Main Manuscript for

DNA metabarcoding reveals metacommunity dynamics in a threatened boreal wetland complex.

Bush A.*[1,2], Monk W.A. [3], Compson Z.G. [1,4], Peters D.L. [6], Porter T.M. [5,7], Shokralla S. [7], Wright M.T.G. [7], Hajibabaei M. [7**], Baird D.J. [1**]

1. Environment and Climate Change Canada @ Canadian Rivers Institute, Department of Biology, University of New Brunswick, Fredericton, NB, Canada
2. Lancaster Environment Centre, University of Lancaster, Lancaster, UK
3. Environment and Climate Change Canada @ Canadian Rivers Institute, Faculty of Forestry and Environmental Management, University of New Brunswick, Fredericton, NB, Canada.
4. Centre for Environmental Genomics Applications (CEGA), St. John's, NL, Canada
5. Great Lakes Forestry Centre, Natural Resources Canada, 1219 Queen Street East, Sault Ste. Marie, ON, Canada
6. Watershed Hydrology and Ecology Research Division, Environment and Climate Change Canada, University of Victoria, Victoria, BC Canada
7. Centre for Biodiversity Genomics and Department of Integrative Biology, University of Guelph, 50 Stone Road East, Guelph, ON Canada

* Corresponding author email**:** alex.bush@lancaster.ac.uk
** Joint last authors

ORCID ID: 0000-0002-0679-6666

Classification: Biological Sciences / Ecology

Keywords: occupancy, detectability, taxonomic resolution, stochasticity, power analysis

This PDF file includes: Main Text, Figures 1 to 4 and 0 Tables

Abstract

The complexity and natural variability of ecosystems present a challenge for reliable detection of change due to anthropogenic influences. This issue is exacerbated by necessary trade-offs that reduce the quality and resolution of survey data for assessments at large-scales. The Peace-Athabasca Delta (PAD) is a large inland wetland complex in northern Alberta, Canada. Despite its geographic isolation, the PAD is threatened by encroachment of oil sands mining in the Athabasca watershed, and hydroelectric dams in the Peace watershed. Methods capable of reliably detecting changes in ecosystem health are needed to evaluate and manage risks. Between 2011 and 2016, aquatic macroinvertebrates were sampled across a gradient of wetland flood frequency, applying both microscope-based morphological identification, and DNA metabarcoding. By using multi-species occupancy models, we demonstrate that DNA metabarcoding detected a much broader range of taxa and more taxa per sample compared to traditional morphological identification, and was essential to identifying significant responses to flood and thermal regimes. We show that family-level occupancy masks high variation among genera, and for the first time, quantify the bias of barcoding primers on the probability of detection in a natural community. Interestingly, patterns of community assembly were near random, suggesting a strong role of stochasticity in the dynamics of the metacommunity. This variability seriously compromises effective monitoring at local scales, but also reflects resilience to hydrological and thermal variability. Nevertheless, simulations showed the greater efficiency of metabarcoding, particularly at a finer taxonomic resolution, provided the statistical power needed to detect change at the landscape scale.

Significance Statement

Too often, ecological monitoring studies are designed without understanding whether they have sufficient statistical power to detect changes beyond natural variability. The Peace-Athabasca Delta is North America's largest inland delta, within a World Heritage area, and is currently threatened by human development. By using multi-species occupancy models we show that the wetland macroinvertebrate community is highly diverse, and spatial and temporal turnover are so high that composition is nearly random, emphasising stochastic processes of assembly. Using DNA metabarcoding, our study detected more taxa, both overall and per sample, than traditional morphology-based sample processing; increasing our power to detect ecosystem change. Improving data quality and quantifying error are key to delivering effective monitoring and understanding the dynamic structure of the metacommunity.

Main Text

Introduction

Tackling the global loss of biodiversity (1) is hindered by a lack of basic biological information needed to guide sustainable management strategies (2). Despite legal protections, freshwater ecosystems are increasingly degraded by multiple stressors (3). In addition, the quality and volume of data collected by monitoring programs often fails to support evidence-based management decisions (4-6). Here, we demonstrate how DNA metabarcoding can resolve challenges faced by traditional monitoring, alter our perspectives on ecosystem dynamics, and improve our understanding of natural variation and sampling error, supporting evidence-based decision making.

DNA barcoding uses short genetic sequences to identify individual taxa. By contrast, DNA metabarcoding supports simultaneous identification of entire assemblages via high-throughput sequencing (7, 8). Using metabarcoding for ecosystem monitoring provides an opportunity to identify organisms in bulk samples at a high taxonomic resolution consistently and accurately ("Biomonitoring 2.0"; 9). The accuracy, consistency, and resolution of taxonomic identification remains a constraint for many biomonitoring programs that must trade-off data quality to make

2

assessment protocols rapid and cost-effective (10). Aquatic macroinvertebrates exemplify this challenge, as their diversity of forms and functions are sensitive to multiple drivers of ecosystem condition. Thus, ecosystem degradation can be identified based on changes in assemblage composition due to environmental filtering (5). Despite decades of development, the challenges associated with traditional methods of sample processing limit inference of biomonitoring programs to gross status classifications (e.g. 11). Metabarcoding presents an opportunity to describe community composition more accurately and consistently, supporting more effective and informative biomonitoring (12, 13).

