choudhury, 2014), parasites are key components of natural communities and ecosystem functioning (Dobson & Hudson, 1986; Frainer et al., 2018; K. D. Lafferty et al., 2008; Marcogliese, 2004; Minchella & Scott, 1991; Pascal et al., 2020; Poulin, 1999). Parasites can induce physiological and behavioral changes in their hosts (Barber et al., 2000; Iwanowicz, 2011), are major elements of food webs (K. D. Lafferty et al., 2006, 2008; Marcogliese & Cone, 1997a) and are impacted by global change (Brooks & Hoberg, 2007; Cable et al., 2017; Carlson et al., 2020; Mostowy & Engelstädter, 2010). Despite this recognition, infection patterns within a population or community can vary across taxa, time and/or space making it difficult to understand and model the drivers of infection within ecological systems (Happel, 2019; Poulin, 2006; Poulin & Dick, 2007; Thieltges & Reise, 2007; Villalba-Vasquez et al., 2018; Young & Maccoll, 2017). Consequently, our understanding of the sources of variation in parasite infection across spatial scales remains poor (Bolnick et al., 2020; Poulin, 1998). Since studies are constrained by context-dependencies such as spatial scaling, biological scaling, and sampling design, a multi-scale approach to investigating host-parasite ecology is imperative. As humans modify ecosystems (physically and biologically) it is essential to understand the full extent of complexity of disease dynamics to improve future predictions and select appropriate scales for management and conservation strategies for both parasites and their hosts (Cohen et al., 2016).

The distribution of organisms in space is constrained by species interactions, dispersal limitations and environmental filtering (i.e. the range of ecological and physiological conditions a species can tolerate) (Leibold et al., 2004). Spatial scale-dependance of these processes has long been established in ecology (Rietkerk et al., 2002; Viana & Chase, 2019). In this regard, single-scale studies tend to overlook the full extent of ecological processes shaping species distribution and interactions, leading to biased conclusions (Levin, 1992; Peterson & Parker, 1998). This issue has also been highlighted in disease ecology. For example, Cohen et al., (2016) found that the relative importance of processes driving the distribution of disease emergence (host richness, abiotic factors, and human population density) varied depending on the spatial extent of the study. The

scaling-level on which empirical data is used for estimating infection parameters can mislead conclusions if cross-scale dependencies are unknown. For instance, in a study on nematodes associated with alfalfa (*Medicago sativa*) at three spatial scales (i.e., river basin, field and microhabitat), the abundance of parasites at one scale did not explain the abundance at another scale revealing the importance of cross-scale processes (e.g., dispersion by irrigation water) in heterogeneous distribution patterns (Simmons et al., 2008). Broadly, communities were similar with *Panagrolaimus* sp. dominating communities at river basin scale while parasite abundances were mostly explained by moisture gradient at field scale (Simmons et al., 2008). One of the first steps to clarify these context-dependencies in infection dynamics is to address the importance of spatial scaling for the estimation of infection parameters such as the community prevalence (i.e., proportion of infected host in a host community). Prevalence is a common parasitological parameter that gives information on the proportion of infected individuals in a given group, and can be measured for both populations and communities of individuals (Bush et al., 1997; Rózsa et al., 2000). Patterns of regional-scale infection parameters can be caused by different underlying local patterns of infection. For instance, a regional prevalence of 20% could be caused by all sites having the same prevalence with no variance (i.e., all sites at 20%) or by the average prevalence of all sites being at 20% with large inter-site variance (e.g., half at 0% and half at 40%). Determining which underlying pattern at the local-scale drives patterns at the regional scale is thus critical for understanding how infection is distributed across a landscape.

The importance of scaling in explaining observed trends is especially relevant in parasite ecology given the hierarchical organization of both the parasite and host component. For example, parasites can be studied at different levels both within and among hosts. The simplest clustering organization of parasites is an infrapopulation, a parasite population within a host individual. Different parasite species can also co-exist within an individual host (infracommunity) or within a host population (component community) (see Bush et al., 1997 for more extended terminology). Component communities of metazoan parasites in fish host populations have been the focus of many studies exploring spatiotemporal patterns of infection (e.g., Carney & Dick, 2000; González & Poulin, 2005; Pérez-del-Olmo et al., 2009; Villalba-Vasquez et al., 2018). For example, Villalba-Vasquez et al. (2018) looked at infracommunities and component communities of parasites in

Panama spadefish (*Parapsettus panamensis*) collected over 6 years from 4 locations. Although they found similar patterns of low species richness, low diversity and dominance from a single parasite species across locations, there were high variations in species composition and community structure between years and locations (Villalba-Vasquez et al., 2018). Conversely, species-level infection patterns in host communities (i.e., one parasite species infecting many host species in a community) remains largely unexplored across spatial scales (but see Bielby et al., 2021; Moore et al., 2014; Olsson-Pons et al., 2015). Since infection parameters vary across time, space and species studies on disease ecology should focus on more than one species to disentangle processes explaining infection dynamics on a broader scale. For example, studying community-level infection dynamics would allow exploring a broad array of sources of variations in infection metrics across spatiotemporal scales.

Predictors are important to assess characteristics of infection hotspots and predict extinction risk with growing anthropogenic pressures. Despite recent efforts, it remains unclear whether infection patterns result from stochasticity (i.e., infection patterns that do not differ from randomness) or determinist filtering (i.e., infection patterns that are explained by a set of known drivers and mechanisms) (e.g., Carney & Dick, 2000; González & Poulin, 2005; Kennedy, 2009; Poulin & Valtonen, 2002). However, numerous studies have found correlation between infection metrics and various environmental and host components (e.g., Altman & Byers, 2014; Falke & Preston, 2021; Lagrue et al., 2011; Lagrue & Poulin, 2015; Poulin & Morand, 1999; Thieltges et al., 2008). On one hand, individual host characteristics such as sex (Zuk & McKean, 1996), age/size (Marcogliese et al., 2001; Poulin, 2000), genetics (Williams-Blangero et al., 2012) and personality traits (Barber & Dingemanse, 2010; Hart, 1990) drive host susceptibility leading to sources of variation in infection parameters of host populations. For example, Poulin (1996) found that infection prevalence of trematodes was higher in males across bird and mammal hosts. Infection prevalence can thus be sex-biased within host populations (Kowalski et al., 2015). On another hand, host community metrics such as species richness and host population density can create a "dilution effect" by reducing a parasite's encounter rates with target hosts (Ahn & Goater, 2021; Buck & Lutterschmidt, 2017; Civitello et al., 2015; Dargent et al., 2013; Lagrue & Poulin, 2015). For instance, incompetent species (i.e., non-host species) in a community can reduce infection in

target host species (Ahn & Goater, 2021) because failed infection attempts on incompetent species reduce remaining energy budget, shorten the infective temporal window or kill the infective stages. Several physical and chemical habitat characteristics have also been associated with infection patterns. For instance, in American eels (*Anguilla rostrata*), parasite diversity decreases when the water pH is below 5.4, with some parasite families such as digenean trematodes being absent below pH 4.7 (Marcogliese & Cone, 1997b). Similarly, lentic waterbodies with dense vegetation display higher prevalence and abundance of *Posthodiplostomum cuticola* parasites in host fish than rivers and reservoirs with steep banks (Ondrackova et al., 2004), suggesting that local habitat and morphometry of the waterbody are drivers of encounter filtering (spatiotemporal window that allow encounters between the parasites and hosts). Furthermore, spatial features of an ecosystem (e.g., watershed, connectivity) act as large scale filtering by limiting dispersal of both parasites and hosts (Bolnick et al., 2020) thus creating infection clustering at the landscape level. For instance, Happel (2019) provided evidence for consistent geographical patterns in the prevalence of trematodes in fish communities between watersheds. Bolnick et al. (2020) also found that increased distance between waterbodies leads to greater differences in parasite community composition in threespine stickleback (*Gasterosteus aculeatus*). Investigating a wide range of predictors has the potential to enlighten processes driving infection clusters (i.e., infection hotspots) within a system that appears homogenous (Bolnick et al., 2020). Consequently, a complete ecological perspective on host-parasite systems must incorporate abiotic, biotic and spatial ecosystem elements to highlight mechanisms shaping patterns of infection (Bolnick et al., 2020; Cohen et al., 2016).

