# 4 | DISCUSSION

Prevalence is a common parasitological parameter that gives information on the proportion of infected individuals in a given group (Bush et al., 1997; Rózsa et al., 2000). Our main goal here was to investigate the estimate of prevalence in fish communities across multiple spatial scales using data on the black spot disease. This approach allowed us to skim over spatial patterns of infection, sampling effort effect, sampling methods bias and, environmental predictors of the prevalence of infection in fish communities.

Despite that fish hosts have been the focus of many empirical studies on aquatic parasites, sampling design and methods bias are not usually discussed although they can induce substantial differences in measurements derived from wildlife samples (Biro & Dingemanse, 2009). For example, Wilson et al. (1993) reported difference in parasite infection in pumpkinseed sunfish (*Lepomis gibbosus*) caught by two different fishing methods. They mentioned that black spots were twice as abundant in trapped fish against seined fish, and white grubs twice as abundant in seined fish, indicating method-induced bias. Our findings showed prevalence estimate variations between sampling methods up to 16.4% at landscape-scale (Table S12) and 46.7% at lake-scale (Table S13). These results suggest substantial method-induced bias in the measurement of community-level infection prevalence. However, while transects measured the highest and minnow traps to lowest prevalence at landscape-scale, this tendency was not consistent at-lake-scale.

Sampling wildlife is challenging, and all methods come with their load of advantages and inconveniences. While fishing methods require a more extensive sampling than field observations (they sample fewer individuals), assessments of infection are accurate and precise because the observer can take the necessary time to inspect fish individuals. Minnow traps are typically selective on active, bold and curious behavior because fish must enter voluntarily in the trap (Biro et al., 2006; Wilson et al., 1993). Besides, individual variations in behavior often lead to variation in infection risk (Hart, 1990), active and bold individuals usually associated to higher infection risk (Boyer et al., 2010; 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 “trappability” (Barber & Dingemanse, 2010). 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). Even if seine net is not directly a behavior-selective gear, 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) because the net would get stuck in it, resulting in tearing the net or, creating openings where fish can escape while researchers unstuck it. Consequently, seine sites might have similar habitat features despite effort to sample habitat heterogeneity. Since species have different habitat preferences, seine net might be indirectly selective on certain fish species and personality because of its habitat restrictions. 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 (Lemly & Esch, 1984).

Sampling effort must be sufficient to measure adequate infection metrics. However, parasitological parameters are often estimated with small sample size because of ethic, time, and monetary constraints. Our results showed that low sampling effort tend to overestimate the prevalence value at landscape-scale (Figure 6). This is because a high proportion of local sites had prevalence values above the regional average (see Table S14), thus at low sampling efforts the probability of randomly sampling a site with high prevalence is high. However, because many sites had a prevalence of zero, higher sampling efforts led to a quick decline in prevalence estimate until the true regional average was reached. Our results showed that low sampling effort lean toward overestimating the prevalence value at landscape-scale (Figure 6). If most samples inhere a prevalence value above the landscape prevalence estimate, there is a higher chance to sample those at low sampling effort thus overestimating the mean landscape prevalence. However, presence of few zero prevalence samples lower the prevalence estimated across the landscape as sampling effort increase. This evidence presence of high and low infection clusters (infection hotspots and coldspots). Parasites have an aggregated nature, where more individuals have low infection intensity, and few individuals are highly infected. This means highly infected individuals might occur in habitats favoring parasite occurrence and transmission. Thus, heterogenous landscape should present infection hotspots where prevalence is high as it correlates with parasite abundance (Šimková et al., 2002). The relationship between prevalence in fish communities and sampling effort have not been addressed before. Although, some researchers investigated the effect of sample size (number of sampled hosts) on prevalence estimate. Marques & Cabral (2007) found no significant effect of sample size on prevalence values, yet they reported that simulated median values tended to over or underestimate the real value at low sample size. Jovani & Tella (2006) also reported that low sample size resulted in greater inaccuracy and suggested an optimal sample size around 15 as a trade-off between losing data and maintaining accuracy. Our results evidence that minimum number of samples needed to adequately estimate landscape prevalence differed between methods. The only sampling method that reached a stable value was the observational snorkeling transects (around 10 samples) although the seine and methods combination showed less than 2% of difference between resampled prevalence and actual prevalence after 35 samples (Table S17). Indeed, snorkeling transects allow observations on more individuals than the fishing methods, explaining why it reaches stability faster and thus, need fewer sampling effort to measure accurate prevalence. 35 sampling efforts were not exhaustive enough to capture the actual infection landscape prevalence (Table S17). Ultimately, «magic cutoff» sampling effort for estimating accurate prevalence does not exist as it can vary with parasite aggregation level (Poulin, 2007; 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 5). Prevalence estimates varied 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 scale (site-scale) (Table S18). Additional lakes would allow us to confirm whether infection is spatially random, or patterns could not be detected because of our small lake sample size (N=14).

