**ABSTRACT**

Detecting the presence of rare species has interested ecologists and conservation biologists for many years. A particularly daunting application of this problem pertains to the detection of non-indigenous species (NIS) as they colonize new ecosystems. Ethical issues prevent experimental additions of NIS to most natural systems to explore the relationship between sampling intensity and the detection probability of a colonizing NIS. Here we examine this question using a recently introduced water flea, *Cercopagis pengoi*, in Lake Ontario. The species has biphasic population development, with sexually-produced 'spring morphs' developing prior to parthenogenetically-produced 'typical' morphs. Thus, this biphasic morphology allows distinction between new colonists (spring morphs) from subsequent generations. We repeatedly sampled Hamilton Harbour, Lake Ontario for the presence of both spring and typical morphs. Probability of detection was positively related to both the number of samples taken and animal density in the lake; however, even highly intensive sampling (100 samples) failed to detect the species in early spring when densities were very low. Spatial variation was greatest when densities of *Cercopagis* were intermediate to low. Sub-sampling, which increased space between adjacent samples, significantly decreased the number of samples required to reach greater, calculated detection probabilities on these dates. Typical sampling protocols for zooplankton have a low probability (< 0.2) of detecting the species unless population density is high. Results of this study suggest that early detection of colonizing, aquatic NIS may be optimized through use of a risk-based sampling design, combined with high sampling intensity in areas deemed most vulnerable to invasion, rather than less intensive sampling at a wider array of sites.

**Keywords**

Biological invasions, colonization, detection, invasive species, non-indigenous species, rare species, sampling intensity.
et al., 2002 and refs. therein) (Fig. 1). Thus, most NIS discoveries occur once they have established and attained ecologically significant densities (Myers et al., 2000; Bax et al., 2002; Inglis et al., 2006). Once species achieve these higher densities, the problem switches from detection to control. Methods that functionally reduce this lag period potentially represent powerful management tools (Fig. 1). The technique used most often to decrease detection thresholds is an increase in sampling intensity (Gotelli & Colwell, 2001; Hortal et al., 2006). The purpose of this study was to quantify the sampling intensity required to detect a newly colonizing NIS in an aquatic habitat and determine how rarity affects the probability of detection.

**Study system**

Ethically, we could not introduce an NIS into a natural system to study the probability of detection in relation to sampling intensity. Consequently, we used an established NIS with a unique life cycle that still allowed us to address our goal. *Cercopagis pengoi* (henceforth *Cercopagis*) is a non-indigenous water flea in the Laurentian Great Lakes. Originating in the Ponto-Caspian region, it was first discovered in North America in the summer of 1998 (MacIsaac et al., 1999). Its cyclic, parthenogenetic reproductive cycle consists of two distinct morphological forms that represent progeny from sexual and asexual reproduction (Laxson et al., 2003). The species overwinters as sexually produced ‘resting eggs’ in lake sediment. In spring, these eggs hatch into individuals possessing a short, blunt caudal appendage (Simm & Ojaveer, 2006). Asexually produced offspring of this ‘spring’ morph have the more ‘typical’ caudal appendage that is longer and hooked near its terminus (Grigorovich et al., 2000). The difference between morphs is so profound that they were originally classified as distinct subgenera: *C. (Apagis) ossiani* (the spring morph) and *C. pengoi* (the typical morph). This morphological difference makes this the ideal species to address the objectives of this study. No adult individuals overwinter in the water column, thus colonization by the blunt-tailed individuals from resting eggs in spring is equivalent to a new invasion. In this study, we sample intensively before, during and after the period during which emergence was expected to occur.

**METHODS**

On six dates from 30 April to 28 August 2007, we sampled Hamilton Harbour, Lake Ontario for *Cercopagis* with hauls of a vertical zooplankton net (750 μm mesh, 50 cm diameter, 150 cm length). *Cercopagis* does not display a typical diel, vertical migration. We used a vertical haul, as opposed to a potentially more efficient oblique haul (Hayes et al., 2005), given the vertical positioning of this species in the water column (i.e. sub-5 m to a maximum of 20 m; Laxson et al., 2003). On each date, 100 samples were taken from 20 m to the surface within a specific 1 km² sampling grid, comprising 100 GPS-derived points (±10 m) spaced 100 m apart. On 28 August, only 80 samples were taken because of inclement weather. After each sample, the net was rinsed to avoid subsequent sample contamination. All samples were preserved in 70% ethanol and stored in the laboratory at room temperature until processed.