Sampling wild populations is often a tradeoff between minimizing the stress of focal species and the impact of sampling techniques on natural habitats while generating accurate estimates of populations and communities when designing their sampling protocols. Accuracy of population parameters (e.g., abundance, recruitment, age classes, sex-ratio) can then be influenced by the sampling method and/or the sampling effort (i.e., the number of sampling). For example, Khaemba et al., (2001) simulated three sampling designs based on empirical data and showed that the intensity of sampling (i.e., the closeness of the transects) improved the precision of the estimated population in all three designs. Wildlife sampling methods often rely on animals' risk-driven

decision. Thus, devices that require a higher level of interaction with the animal for it to detect, or capture it (e.g., a trap requires a higher level of interaction than a camera), tend to sample fewer and bolder individuals inducing a personality filtering of the sampled population (Johnstone et al., 2021). For example, Biro & Post (2008) noted that bold rainbow trout were caught three times as often as shy individuals in gillnets indicating a personality bias in trappability of fish. Consequently, estimations based on empirical surveys are context-dependent on how the methodology used is personality-biased. Although it has not thoroughly been investigated yet, there is no reason indicating different outcomes for estimates of infection parameters in the wild. A study on the chronic wasting disease in deer species showed that the survey method used to estimate the disease prevalence can lead to overestimation or underestimation of the true infection parameter (Nusser et al., 2008). Wilson et al. (1993) also observed different infection levels in Pumpkinseed sunfish (*Lepomis gibbosus*) depending on the fishing method. Therefore, sampling design elements such as sampling effort and sampling methods should be considered when studying host-parasite systems as they might lead to bias in infection parameter estimates.

Here, we explore sources of variation in infection prevalence estimates across sampling methods focusing on data on the black spot disease in littoral fish communities across 15 lakes varying in morphometric attributes and, local biotic and abiotic conditions. We analyze infection prevalence data at the fish-community level (all fish hosts) for context-dependencies through three spatial scale levels (landscape-, lake- and site-scale) in order to investigate (i) the effect of increasing sampling effort on landscape prevalence estimation (random resampling accumulation curves) through different sampling methods (ii) the distribution of observed lake prevalence estimates across the landscape and (iii) the importance of ecological predictors on infection prevalence on a site-scale sampling. For landscape-scale analysis, we generated random prevalence accumulation curves by resampling across all lakes. This approach allows us to compare regional-scale prevalence estimates among different sampling methods along a gradient of increasing sampling effort (or area sampled) in order to evaluate the number of samples needed to correctly estimate the landscape prevalence (Figure 1.A). For lake-scale analysis, we used a frequency and spatial distribution of observed infection prevalence in lakes to better understand the infection dynamics across the landscape (Hess et al., 2002; McGeoch & Gaston, 2002)(Figure

1.B). The final step was to test a suit of deterministic environmental predictors underlying different scale-dependent processes to understand environment-driven variations in observed site-scale infection prevalence. The shapes and strength of the relationships between infection parameters and drivers are a key element to better understand the mechanisms behind infection patterns (Poulin et al., 2011).
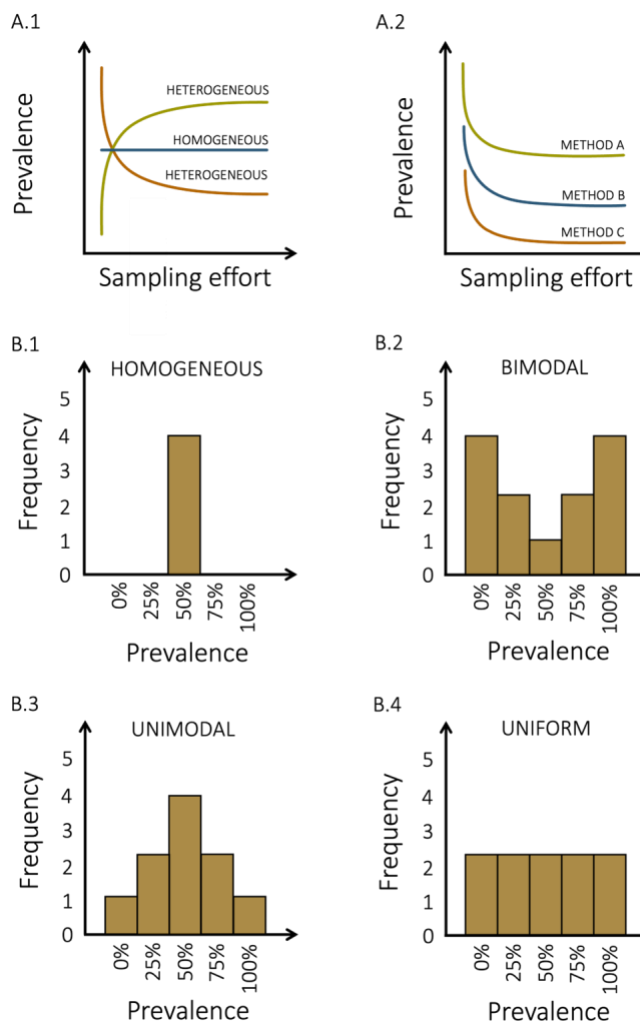
Figure 1. Case scenarios of (A) sampling effort effect on landscape prevalence estimates and (B) frequency distributions of lake prevalence estimates. A.1) Prevalence could be homogeneous (blue line) or heterogeneous (green and brown lines) across the landscape. In heterogeneous scenarios, prevalence is either overestimated (brown line) or under underestimated (green line) before stabilizing around the accurate prevalence. A.2) Sampling methods could lead to different prevalence estimates across the landscape. Prevalence difference between stabilized lines would represent method biases. B.1) All lakes could be infected at the same prevalence level. B.2) Lake's frequency distribution could follow a bimodal pattern where low and high prevalence are more common across the landscape. B.3) Lake's frequency distribution could follow a unimodal pattern where mean landscape prevalence is more common. B.4) Lake's frequency distribution could follow a uniform pattern where all infection levels are equally represented.

2 | Methods

2.1 Host-parasite system

Black spot disease is a common fish infection in marine and freshwater ecosystems caused by Digenean trematodes (flukes) (Kurochkin & Biserova, 1996). While some aspects of the ecology of these parasite species (e.g., *Posthodiplostomum cuticola, Uvulifer ambloplitis*, *Crassiphiala bulboglossa* and *Apophallus brevis*) are relatively well studied (e.g., Hoffman, 1956; Hoffman & Putz, 1965; Hunter, 1933; Miller, 1946; Ondrackova et al., 2004; Sinclair, 1972), spatiotemporal infection patterns and their relation to environmental conditions remain poorly understood. Black spot-causing trematodes all have a similar complex life cycle requiring snails, fish and piscivorous birds as hosts (Figure 2). The parasite encysts under the skin, in the fins or the muscles of the fish host (Hoffman, 1956; Krull, 1932, 1934) where it can survive in a dormant form for a few years (Hoffman & Putz, 1965) until the infected fish gets eaten by a bird. The distinctive black cyst symptom results from melanin pigment stimulation in the fish tissues induced by trematode cercaria penetration (Davis, 1967; Lemly & Esch, 1984a). Here, black spot-causing trematodes will be referred as a guild because they have similar ecological requirements and cause the same symptoms in fishes.