Frequency distributions are a great tool to investigate mechanisms behind spatial occurrence of species (McGeoch & Gaston, 2002). For parasites, frequency occupation distributions have been revised so that host species 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. Accordingly, histograms do no show parasite frequency of species but rather frequency of lakes’ community with corresponding prevalence level. Results showed that frequency distributions differed between sampling methods, some suggesting a bimodal distribution tendency (Figure 5.A, B) and others with no obvious pattern (Figure 5.C, D). While binomial shapes are frequently observed in frequency occupation distributions, some argue that they are an artifact of small sampling scale (Nee et al., 1991). However, it revealed that lake community tend to display a high or low infection prevalence. Presuming that black spot trematodes dispersion is not limited by the final hosts’ movements (birds) across our study area, these high- and low-prevalence groups could present environmental characteristics that favor (or restrict) parasite abundance and transmission to fish hosts. Alternatively, this could mean that when the parasite is present, there are regulating pressures that limit the parasite abundance and transmission. Yet, if the parasites overcome the selective pressure, they become highly prevalent in the fish community. A larger lake sample size might clarify if these patterns of infection dynamics.

Differences in frequency distributions imply that method-biases are also perceptible at lake-scale and that they influence observed prevalence distribution patterns. Accordingly, maps (Figure 5) clearly show that lakes’ prevalence can be over or underestimated according to the method. For example, in lake Pin rouge, the transect method underestimated the community prevalence while in lake Croche, seine net overestimated the community prevalence (see Table S13). This evidence method-biases in field sampling of fish communities. If we take back the previous example, this means that the proportion of infected fishes sampled by seine nets in lake Croche was higher than the proportion sampled by the other methods. The method might have selected fish species with high host specificity, sampled infection hotspots, sampled older individuals with higher infection exposure or sampled active and bold individuals that have a higher infection risk. Our approach revealed method-biases in prevalence estimates, but do not allow us to establish the mechanism behind this causality. 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 7, Table S18). Predictors are important to assess characteristics of infection hotspots and predict extinction risk with growing anthropogenic pressures. Our results evidence some non-linear patterns supporting the use of flexible approach to model empirical (data-driven) infection dynamics. All physico-chemistry parameters (turbidity, temperature, dissolved oxygen, conductivity, and pH) were good predictors of infection prevalence at site-scale (Figure 7.A, B, C, D, E). Comparatively to infection assessment of fish populations and communities by field surveys (or biological samples), water physico-chemistry measurements do not require a lot of equipment, time and are non-invasive to organisms. Thus, water physico-chemistry might be used 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 only 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 (D2 = 84.17%) negative relationship between the macrophyte cover and the prevalence of infection in transect sites (Figure 7.G). Contrary to our findings, heavier infection has been measured in fishes associated with vegetated zones in previous studies (Hartmann & Nümann, 1977; Marcogliese et al., 2001; Ondrackova et al., 2004). On one hand, because cercariae have a short life-span (Combes et al., 1994; Pietrock & Marcogliese, 2003), physical barriers might prevent the free-living stage from encountering a suitable host. Macroalgae have been shown to reduce the number of trematode cercariae (Bartoli & Boudouresque, 1997; Welsh et al., 2014) thus interfering with the parasite transmission to the next host. For instance, Prinz et al., (2009) demonstrated that inert seaweed can obstruct transmission of trematodes cercariae to their mussel host. Similarly, Buss et al., (2022) suggested that microplastic fibers might reduce success of infection by impeding cercariae to encounter their amphibian host. A treatment of complex shell surface of oysters also lead to reduction of 44% of cercariae in a experimental study (Welsh et al., 2014). In short, structural barriers might impede transmission by trapping cercariae, damaging them or consuming time of their short infective window. On the other hand, infected snails releasing cercariae in our lake system might not live or feed on macrophytes, explaining why we did not find a positive relationship between macrophyte cover and prevalence of infection in fishes. Indeed, if the association between macrophyte and snails don’t stand, there is no reason to think that increasing macrophyte would lead to an increase in cercariae release and accordingly prevalence of infection in the fish community. 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 graze on periphyton (Kesler, 1981). Consequently, we would not expect a strong relationship with macrophyte cover.