The Peace-Athabasca Delta (PAD) in northern Alberta, Canada (Figure 1; and 14). The PAD is North America's largest inland delta (approximately 6,000 km^2), and is located at the confluence of the Peace and Athabasca Rivers, consisting of hundreds of lakes and wetlands that become connected during flood events, particularly when spring snowmelt leads to ice-jams (15, 16). The PAD is a Ramsar wetland, protected within Wood Buffalo National Park, a UNESCO World Heritage site. Nonetheless, there have been concerns that the PAD could be affected by upstream developments, including current and proposed hydroelectric dams on the Peace River, continued expansion of oil sands mining on the Athabasca River to within 30km of the park boundary, and climate change (17). Assessing how such factors influence the integrity of a natural wilderness is made more challenging by the paucity of biological surveys that have been conducted, and the logistics of working in such a remote region. To gain a better understanding of the PAD's ecology, rapid assessments of aquatic macroinvertebrates have been conducted in since 2011 to establish a baseline understanding of the ecosystem's diversity and structure (14, 18). Importantly, while surveys have followed established protocols from the Canadian Aquatic Biomonitoring Network (hereafter CABIN)(19), samples were processed using both traditional and DNA metabarcoding approaches, allowing us to test the power of each approach to support environmental management of the delta for the first time.

Sampling error is a ubiquitous feature of any ecological survey, irrespective of the methodology, and of particular concern is the frequency of false absences (20). Depending on the covariance of species' detectability with other environmental characteristics, models can be structurally biased and their confidence overestimated (21). Although imperfect detection is very common, and freshwater biomonitoring protocols have a long history of standardization to maintain comparability (5), there are few examples of research explicitly quantifying the nature of sampling error (e.g. 22). Instead, variability due to sampling error is usually combined with that from natural sources (i.e. as "noise"; 23). An alternative is to specify the likelihood of detection (the observation process model) and simultaneously correct our estimates of species occurrence (the ecological state model) within a single hierarchical framework (24). In this study, we employed multi-species occupancy models (MSOMs; 25, 26) to account for the effects of imperfect detection on estimates of macroinvertebrate diversity, drawing upon data from six years of macroinvertebrate surveys in the PAD. We quantify the efficiency with which the macroinvertebrate community can be surveyed using both traditional morphological identification and DNA metabarcoding, and demonstrate that these approaches make a qualitative difference to our view of the metacommunity is structured, to the efficiency of monitoring, and consequently our power to detect change (27).

Results

A key difference between our sampling approaches was that the standard CABIN wetland protocol (19) provided estimates of relative abundance based on counts from a subset of each sample, whereas sequences identified using DNA metabarcoding were converted to presence-absence data (13, 28). In addition, CABIN identified 74 families based on morphological features, but metabarcoding could identify 109 families, as well as 263 genera (see SI Appendix 1, Fig. S1.6). As a result, we trained four hierarchical MSOMs for each data type: 1) counts of macroinvertebrate families from CABIN (CABIN *Fcount*), 2) the presence-absence of macroinvertebrate families from

CABIN (CABIN *Fpa*), 3) the presence-absence of macroinvertebrate families from DNA data (DNA *Fpa*), and 4) the presence-absence of macroinvertebrate genera from DNA data (DNA *Gpa*). Although metabarcoding can discriminate among taxa at even finer resolution (i.e. species), given the prevalence was lower than the prevalence of genera and the available sample size, we did not feel the detectability and occupancy of those taxa could be estimated reliably.

Occupancy and detectability
The CABIN *Fcount* model predicted total abundance was dominated by four taxa (two Chironomidae subfamilies, Oligochaeta and Planorbidae), but also suggested that almost all taxa were present everywhere within the PAD (i.e. site occupancy ≈ 1), with no environmental covariates retained in the final model. This scenario is plausible, but if we apply the predicted probabilities of detection and same survey effort (number of individuals counted), and assume taxa are sampled at random from the pool of individuals, the CABIN *Fcount* model suggested we should have observed 38 taxa on average instead of 18. Non-random aggregation of individuals is typical of ecological communities (29), and may be why the model appeared to be misspecified.