Akin to other parasitic diseases, the infection parameters of black spot disease can vary across time, space, and species. For example, the infection prevalence in bluegill sunfish (*Lepomis macrochirus*) in ponds is typically at its lowest in April/May and reaching its maximum around September/October (Lemly & Esch, 1984a, 1984b). This seasonal tend occurs because heavily infected fish are more susceptible to overwintering mortality leading to reduced infections at the population level in the spring. As cercariae are released throughout the summer and infect more hosts through time, prevalence reaches its maximum in the fall (Lemly & Esch, 1984a, 1984b). Habitat characteristics that favor snail populations such as shallow waters, low water flow and abundant macrophytes have also been associated with higher abundance and prevalence of infection in fish hosts (Marcogliese et al., 2001; Ondrackova et al., 2004). Hence, the infection parameters of black spot disease are usually higher in lentic systems and in the littoral zone rather than the pelagic zone (Ondrackova et al., 2004) because of a higher encounter probability with

cercaria in these regions. Since the black spot disease is caused by a parasite guild, a large range of fish species can be infected with some species more heavily infected than others. For instance, Lemly & Esch, (1984a) found that bluegill sunfish were more heavily infected than largemouth bass (*Micropterus salmoides*) because the two species might differ in habitat preferences and suitability for *U. ambloplitis*.
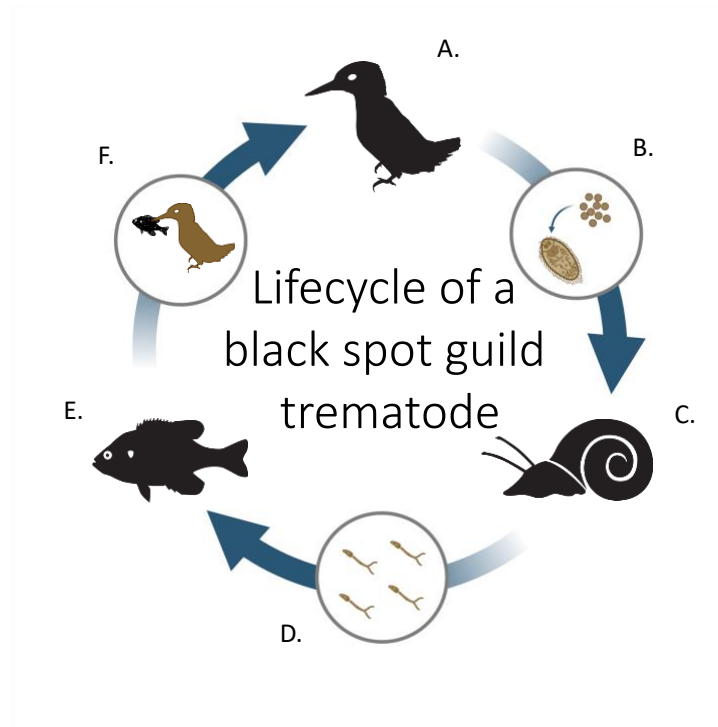
Figure 2. The typical lifecycle of a parasite species causing the black spot disease in North American freshwater fishes. (A) The definitive host is a piscivorous bird, usually a kingfisher (*Megaceryle alcyon*) or a great blue heron (*Ardea herodias*). The metacercariae mature to its adult form and reproduce in the bird digestive system. (B) The eggs are released in the water through the bird feces and grow into miracidia (first larval form). (C) The miracidia infect the first intermediate host, a freshwater snail (typically of Planorbidae or Amnicolidae families). The larvae develop into free-living stage cercariae (second larval form) and leave the snail. (D) The cercariae swim until they encounter a potential fish host. (E) The second intermediate host is a fish. The cercariae penetrate under the skin or the fins and transform in the metacercaria form. (F) The fish is eaten by a definitive host.

2.2. Study area

We sampled 15 lakes from six watersheds around Saint-Hippolyte, QC, Canada (45.92°N, 73.97°W). The lakes Corriveau, Croche, Cromwell and Triton are located in the protected area of the Université de Montréal's Biology Station of the Laurentians. The remaining lakes are surrounded by private properties and are accessible only upon consent of landowners or municipal associations. Considering their geographical proximity, we assumed that all the sampled lakes are exposed to the same broad climatic conditions. The lakes were selected nonrandomly according to

their accessibility (i.e., shore access on private properties granted by landowners) and the availability of morphometric data (see Table S1 for the morphometric characteristics).

2.3. Data acquisition

Infection prevalence was estimated as the number of hosts infected with black spot disease divided by the total number of hosts sampled in a given fish community. Since we did not identify black spot causing trematode species, using species-specific infection metric would be inadequate to assess environmental characteristics driving infection risk. Prevalence estimates are less sensitive to host and population identity (Poulin, 2006) and are thus, are an appropriate metric to study community-level sources of variation in disease dynamics. Three sampling methods (minnow traps, seine nets and transects) were used to assess infection prevalence and maximize the chances of catching different fish species. Field work took place from mid-June to the end of August 2022 and was restricted to days without precipitation to limit sampling bias due to meteorological effects. Animal handling was approved by the Université de Montréal's animal care committee (protocol number 22-025) and the scientific fishing permit was granted by the Ministère de l'Environnement, de la Lutte contre les changements climatiques, de la Faune et des Parcs (MELCCFP) of Québec (2022-05-16-1971-15-S-P).

Sampling effort within lakes was determined according to the lake surface area, except for minnow traps (see Table S2) as setting these traps requires less time and manipulation than the other methods. Seine sampling varied between four and eight fishing efforts while the number of transect sites varied between two and six per lake. 15 minnow traps were set in each lake to maximize the number of captures.

The minnow trap and seine net samplings were both conducted on the same day for each lake. The seine samplings occurred during daytime (usually between 10AM and 4PM) in habitats without large obstacles like trunks and rocks to prevent net obstruction. Two sizes of minnow traps were set at approximately equal distance along the shore at 4PM and pulled four hours later, to target species that are most active at dusk. All traps sat on the bottom substrate, but there were no other habitat restrictions for the minnow trap placement. Eight of the traps (3 large and 5 small

traps) were baited with bread to sample various feeding preferences and behavioral traits. The fishing gear dimensions are available in Table S3. All the fishing gear was pressured washed and sun dried between lake sampling following MELCCFP recommendations to prevent the spread of exotic species contamination.

All captured fishes were counted, identified to the species level, and measured (estimation of the total length to the nearest centimeter) directly after capture and subsequently released. Northern redbelly dace (*Chrosomus eos*) and finescale dace (*Chrosomus neogaeus*) individuals were identified as *Chrosomus* spp. as the two species hybridize in these lakes and cannot be distinguished based on morphology in the field (Leung et al., 2017). The presence of black cysts was assessed by first examining the left side of the fish. The right side was then examined for the presence of black spots if the left one did not have any. Juveniles and adults were included in this study as both life stages are vulnerable to black spot disease.