Our results suggest that community structure is a driving factor of prevalence of the black spot disease at small spatial scale. We found negative relationships with the total fish abundance, the non-host abundance and the Simpson’s diversity index (Figure 7.J, K, L). These results support the 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 transect with higher fish abundance. First, a higher non-host fish abundance might reduce prevalence estimate in fish communities by an unsuccessful infection attempt on a non-host individual, by encountering individuals playing a structural barrier role (explained above) or by inducing a behavior change that reduces encounter rate with fish hosts. For example, Ahn & Goater (2021) showed that abundance of brain worms decreased in fathead minnows (*Pimephales promelas*) when mixed with non-host emerald shinners (*Notropis atherinoides*), but that there was no difference when mixed with other non-host species, suggesting a behavioral change when *P. promelas* and *N. atherinoides* co-occur. This emphasis fish species identity role in dilution effect mechanisms. Accordingly, diverse communities in our system tend to be less infected than communities with dominant species. In our lake system, *L. gibbosus* is a dominant species of littoral communities and the species with the higher infection prevalence across the landscape (Table S19). The presence of competitor species (e.g., *P. flavescens*) might then decrease the overall prevalence in the community by modifying pumpkinseeds sunfish’s behavior that increases encounter with cercariae. Indeed, Dargent et al. (2013) found that 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, local diversity of non-fish organisms could influence the infection in fish. For instance, Lagrue & Poulin, (2015) found a negative association of non-host benthic invertebrates diversity with infection prevalence in second intermediate host. 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 process explaining dynamics of parasitic disease across scale is fundamental to understand future infection and extinction risk in natural systems. Here we investigated prevalence of the black spot disease across multiple scales to skim over sampling effort relationship, sampling method bias, spatial occurrence patterns and environmental predictors. Results suggested that prevalence is distributed in a non-random heterogeneous way in the landscape with small sampling size largely overestimating regional prevalence. We evidenced inconsistent method-induced bias in prevalence estimates at both lake and landscape-scale, sometimes leading to high variation in estimates. Method also influenced sampling effort needed to reach accurate landscape prevalence estimate, the observational snorkeling transects requiring the least. Best predictors of site-scale prevalence were water physico-chemistry and community structure metrics. Our findings suggest encounter dilution effects by host decoy or physical obstruction. However, we observed many non-linearities in the effects suggesting complex interactions that we could not disentangle with our approach. Therefore, we encourage research to deepen knowledge to understand processes linking these predictors to infection parameters for future integration into sampling surveys and conservation plans. In short, we evidenced variations of infection prevalence across spatial scale, sampling effort and sampling methods. Hence, we urge researchers to take these into consideration in the context of comparative studies.