To determine *Cercopagis* presence, each sample was sieved through 40 μm mesh and rinsed into a sectioned Petri dish, and counted in total. Presence/absence data for each sampling date was bootstrapped without replacement (1000 iterations) to generate sample-based rarefaction curves relating sampling intensity to the probability of finding at least one *Cercopagis* (Gotelli & Colwell, 2001). *Cercopagis* density and thus the points within our sampling grid where we detected *Cercopagis* varied with sampling date (see Results). To determine the potential effects of sampling intensity (number of samples) and our sampling method (vertical hauls taken within a regular sampling grid) for detecting *Cercopagis*, additional rarefaction curves were generated from two subsets of samples spaced 200 and 300 m apart, respectively, for each sampling date. There were four different arrangements of sub-samples equally spaced 200 m apart. Thus, we used the mean detection probability from all four sampling arrangements for any further analyses. There was only one possible arrangement for sub-sampling points equally spaced 300 m apart.

Effects of *Cercopagis* density and number of samples taken on the relationship between sampling intensity and detection probability were assessed by fitting a type II functional response curve to sample number and detection probability data for each sampling date. The linear form of a type II functional response identifies the rate of saturation (slope) and the point of saturation or asymptote (y-intercept) for any saturating curve (Fan & Pettit, 1994). We then compared these parameter estimates for the full sampling intensity between dates and between sub-samples within each date with Tukey-Kramer linear contrasts.

To test for spatial patterns of *Cercopagis* presence/absence, we used kriging to predict variation in detection probability across our sampling grid (Legendre & Fortin, 1989; Fortin et al., 2002). We estimated the variance between sample points with semivariograms from the most parsimonious of five covariance models (Gaussian, exponential, spherical, power or linear). Covariance model
appropriation was determined by selecting the model with the weighted sum of squares closest to zero. Spatial analyses were only performed on the full sampling data set as a result of sampling size limitations (200 m spacing, $n = 25$; 300 m spacing, $n = 16$; Fortin et al., 1989; Legendre & Fortin, 1989). All analyses and comparisons were performed with R v. 2.7.1 (R Development Core Team, 2008).

**RESULTS**

*Cercopagis* was detected on four of six sampling dates. No individuals were found on the two earliest sampling dates (30 April or 15 May), whereas on the third date (29 May) 3 of the 100 samples contained *Cercopagis*. Two of these three individuals were of the spring morph, indicating the initial generation. On the following sampling dates, detection frequency progressively increased from 25% to 91% until the final sampling date when detection frequency was again quite low (5%; 28 August).

The total number of individuals found on any one sample date varied from 0 to 217, with a density range of 0.00 to 0.55 individuals m$^{-3}$. The highest density of *Cercopagis* (0.55 ind. m$^{-3}$) was observed on 19 June, and the population declined thereafter (28 August; 0.03 ind. m$^{-3}$). Thus, our sampling period spanned the emergence, peak density and decline phases for *Cercopagis* in Hamilton Harbour.

Comparison of rarefaction curves from the full sample sets (100 m spacing) suggests that as *Cercopagis* density increased, the slope of the rarefaction curves increased concomitantly (Fig. 2). All slope estimates from fitted type II functional response curves were significantly different across dates; however, there were no detectable differences in the probability saturation points, except the first two dates when no *Cercopagis* was detected (Table 1).

The resulting rarefaction curve from 30 April and 15 May (combined) had a flat or zero slope and no saturation point (Fig. 2a). *Cercopagis* was detected on 29 May but at an extremely low density (< 0.01 ind. m$^{-3}$) (Fig. 2b), generating a rarefaction curve with a shallow slope and a saturation point greater than 1 (y-intercept > 1; Table 1). This suggests more than 100 samples were needed to reach a detection probability of 1. Density increased to (0.09 ind. m$^{-3}$) on 12 July. The slope of this curve was significantly greater than that of 29 May (Table 1) and saturation was reached after c. 15 samples, with a probability asymptote of 1 after c. 35 samples (Fig. 2c). The highest *Cercopagis*
density was observed on 19 July (0.56 ind. m$^{-3}$). This density produced a very steep curve (slope 0.829), saturated after only four samples and reached a detection probability of 1 at only six samples (Fig. 2d). The shape of the rarefaction curve on the last sample date, 28 August, was similar in general shape to that for 29 May, but the density was approximately three times greater (0.02 ind. m$^{-3}$). Additionally, the saturation point for 28 August was close to that of 12 June, but the slope was shallower (Table 1; Fig. 2). This shallow slope greatly increased the number of samples (∼60) required to reach saturation and the curve never reached a probability of 1 (max. 0.983), although only 80 samples were acquired on this date.