Fifty meters snorkeling transects positioned along the shore were conducted to assess black spot infection prevalence in the fish communities. We selected sites that were between approximately 0.5 and 3.0 m deep, not fully covered by macrophytes, not obstructed by docks and preferably with some vegetation, rock, or trunk refuges. The sites were selected and flagged every ten meters at the end-May 2022. The transects were performed simultaneously by two observers each monitoring one-meter field of view in front of them and moving forward at a pace of three minutes per ten meters. The fishes coming from behind the observer were not counted to avoid recounting curious fishes that often follow the observers. All the fishes were identified to the lowest taxonomic level possible and visible black spot infection was noted. All cyprinids were grouped into the same category due to the difficulty in accurately identifying fish to species level while swimming.

Following the prevalence transects, we completed the habitat description. The percentage of four substrate categories (silt, sand, rock, and boulder), the macrophyte coverage, the number of trunks (or large branches), and the mean depth was estimated for each 10-meter transect section. The mean estimate of each category was then calculated for all variables at the transect-scale. Temperature, dissolved oxygen, conductivity, turbidity, and pH were measured at mid-depth

(around 0.3 to 2 meters) at the beginning of every transect with a YSI ProDSS Multi-Parameter Water Quality Meter. 1 L of unfiltered water was taken at mid-depth at each transects in previously acid-washed HDPE bottles for carbon and nutrient content analyses. There, each transect samples were placed on ice in a dark cooler and transported back to the field station laboratory where the 1 L samples were separated in one previously acid-washed 40 mL vials to measure total organic carbon (TOC), and one 500 mL HDPE bottle for total nitrogen (TN) and total phosphorus (TP). TOC samples were placed in a 4°C refrigerator and analyzed within the week while TN and TP samples were kept frozen at a-20°C until processing. Every week, samples were brought back on ice at Université de Montréal (MIL campus) for processing. TOC sample contents were measured on a Sirevers M5310 C Laboratory Total Organic Carbon Analyzer. TN and TP samples were oxidized with persulfate and autoclaved the day before analysis following EPA methods 353.2 and 365.3 respectively (US EPA, 1978, 1993). TN concentrations were measured on a Lachat QuikChem 8500 analyzer and TP concentrations were analyzed on a Asoria-Pacific Astoria2.

2.4. Statistical analyses

We described the black spot prevalence patterns across multiple scales (landscape, lake and site). All data manipulations and analyses were conducted using R (version 4.2.3). Lake Tracy was excluded from the lake-scale analysis because of low fish abundance data obtained with the fishing methods (N = 1 individual sampled). Prevalence distributions were compared between methods using a Kruskal-Wallis rank sum test and Dunn test with a Benjamini-Hochberg correction (Dinno, 2017).

For the landscape-scale analysis, we looked at the effect of the sampling design (here sampling method and sampling effort) on the estimation of landscape fish infection prevalence. We used a resampling approach on the data from the different sampling methods (minnow trap, seine net, transect, and all methods combined) to investigate how the infection prevalence estimates change along a gradient of increasing sampling effort. The sampling effort associated prevalence was calculated as the number of infected individuals divided by the total number of individuals regardless of the origin lake. For each sampling method, we randomly selected N from our data to estimate the landscape prevalence (we used a weighted mean to estimate the landscape

prevalence across samples to avoid the effect of different fish abundance in samples). The operation was repeated 999 times for a sampling gradient N from 1 to 35 samples. We set the largest sampling effort at 35 because we only had 39 total transect samples across all lakes, and we did not want to resample the entire data set. A local regression (loess) was then fitted to the landscape estimates generated by our simulation of each sampling effort value (N) for visualization of the trend of mean landscape prevalence through the sampling gradient. We then compared for each method the observed prevalence (calculated with all samples) with resampled landscape prevalence at N = 5, 10, 15, 20, 25, 30 and 35 to see when accurate prevalence (i.e., when resampled preference is not different from observed prevalence) is reached.

For the lake-scale analysis, we examined the frequency distribution of the fish infection prevalence according to the sampling method. The mean infection prevalence was previously calculated for each lake, and visualization was made with histograms set at six bins to avoid gaps. The distributions were then compared with the maps of the study area showing prevalence gradient in order to look for spatial patterns.

For the site-scale analysis, we modeled the relationships between environmental drivers and the community infection prevalence on a transect level. We used the prevalence data from the transect method because it is the only method that allows an accurate association of infection prevalence with all environmental variables (including those measured at the site-scale), thus can best reveal the key drivers of infection prevalence. Prior to analysis, we explored the data for outliers in both the response and explanatory variables, for collinearity between explanatory variables, and for non-linearity in the relationships between the response and the explanatory variables following recommendations by Zuur et al. (2009). Since our data showed some non-linear patterns, we applied generalized additive models in a mixed-model framework (GAMMs) using the *mgcv* package (Wood, 2017). We modeled one environmental variable at a time using a cubic spline (smooth function) and a random effect on the lake variable to account for spatial clustering of sites from the same lake. Maximum likelihood method (method = "ML") was used for estimating penalties. We used a quasi-binomial family distribution because our response variable (infection prevalence) is a proportion constrained between 0 and 1 and, to account for overdispersion in the

data. The optimal amount of smoothing (λ) and the number basis functions were internally determined by the modeling function from the *mgcv* package (Wood, 2017). The deviance explained (D^2) was used as a measure of model fit. The model validations were conducted with *gratia* package (Simpson, 2023) and visualization of partial effects was made with *gratia* (Simpson, 2023) and *ggplot2* (Wickham, 2016) packages."

3 | Results

Our main goal here was to investigate how prevalence of black spot disease in fish communities is perceived across three scale levels (landscape, lake, site) and assess context-dependencies. 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 and 8 families for this study (Table S4 for details of species abundance). The minnow traps caught 1906 individuals from 10 species while seine nets caught 2427 individuals from 14 species. 6964 individuals belonging to five taxonomic groups including four species and one separate family were observed in the snorkeling transects. The mean length of all fish captured though fishing methods was 5.59 ± 2,96 cm (mean length ± SD; N = 4333, see Table S5 for mean length according to all methods). The mean length of lake-scale fish communities is presented in Table S6 by fishing method.

Scientific literature supports black spot disease susceptibility in all fish species sampled except for *Ameiurus nebulosus* and *Esox masquinongy* (see Table S7). However, our field 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*). Consequently, *A. nebulosus* and *E. masquinongy* individuals were excluded from the prevalence estimation as there is no evidence they can host a trematode causing the black spot disease. Details on infection prevalence of fish communities are presented for every sampling method in Table S8 (landscape-scale), Table S9 (lake-scale) and Table S10 (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 3). Accumulation curves built for each sampling method tended to overestimate the mean prevalence of infection at low N values (number of samples). The estimate of landscape resampled prevalence differed between the sampling methods (Kruskal-Wallis chi-squared = 1288, p-value = 0, Table S11), varying between 21% and 36% (Table S12). After 35 samples, the transect method generated the highest mean prevalence (36 %) followed by the methods combination (31 %), the minnow trap method (25 %) and the seine method (21 %). However, the minnow trap curve did not stabilize after the 35 random samples implying that the prevalence value obtained is higher than the actual estimate (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.
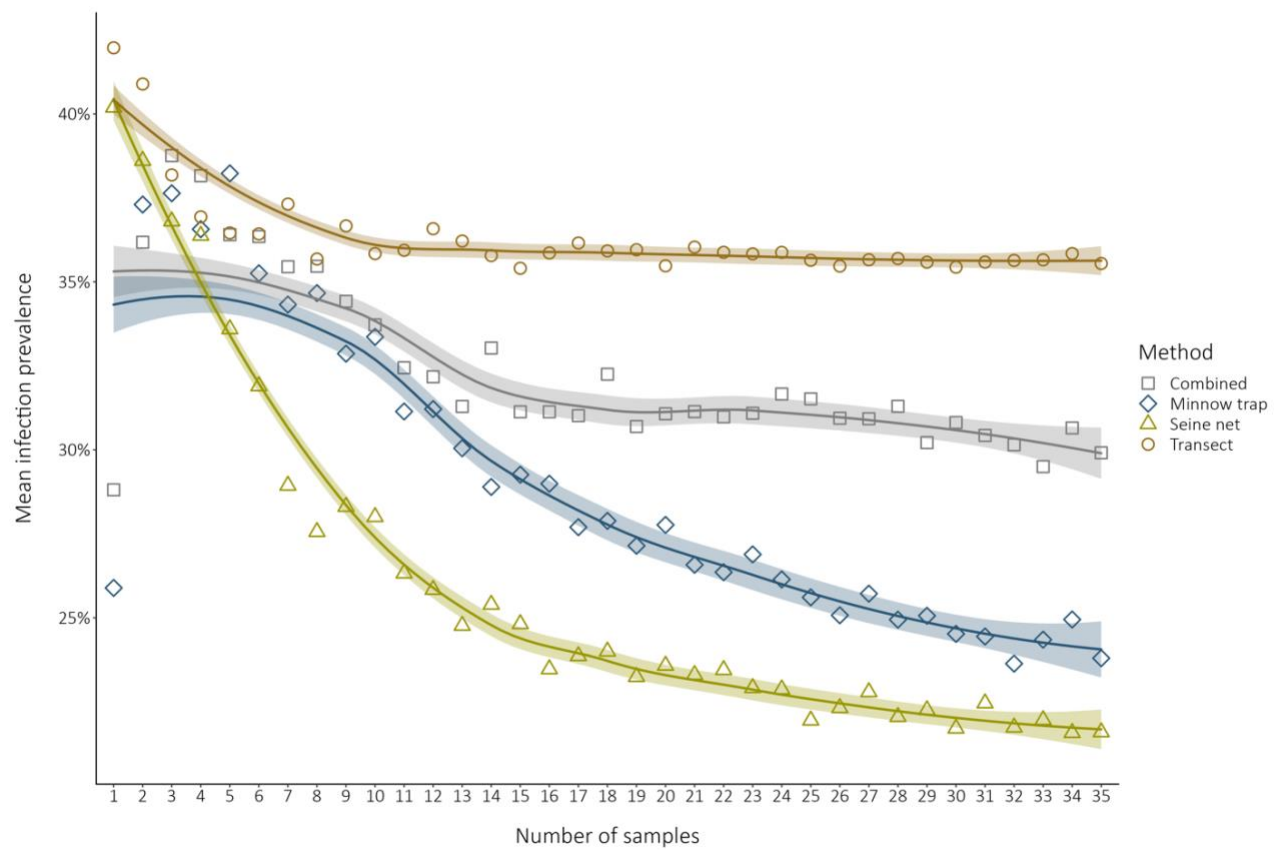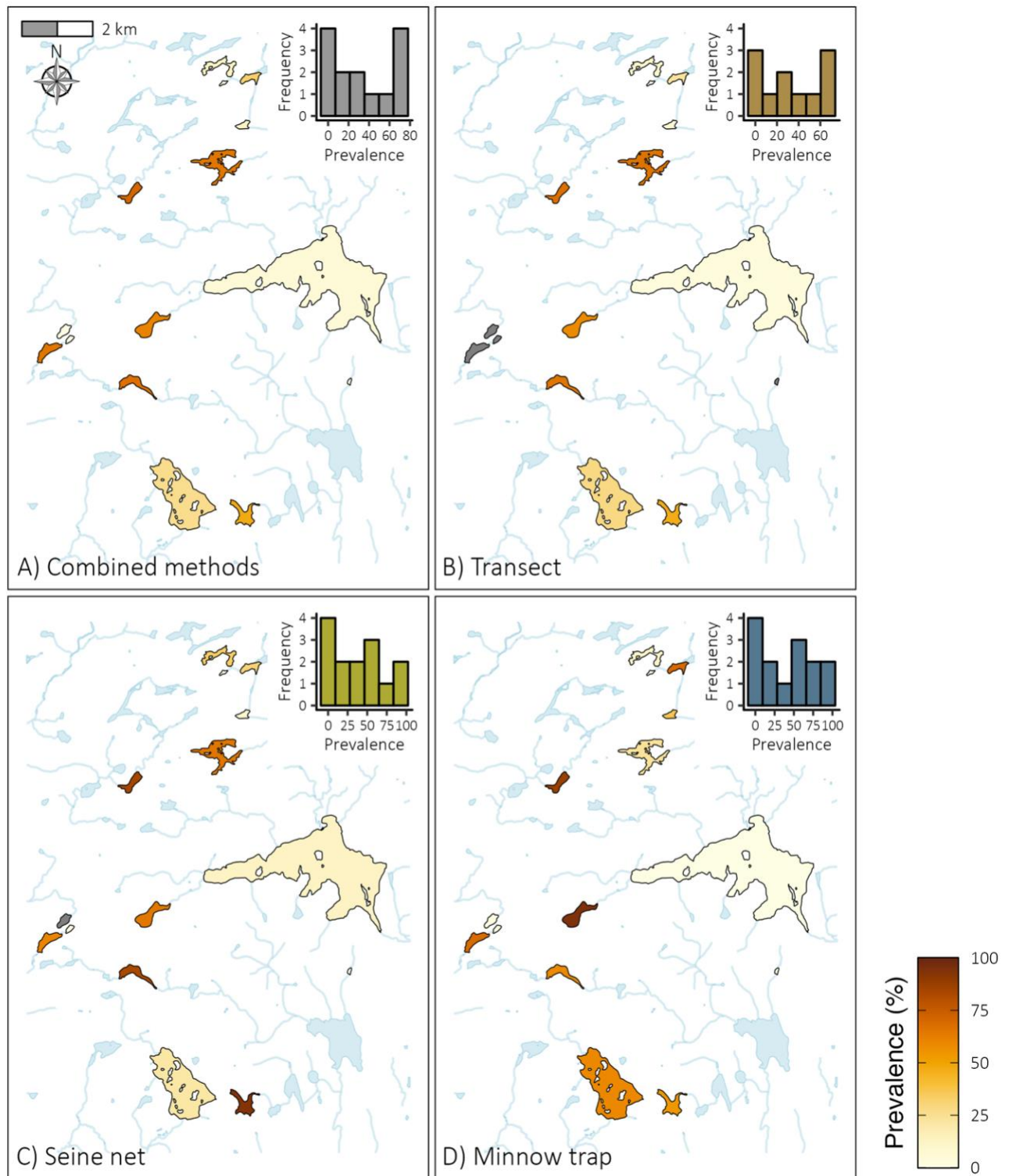
Figure 3. Method comparison of landscape prevalence estimates through an increasing random sampling effort. The data points indicate the mean prevalence for a given number of samples (N) randomly selected. The shadings indicate the 95% confidence interval on the loess smooth. The pool data used for the resampling with the minnow traps, the seine nets, the transects and all methods combined respectively contained 225, 75, 39 and 339 samples.

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 4). Fish community infection prevalence estimates measured at the lake level are not constant across the landscape. Both frequency distributions and prevalence maps support a difference between methods leading to biases in infection estimates. The lake prevalence frequency distributions for the combined methods and for the transect method showed 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 had similar estimates of landscape-scale prevalence (20.45% and 19.46% respectively, Table S8). Although spatial patterns vary across sampling methods, heavily infected and less infected lakes do not appear to be clustered in space at the regional scale (see maps in Figure 4). Moreover, close and connected lakes do not appear to follow a spatial infection gradient. Nevertheless, lake-scale geographic patterns were not statistically tested.