We sub-sampled the data sets for each date to increase distance between adjacent samples. Doing so only affected the relationship between mean detection probability and sample number when distance between adjacent samples was increased to 300 m. As observed with the full data sets, there was no difference in the calculated saturation points (Tukey-Kramer contrasts, $P > 0.05$). There was no difference in saturation rates between sub-sampling at 200 m and 100 m for any date, nor any difference between any sub-sampling data sets for 19 June (all Tukey-Kramer contrasts, $P > 0.05$; Fig. 2). However, increasing inter-sampling distance from 100 m or 200 m to 300 m on 29 May (Tukey-Kramer contrasts; 100 m to 300 m, d.f. = 114, $P < 0.0001$; 200 m to 300 m, d.f. = 39, $P < 0.0001$), 12 June (Tukey-Kramer contrasts; 100 m to 300 m, d.f. = 114, $P < 0.05$; 200 m to 300 m, d.f. = 39, $P < 0.05$) and 28 August (Tukey-Kramer contrasts; 100 m to 300 m, d.f. = 114, $P < 0.0001$; 200 m to 300 m, d.f. = 39, $P < 0.0001$) significantly decreased the saturation rate of the curves (Fig. 2b–e).

Fitted semivariograms were consistent with trends established in the parametric analysis; spatial variation was greater at intermediate densities of Cercopagis. Although all standard error surfaces were rather flat and well behaved (Fig. 3e–h), semivariograms suggested spatial patterns within our data were non-stationary (Table 2; Fig. 3). A linear semivariogram best fit the data for 29 May and 12 June, whereas a Gaussian model fit best on 19 June and 28 August (Table 2). The greatest spatial variation was observed on 12 June, and generated a flat semivariogram (Fig. 3b,f), indicating no spatial structure at the scale of our sampling. Fitted semivariograms for 29 May, 19 June and 28 August had no discernible sill (estimated for 19 June and 28 August, Table 2), suggesting all semivariograms exhibited a nugget effect (Table 2; Fig. 3).

Overall patterns suggest that as relative density of Cercopagis increased, there was a proportional increase in the probability of its detection (Fig. 4). At very low density (e.g. 29 May), the probability of detection never exceeded 0.2, even with 100 samples. By contrast, when Cercopagis was present at high relative density, detection probability always exceeded 0.8 even when few (<5) samples were taken. We utilized high sampling intensity throughout the study (80–100 samples), thus the density of Cercopagis appeared to ultimately determine detection probability.

### DISCUSSION

Detection probability was strongly affected by both sampling intensity and Cercopagis abundance in Lake Ontario. Many previous studies have determined an analogous relationship between species richness and sampling intensity (e.g. Caley & Schluter, 1997; Rohr et al., 2006; Soria-Auza & Kessler, 2008). As the number of samples increases, or as the density of a species increases in a specific area, there is a concomitant increase in detection probability. However, our study provides novel, empirical insights into the dynamics of detecting NIS at the colonization stage. Even with sampling intensity 20 times higher than that typically employed by plankton ecologists, the probability of detecting emerging Cercopagis was never higher than 0.20 when the species was present at low abundance. This finding has major implications for the design of monitoring programs that prioritize early detection of NIS in high-risk habitats, such as ports with significant ballast water discharge (e.g. Drake & Lodge, 2004) or areas where large volumes of wood dunnage are dumped or shipping containers unloaded (e.g. Work et al., 2005).

The purpose of this study was to determine the sampling intensity required to detect a colonizing NIS. We were able to ethically pursue this question by exploiting the seasonal life cycle characteristics of Cercopagis, which effectively re-colonizes the system each spring. Cercopagis was not detected on either of our first two sampling dates (Table 1; Fig. 2a), indicating that sampling commenced before Cercopagis began to emerge, or that emerged individuals were present a sub-threshold levels. Detection of the spring morph at a very low frequency (3 of 100 samples) on our third sampling date indicates that our sampling protocol was sufficient to capture the early colonization period of Cercopagis in Hamilton Harbour (Table 1). The spring morph of Cercopagis was detected only on the first date the species was detected in our samples; thereafter all collections contained only the typical morph (Table 1), suggesting sampling spanned the colonization, establishment and decline phases.

Although we observed typical seasonal wax and wane patterns for Cercopagis, the maximum density achieved was orders of magnitude lower than that reported in previous studies on Lake Ontario (e.g. Makarewicz et al., 2001; Laxson et al., 2003). Introduced populations can exhibit a boom and bust dynamic
after a period of initial invasion (Simberloff & Gibbons, 2004), and Cercopagis densities in Lake Ontario, although variable, have exhibited a declining trend since first being reported in the lake in 1998. This trend could be attributable to fish predation, as numerous planktivores including alewife (Alosa pseudoharengus) and rainbow smelt (Osmerus mordax) prey upon Cercopagis in the Great Lakes (Bushnoe et al., 2003). Alternatively, the low observed density may have resulted from a lack of sampling during July and early August when peak Cercopagis abundance is typically reported in Lake Ontario and other regions where it has been introduced (Krylov et al., 1999) or may have been limited by the sampling methods and design we employed. Ultimately, however, these comparatively low densities were useful, given our objectives.

The probability of detecting Cercopagis was strongly influenced by sampling intensity, but even more so by population density.

Figure 3  Summary of spatial analyses displaying semivariograms (a–d), and kriged variance (SE) surfaces (e–h) with reference to Cercopagis abundance plots from the full sample (100 m) data set (i–l) for all dates where Cercopagis was detected. Abundance increases with lighter colour within each abundance plot.
and the location of *Cercopagis* within the area we sampled (Figs 2 and 3). Sampling intensity across this study was well maintained, even with only 80 samples being collected on the last date. On dates when *Cercopagis* was detected (Fig. 2b–e), there was no difference in the point at which any of the rarefaction curves reached saturation. Yet, significant differences between the slopes of all curves indicated that a different number of samples were required to achieve the same probability of detection.

A striking contrast emerges when comparing saturation dynamics from the two June dates with those of the May and August dates. Comparatively few samples were required to achieve a saturating (100%) probability of detecting *Cercopagis* (samples 15 and 4, respectively; Fig. 2c,d) on the June dates, whereas saturation was never achieved on the latter dates when full sampling intensity (100 or 80 samples) was applied. These latter samples are more characteristic of the conditions and problems encountered when attempting to detect a colonizing NIS. Sub-sampling the data changed this trend, however, by increasing the rate of saturation when samples were spaced 300 m apart. When this was done, 100% detection probabilities were estimated at c. 20 samples, even when *Cercopagis* was at lower densities (e.g. Fig. 2b,e).

This variation in detection is probably a result of the greater spatial variation in *Cercopagis* at intermediate to low densities (Fig. 3). We chose to use a systematic sampling grid with the purpose of intensely and repeatedly sampling a finite area. Previous studies have noted that systematic, regular sampling is the most effective design for detecting presence of NIS (Hirzel & Guisan, 2002; Rew *et al.*, 2006). In a spatial context, systematic sampling is easier to employ and does not require previous knowledge of the area to be sampled (Fortin *et al.*, 1989). However, they may lead to oversampling non-informative areas and undersampling more informative areas, as opposed to random or stratified random designs, which can be more useful in detecting significant spatial structures (Fortin *et al.*, 1989; Legendre & Fortin, 1989). The use of a systematic sampling design is likely the main factor leading to the apparent non-stationarity, where the processes determining spatial pattern change with spatial scale, found in our spatial analyses. Combining the results of our analyses indicates that it is the spatial arrangement rather than the number of sampling points that may most influence detection of colonizing NIS (e.g. Fortin *et al.*, 1989). That is, with an increased probability of detection from samples taken 300 m apart and our spatial analyses indicating some variation at even small spatial scales (nugget effect), sampling a larger area with samples spaced farther apart may have been a more sensitive approach to detecting this colonizing NIS.

In reality, standard zooplankton sampling protocols usually entail collecting five or less samples within a single system on a given date. On dates when *Cercopagis* was present at low abundance, this sampling intensity would generate a detection probability of 13% or less. Most inspection and environmental monitoring programs require the maintenance of a 95% detection probability for known pest species with an infestation frequency of 10% (Venette *et al.*, 2002). The only date for which we reached a 95% detection probability with only five samples was on 19 June, when *Cercopagis* density was highest (Figs 2d and 3c). Less than 10% of the relative density range observed from this study would meet most inspection requirements using standard sampling protocols (Fig. 4).