Figure 4. 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 29.62 %). (B) The transect method (weighted mean regional prevalence is 35.55 %). (C) The seine nets (weighted mean regional prevalence is 20.45 %). (D) The minnow traps (weighted mean regional prevalence is 19.46 %).
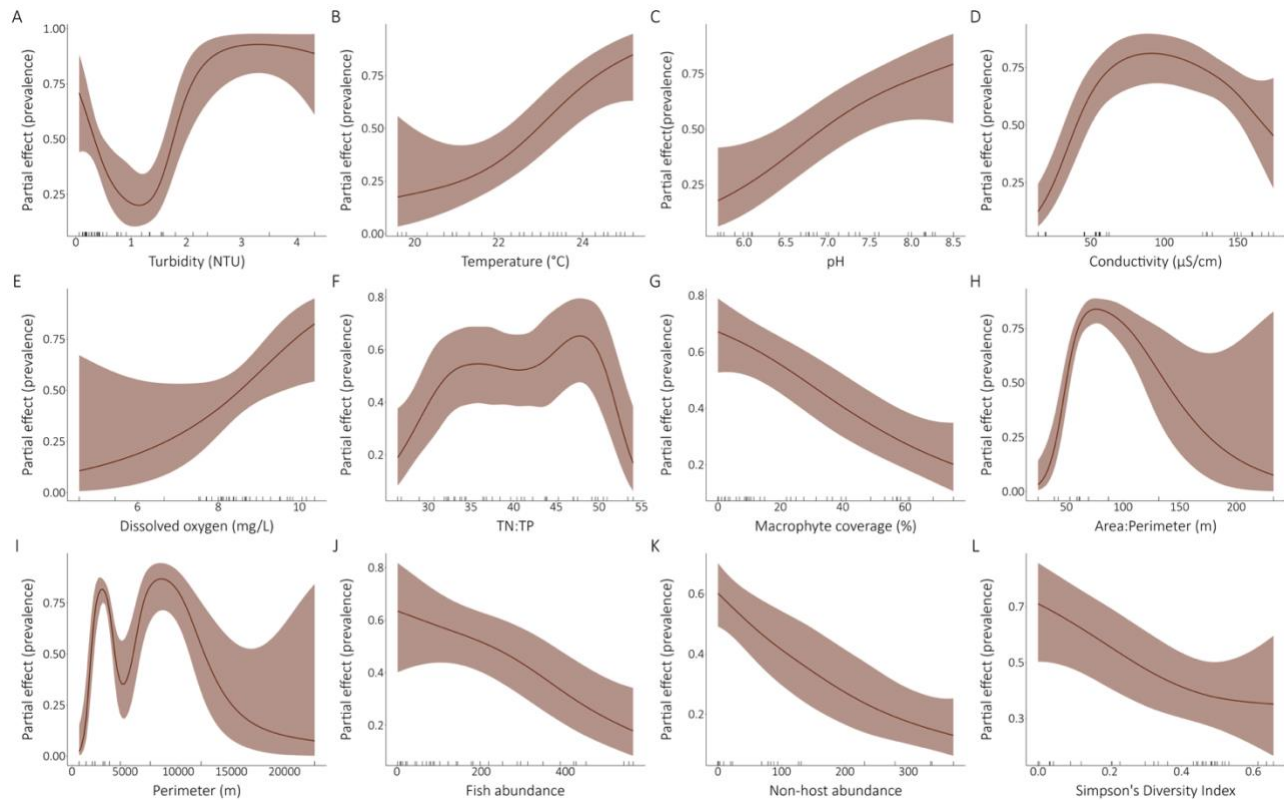
3.3. Site-scale

The relationships between the potential predictors and the site-scale infection prevalence (transect-level prevalence) were assessed with generalized additive mixed effects models (see Table S13 for models' estimates). The environmental characteristics of transect sites used as model predictors are presented in Table S14.

The models for TN, TP, TOC, substrate type (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 (Table S13). We did not find any relationships between black spot prevalence and any spatial attributes of the lakes. However, water physico-chemistry and community metrics were good predictors of infection prevalence of fish communities in our site.

The partial effects of environmental variables on infection prevalence are shown for all significant models in Figure 5. The turbidity model had the best fit (D2 = 88,71%) and is mostly non-linear. The relation revealed a plateau where maximum prevalence saturation is reached for turbidity values above 2.5 NTU. Infection prevalence increased proportionally with both water temperature and pH with prevalence being highest in warmer, more alkaline lakes. The conductivity and prevalence had a non-linear, unimodal relationship peaking around 80 (μS/cm). Prevalence values increased almost proportionally with dissolved oxygen concentration in the water. 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 infection prevalence in site-scale fish communities. A parabolic curve is observed in the Area:Perimeter model although the variation interval increases at high ratio values. The perimeter model is highly non-linear with two distinct 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 a 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 (Gini-Simpson's diversity index between 0 and 0.64) indicating the dominance of some species, namely cyprinids and *L. gibbosus*. Gaps in values of Turbidity, DO, Conductivity, Area:Perimeter and Perimeter environmental gradients have to be taken under consideration for interpretation of the model results.

Figure 5. 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.

4 | Discussion

Our main goal was to investigate the context-dependencies of infection prevalence estimates in freshwater fish communities across multiple spatial scales using data on the black spot disease. This approach allowed us to investigate spatial patterns of infection, sampling effort effect, sampling methods bias and, environmental predictors of the prevalence of infection in fish communities. Our results showed substantial differences in prevalence estimates across methods suggesting major sampling bias in infection estimates. Results showed that minimal sampling effort must also consider the sampling method as they present different outcome in prevalence accumulation curves. Local environmental characteristics such as water physico-chemistry and community structure were found to be the best predictors of infection prevalence at small spatial scale.

Study design in ecology aims to generate reliable data that are representative of natural systems while minimizing stress on the environment and maximizing statistical power. However, researchers often compromise aspects of the idealistic sampling design to overcome other challenges. Therefore, empirical data are biased towards designing and methodology choices. Surveys of disease prevalence among populations and communities often do not consider the importance of sampling design and methods bias despite the potential to substantially induce differences in measurements derived from wildlife samples collected using different techniques (Biro & Dingemanse, 2009). For example, Wilson et al. (1993) found that sunfish infected with black spot disease were twice as abundant in trapped fish compared to seined fish, whereas white grubs were twice abundant in seined fish, suggesting method-induced bias.

Sampling wildlife is challenging, and all methods come with their load of advantages and inconveniences. In our study, prevalence estimates' variations between sampling methods varied up to 16.1% at the landscape-scale (Table S8) and 74.0% at lake-scale (Table S9). These results evidence that method-induced bias can be substantial in the measurement of community-level infection prevalence. However, while transects measured the highest prevalence and minnow traps the lowest prevalence at landscape-scale, this trend was not maintained at the lake-scale. In fact, at lake-scale, methods that led to overestimation or underestimation of the infection prevalence

change from one lake to another. Moreover, method choice did not always lead to strong variation between prevalence estimates (e.g, lake Beaver and lake Montaubois). This suggests that method-induced bias is highly context-dependent on local community structure and composition.