Countries including Australia, Canada, New Zealand, USA and a consortium of countries in Africa have placed early detection as a top NIS priority (ANZEC 2001; Hellström *et al.*, 2003; Environment Canada, 2004; Moore, 2005; USDA APHIS 2007). Despite this objective, the economical and logistical challenges of developing and implementing such a program are often seriously constrained. Government agencies are challenged by a trade-off between monitoring numerous at-risk habitats and the sampling intensity required to adequately census these areas. Based on the results of this study, this trade-off is most influenced by the density at which a colonizing NIS can be detected; spatial variation decreased with an increase in density and there was no effect of sample spacing when *Cercopagis* density was greatest. At relatively high densities (e.g. Fig. 4, > 0.5), it is more efficient to sample a greater number of sites at a lower sampling intensity. At higher densities, the marginal benefit of more intense sampling is reduced as detection probability varies little between c. 10 to 20 samples (Fig. 4). At relatively low densities (Fig. 4, < 0.25), it is more profitable to target fewer sites but sample each more intensively and with a more spatially-minded design (detection prob. < 0.2 with fewer than 15 samples, Fig. 3). This scenario is clearly based upon our study of a single species in a single system. Development of strategic sampling plans require understanding the species-specific nature of the probability of detection in relation to population density and sampling intensity as conspicuousness, aggregation behaviours, habitat utilization and landscape history are factors that can impact the detectability of invading NIS (Higgins *et al.*, 1999; Rejmánek, 2000; Inglis *et al.*, 2006; Tobin, 2007). We expect that similar patterns will occur for most colonizing, aquatic NIS.

Initial detection of an NIS is the most difficult but arguably most important aspect of NIS management (Rejmánek, 2000; Hulme, 2006). With vector-based approaches to estimating NIS propagule pressure and environmental niche modelling to predict which geographical regions are most susceptible,
high-risk habitats are being more clearly identified (e.g. Herborg et al., 2007; Meentemeyer et al., 2008). Once high-risk habitats are identified, one alternative to the application of intensive sampling programs is the use of species-specific molecular markers. These methods seek to identify the presence of species-specific DNA sequences from bulk samples. For example, molecular markers have been utilized to detect an invasive kelp in port water samples (Hayes et al., 2007), American bullfrogs in Europe (Ficetola et al., 2008) as well as a pestiferous moth species in the sub-Antarctic (Chown et al., 2008). Another technique used to decrease detection thresholds, by artificially increasing the local density of a target species, is the deployment of pheromone-baited traps (Brockerhoff et al., 2006; Tobin, 2007). After NIS detection, other novel approaches can be employed to assess their spatial range (Barnett et al., 2007). For example, large areas of terrestrial habitat can be sampled using remote sensing spectroscopy. Spectral images of vegetation patches can be taken from above to indicate distinct spectral signatures for both native and non-indigenous cover (Lass et al., 2005). Although highly effective, each of these aforementioned methods requires technical sophistication and knowledge of potential invaders. Methods for detecting unknown and newly colonizing species require further development.

Employing a combination of predictive and sampling methods is the most effective prevention and control of NIS (Rejmánek, 2000; Hulme, 2006). Ecologists designing programs for early detection of NIS must be aware of sampling intensity and spatial design issues to develop strategies accordingly. A minimum of 65 samples was required to achieve a 95% detection probability when Ceropagis was present at low abundance in Lake Ontario. This sampling intensity is not feasible for active monitoring programs. If, however, monitoring programs were tailored to areas where vector activity is most acute, then comprehensive sampling designs would become more logistically feasible. With increased feasibility of accurate and sensitive monitoring techniques, the reality of preventing NIS establishment as opposed to decreasing their negative impacts may become more tangible.

ACKNOWLEDGEMENTS

We thank Colin Van Overdijk, Dr Francisco Sylvester, Jaclyn Brush and Susan Dennard for field assistance. Kristen Patterson aided in processing samples in the lab. Much appreciation is given to Drs James Forester and Phil Graniero for comments regarding the spatial analyses and Dr Jim Muirhead regarding bootstrapping and rarefaction curves. This manuscript was significantly improved by comments from three anonymous reviewers. We acknowledge financial support from the Canadian Aquatic Invasive Species Network, NSERC, and a DFO Invasive Species Research Chair to HJM.

REFERENCES


cladoceran *Cercopagis pengoi* from Lake Ontario, Baltic Sea and Caspian Sea. *Archiv für Hydrobiologie*, 149, 23–50.


Editor: David Richardson