Different sampling methods are doubtlessly associated with their own sources of bias that might affect prevalence estimates. Minnow traps are typically selective on active, bold and curious behavior because fish must voluntarily enter the trap (Biro et al., 2006; Wilson et al., 1993). Individual behavioral variation can also lead to variation in infection risk (Hart, 1990), where more active and bold individuals are usually associated with a higher infection risk (Boyer et al., 2010; Gradito et al., 2023; Santicchia et al., 2019; Wilson et al., 1993). Conversely, parasite infection can alter host behavior including habitat use and risk taking (Moore, 2002), making them more vulnerable to capture (i.e., "trappability")(Barber & Dingemanse, 2010; Thambithurai et al., 2022). In the end, it remains unclear if host individual behavior define their infection risk and/or whether parasite infection alters their personality (Dubois & Binning, 2022). In fact, the results of (Gradito et al., 2023) suggest that personality-driven susceptibility to infection and personality change due to infection can occur simultaneously in naturally infected sunfish. Although seine nets do not directly select host behavior, this method comes with important habitat-selection restrictions. Sites must not be deeper than the height of the net to avoid fish escaping and, ideally, must not have large obstacles (e.g., branches, trunks, boulders) which might damage or impair the net function. Consequently, seine sites might have similar habitat features that facilitate deployment and capture success. Therefore, seine nets might indirectly select for fish species or personality types with similar habitat preferences. For example, Biro et al., (2006) found that bold trout were more frequently caught in open risk habitat than shy trout, corroborating habitat use segregation based on intraspecific personality. However, habitat use can change in the context of competition and/or predation (Fischer, 2000; Mittelbach, 1986; Werner & Hall, 1977). Therefore, community composition must be considered when inferring species selection bias based on habitat use. Observational snorkeling transect is not a habitat-restricted nor a behavior-selective methodology. Although, quality of observations underwater relies on, meteorological conditions, water transparency (color, turbidity) and expertise of the observers to identify fish species in movement. As a result, precision of infection assessment can vary between sampling days and, between lakes.

Low black spot abundance might be missed because of poor visibility or fish swift movement, leading to underestimation of prevalence estimates. Moreover, since infection assessment is easier in curious and solitary fishes as they usually get closer to the observers, accuracy of prevalence estimates might vary between fish species and personality. Finally, since older/bigger fish have a higher risk of infection than younger/smaller ones age structure and growth curves might influence infection estimates (Lemly & Esch, 1984a). Although seine nets sampled more rare species that the other methods, our data did not show consistent species-specific selection across lakes. Further analyses would be needed to determine how species composition of samples, personality types and age structure influence community prevalence estimates, and how this vary across spatial scales.

Sampling effort must be sufficient to measure accurate infection metrics. However, parasitological parameters are often estimated from small sample size because of ethical, time, and monetary constraints. Furthermore, the relationship between infection prevalence in fish communities and sampling effort has not been addressed before. Our results showed that low sampling efforts tend to overestimate community infection prevalence at the landscape-scale (Figure 3). This is the case when most samples indicate a prevalence value above the landscape prevalence estimate. Therefore, small sample sizes have a higher chance to sample a higher-than-average prevalence thus overestimating the mean landscape prevalence. However, the presence of zero prevalence samples has a big impact lowering the prevalence estimated across the landscape as sampling effort increase. Presence of high and zero prevalence values suggest the existence of high and low infection clusters (infection hotspots and coldspots). Parasites tend to be aggregated in their hosts, where a few individuals in a population or community host the majority infection (Poulin, 2007a; D. J. Shaw & Dobson, 1995). One of the main recognized reasons for this is the environmental aggregation of infective stage leading to heterogenous host exposure risk (Karvonen et al., 2004). Then highly infected individuals might occur in suitable microhabitats for the development of infective stages (Morgan et al., 2004). Consequently, heterogenous landscapes should present infection hotspots where prevalence is high as parasite abundance and intensity positively correlate with prevalence of infection (Poulin, 2006; D. J. Shaw & Dobson, 1995; Šimková et al., 2002). Despite sampling effort effects on prevalence estimates have not been addressed

before, some studies have investigated the effect of the number of sampled hosts on infection prevalence estimate. For instance, Marques & Cabral (2007) found no significant effect of host sample size on prevalence values, but reported that simulated median values using a Monte Carlo procedure tended to over or underestimate the real value at low sample sizes. Jovani & Tella (2006) also reported that small host sample sizes resulted in greater inaccuracy in prevalence estimates and suggested an optimal sample size of around 15 individuals as a trade-off between discriminating samples and maintaining accuracy. Our results provide evidence that the minimum number of samples needed to adequately estimate landscape prevalence differs among methods. The only sampling method that reached a stable value following our sampling effort was the observational snorkeling transects (around 10 samples) although the seine and the combination of sampling methods showed less than a 2% difference between resampled prevalence and actual observed prevalence after 35 samples (Table S12). Indeed, snorkeling transects allow observations on more individuals than the fishing methods (6964 against 4333 observations), explaining why it reaches stability faster and thus, needs less sampling effort to accurately measure prevalence. Conversely, fishing methods require more extensive sampling in terms of area and time to estimate a representative prevalence estimate of infection. Accordingly, 35 sampling efforts (against 225 sampling efforts used to estimate observed prevalence) were not exhaustive enough to capture the actual infection landscape prevalence for the minnow traps (Table S12). However, assessments of infection intensity and prevalence are more accurate and precise with fishing because the observer can take the necessary time to inspect individuals for signs of disease. Ultimately, «magic cutoff» sampling effort for estimating accurate prevalence does not exist as it can vary with parasite aggregation level (Poulin, 2007b; Shvydka et al., 2018), host species target and abundance, sampling design (e.g., timing) and as we have shown, sampling method. However, our results should be used as an empirical tool to guide future determination of sampling effort to assess infection prevalence in natural systems.

Our results show that prevalence of infection is spatially heterogenous in our system (Figure 4 and 5). Prevalence estimates varied both among and within lakes. Regardless, we did not find evidence of geographical patterns. Accordingly, none of the spatial attributes that we tested were good predictors of infection at small spatial scales (site-scale) (Table S13). Future studies should

increase the number of lakes sampled to confirm whether infection is spatially random, or if indeed there are other patterns at play.

Frequency distributions are a useful tool to investigate mechanisms behind the spatial occurrence of species (McGeoch & Gaston, 2002). For parasites, frequency occupation distributions are modified such that hosts represent suitable patches (i.e. habitat)(Hess et al., 2002). In our case, frequency distributions were made to investigate the distribution of prevalence of black spot disease infection in lake fish communities. We found that frequency distributions differed between sampling methods; the transects and the combination of methods suggesting a bimodal distribution tendency (Figure 4.A, B) and both fishing methods showing no obvious pattern (Figure 4.C, D). While binomial distributions are commonly observed in frequency occupation distributions, some argue that they are an artifact of a small number of sites (Brown, 1984). However, for our purpose, bimodal distributions revealed that lake community tend to display a high or low infection prevalence. Since we assumed that the distribution of trematodes is not limited by the final hosts' movements (birds) across our study area, these high- and low-prevalence groups likely represent environmental characteristics that favor (or restrict) parasite abundance and transmission from snails to fish hosts. Alternatively, this could mean that when the parasite is present, there are regulating environmental pressures that limit the parasite abundance and transmission. Yet, if the parasites overcome the selective pressure, they become highly prevalent in the fish community. Sampling a larger number of lakes might clarify these patterns of infection dynamics.

Differences in frequency distributions imply that method biases are also perceptible at the lake-scale and that they influence observed prevalence distribution patterns. Accordingly, map comparisons (Figure 4) clearly show that lake fish community prevalence can be over or underestimated according to the sampling method used. For example, in Lake Pin rouge, the transect method underestimated the community prevalence while in Lake Croche, seine net sampling overestimated the community prevalence (Table S9). This suggests method biases in field sampling of fish communities. If we consider the previous examples of lake-scale prevalence, the methods might have selected for fish species varying in parasite compatibility, environment varying

in quality for infective stage aggregation, specific age classes or sampled different personality traits that led to biased prevalence estimates. Although, our approach revealed method biases in prevalence estimates, but it did not allow us to establish the mechanism behind this relationship. Moreover, method biases seem to be lake-dependent.

At the site-scale, we used GAMMs to inspect relationships between the prevalence estimate and environmental predictors (Figure 5, Table S13). Here, the use of a flexible approach (additive models do not assume linearity) to model empirical (data-driven) infection dynamics was supported by evidence of many non-linear patterns. Our results showed that small-scall parameters such as water physico-chemistry and community metrics are better predictors of prevalence estimates than larger scale metrics (e.g, lake- or landscape-scale parameters)(see Table S13) suggesting the local scale is more appropriate for management and conservation strategies. Turbidity, DO, Conductivity, Area:Perimeter and Perimeter models must, however, be carefully interpreted because of gaps in the corresponding environmental gradient (see Figure 5).

All physico-chemistry parameters models (turbidity, temperature, dissolved oxygen, conductivity, and pH) were good predictors (Figure 5.A-E) of infection prevalence at the site-scale in the sense that their model fit (D^2) is better than predicted by stochasticity (compared to the null model, see Table S13). Water physico-chemistry measurements are relatively easy to measure, do not require a lot of equipment or time and are non-invasive to organisms. Thus, water physico-chemistry might be useful as proxies of parasite infection for future monitoring and managing strategies. Consequently, we encourage researchers to investigate these relationships in other parasite species, geographical localities, and types of habitats as our data provide only a snapshot the complexity of fish-parasite associations in natural systems.

Surprisingly, habitat structure features were not good predictors of prevalence of infection in site-scale fish communities from the littoral zone. However, we found a strong (D^2 = 84.17%) negative relationship between the macrophyte cover and the prevalence of infection in transect sites (Figure 5.G). This was somewhat surprising as heavier infection has been measured in fishes associated with more vegetated zones in previous studies (Hartmann & Nümann, 1977; Marcogliese et al., 2001; Ondrackova et al., 2004). However, trematode cercariae have a short

free-living stage (24-72h depending on species and environmental conditions) (Galaktionov & Dobrovolskij, 2003; Morley, 2012) and thus physical barriers created by aquatic vegetation or other debris might prevent the cercaria from encountering a suitable host. For instance, macroalgae have been shown to reduce the number of trematode cercariae in coastal systems (Bartoli & Boudouresque, 1997; Welsh et al., 2014) thus interfering with the parasite transmission to the next host (fish or mussel). Similarly, Prinz et al., (2009) demonstrated that seaweed can obstruct transmission of trematodes cercariae to their mussel hosts. Buss et al., (2022) also suggested that microplastic fibers might reduce infection success of infection in frogs by impeding host encounters by cercaria. Welsh et al. (2014) also investigated organisms that can interfere with trematode *Himasthla elongate* cercariae and found that complex surface of oysters (*Crassostrea gigas*) led to a reduction of 44% of cercarial trematodes while smoother shells did not significantly reduce the number of cercariae. Another explanation is that the snail species releasing cercaria in our system may not be as reliant on macrophytes as some other aquatic gastropods. For example, the mud Amnicola (*Amnicola limosus*) is the second intermediate host to at least one species of the black spot trematode guild in our lake system (Levet, unpublished). This freshwater snail species live partially in the lake sediment from the littoral zone (Pinel-Alloul & Magnin, 1973) and often grazes on periphyton (Kesler, 1981). Consequently, we would not expect a strong relationship with macrophyte cover and cercaria presence for infections transmitted by this species.

Our results suggest that fish community structure is a driver of black spot disease prevalence at small spatial scales. We found negative relationships between infection prevalence and total fish abundance as well as non-host abundance and the Gini-Simpson's diversity index (Figure 5.J-L). These results support the encounter dilution effect hypothesis. Since total fish abundance and non-host abundance are positively correlated (Figure S1), we assume similar mechanisms explaining why lower prevalence was measured in transects with higher fish abundance. First, a higher non-host fish abundance might reduce prevalence estimates in fish communities by an unsuccessful infection attempts by cercaria on non-host individuals (susceptibility barrier). Next, non-host fish may induce a behavior change in hosts that reduces encounter rate with cercariae. For example, Ahn & Goater (2021) showed that abundance of brain worms decreased in fathead minnows (*Pimephales promelas*) when they were mixed with non-host

emerald shiners (*Notropis atherinoides*), but that there was no difference when minnows were mixed with other non-host species. The authors suggested that behavioral changes occur in minnows when they co-occur with shiners that reduce their likelihood of being infected. This emphasizes the role of fish species identity role in dilution effect mechanisms. Accordingly, more diverse communities (Gini-Simpson's Diversity Index) in our system were less infected than more uniform communities. In our lake system, *L. gibbosus* dominates littoral communities and is the species with the highest infection prevalence across the landscape (Table S15). The presence of competitor species (e.g., *P. flavescens*) in a lake or a site might decrease the overall prevalence of infection in the community by constraining pumpkinseeds sunfish's behavior thus reducing encounter with cercariae. Indeed, Dargent et al. (2013) found that the presence and abundance of *Gyrodactylus* spp. (monogenean) were lower in hosts when they occur in mixed-species groups. Even though it was not considered in this study, the local diversity of non-fish organisms could influence host fish. For instance, Lagrue & Poulin, (2015) found a negative association of non-host benthic invertebrate diversity with infection prevalence in second intermediate hosts. In fact, non-host species can reduce infection success of cercariae by actively or passively feeding on them (Thieltges et al., 2013). This has been observed in many taxa in aquatic ecosystems including zooplankton, Bivalvia, crustacean macroinvertebrates, cnidarian, and small fish (Kaplan et al., 2009; E. Mironova et al., 2019; K. Mironova et al., 2020; Schotthoefer et al., 2007; Vielma et al., 2019).

Understanding the processes explaining parasitic disease dynamics across scales is fundamental to predicting future infection and extinction risk in natural systems. Here, we investigated context-dependencies of fish community prevalence estimates of the black spot disease across multiple scales to elucidate the roles of sampling effort, sampling method bias, spatial occurrence patterns and environmental predictors in estimating host infection prevalence. Results suggested that prevalence is distributed in a non-random, heterogeneous way across the landscape with small sampling effort largely overestimating regional prevalence. We provide evidence for inconsistent method-induced bias in prevalence estimates at both the lake and landscape-scale, sometimes leading to high variation in prevalence estimates. The fish sampling method also influenced the sampling effort needed to reach accurate landscape prevalence estimates, with observational snorkeling transects requiring the least amount to achieve accurate

estimates. The best predictors of site-scale prevalence were local metrics such as water physico-chemistry and community structure metrics, suggesting that small-scale monitoring and management might be appropriate for future conservation strategies. Our findings also suggest that encounter dilution effects by host decoys or physical obstruction reduce infection prevalence in fish hosts. However, we observed many non-linearity in these effects suggesting complex interactions that we could not disentangle with our approach. Therefore, we encourage more research across larger spatial and temporal scales to deepen our understanding of the processes linking these predictors to infection parameters for future integration into sampling surveys and conservation plans.

Author's contributions

J.V., E.H and S.A.B conceived the ideas, J.V. did the field work, J.V. led the data analyses with input from E.H. J.V. led the manuscript writing with input and revision from S.A.B and E.H.

Funding information