In contrast to the count-based model, the presence-absence models all suggested taxon site-occupancy was below 1 (Fig. 2, SI Appendix 1, Fig. S1.10), although the "U-shaped" form of the hyper-parameters in Fig. 2a and 2c was an artefact of the bounded distribution (29). The CABIN *Fpa* model estimated the probability of detecting macroinvertebrate families was lower than the DNA *Fpa* model (Fig. 2b right-skewed relative to 2d,e; see also Figure 3). Models must balance the expected occupancy to fit with the detections, and probability of detection made in each survey, and the CABIN *Fpa* model therefore also predicted higher occupancy than the DNA *Fpa* model (Fig. 2a left-skewed relative to 2c). The differences in occupancy and detectability of specific families were not associated with prevalence, although many taxa were not recorded by both approaches, and therefore cannot be compared (red points in Fig. 3; see SI Appendix1 for detail). In addition, detectability using DNA metabarcoding is intrinsically linked to the genetic primer used (30), and the importance of primer bias is well known from mock laboratory samples (e.g. 31). Here we show biases in detectability can be quantified as part of the observation model, either at the community-level (Fig. 2d and 2e), or for individual taxa (SI Appendix 1, Fig. S1.11). Lastly, neither the CABIN *Fcount* or *Fpa* model retained environmental variables to explain changes in occurrence, whereas both DNA occupancy models did so consistently. The covariates retained were 1) the frequency of spring and summer floods (i.e. connections between the wetland and river), 2) time since the ice melt, and 3) maximum water temperature prior to each survey. Responses to environment at the community-level were almost neutral (SI Appendix 1, Fig. S1.13), and the posterior distribution of coefficients differed from zero for only a minority of taxa (SI Appendix 1, Fig. S1.14), but their inclusion in the model suggests the high inter-annual turnover (SI Appendix 1, Fig. S1.7) may be explained in part by deterministic factors.

Alpha, beta and gamma diversity
Recognizing that imperfect detection is commonplace in ecological surveys, it follows that regional (gamma diversity, SI Appendix 1, Fig. S1.8) and local (alpha diversity, SI Appendix 1, Fig. S1.9) diversity is routinely underestimated. As the CABIN *Fcount* model effectively assumed alpha and gamma diversity were equal, it estimated that only two families were likely to have gone undetected in the metacommunity. Conversely, the CABIN *Fpa* model estimated ~20 families were missed (i.e. γ = 95), a 28% increase on the observed total. Interestingly, this estimate was still short of the richness observed using metabarcoding (n=109), and based on the distribution of detection probabilities, the DNA *Fpa* and *Gpa* models estimated the metacommunity could potentially contain 130 families and 360 genera, a 19% and 37% increase (SI Appendix 1, Fig. S1.8).

Although imperfect detection always underestimates richness, its effect on the observed compositional dissimilarity between sites (beta diversity) is less predictable. The observed pairwise-dissimilarity of samples consistently exceeded 40%, both within and between years, with no consistent increase over time (SI Appendix, 1 Fig. S1.7). Our analysis showed that compositional

turnover in the CABIN dataset was overestimated, whereas for the DNA models the corrected and observed dissimilarities were similar (SI Appendix, 1 Fig S1.15), although temporal turnover (i.e. inter-annual, within-site dissimilarity) was marginally overestimated by the DNA dataset. This implies that although metabarcoding underestimated alpha diversity at each site, the proportions of the taxa missed that were shared or unique to site-pairs were similar. Finally, one predictable aspect of turnover is that as the taxonomic resolution is increased, sub-taxa are on average less prevalent than their parental ranks (see SI Appendix, 1 Fig. S1.12a), typically harder to detect (presumably because they are also less abundant than parental ranks), and therefore dissimilarity among sites at the genus-level was 7% higher compared to the family-level.

<u>Power analysis</u>
The power to detect statistically significant changes depends on the strength of the ecological signal relative to other natural variability, as well as the efficiency with which we can accurately describe ecological state, factors directly related to taxonomic resolution, and detectability (27). We simulated the PAD metacommunity based on a fitted distribution of occupancy and estimated gamma diversity to represent its baseline condition, and then took subsamples that reflected the observed biases in each sampling approach. Note that the true state and behaviour of the system are unknown, and underlying processes were instead inferred by the MSOMs after quantifying observation biases. Human impacts that might affect the PAD system in the future were also unknown, and this analysis therefore aimed to identify our power to observe a generalized stressor effect. To keep the process model consistent, we based simulations on the most detailed DNA *Gpa* model, and then aggregated taxa to higher ranks to compare power among sampling approaches. A complete description of the simulation and power analysis is provided in SI Appendix2.

A natural consequence of high, near-random, background variation in composition is that degradation of a wetland site would need to be severe to raise concerns. Instead, it is more effective to measure when there is a shift away from our expectation of the PAD metacommunity aggregated across sites (i.e. changes in occupancy of many taxa). Even so, based on the high natural variability of the PAD, the survey effort needed to confidently detect shifts in occupancy in any year would be prohibitive. As a result, we considered a monitoring system to be adequate if significant differences in composition were detected within 2 years (at least 50% of the time, SI Appendix 2, Fig. S2.4). Our results demonstrated that our power increased: 1) as the number of sites sampled increased (but the rate of increase declined beyond 8-10 sites per year); 2) with DNA metabarcoding compared to CABIN sampling, and with genus- compared to family-level data; and 3) if we sampled sites multiple times (but gains depended on the number of sites and sampling approach) (Fig. 4). Statistical power also varied by stressor type because metacommunity shifts were readily apparent if the stressor impacted prevalent taxa, whereas changes were challenging to observe if prevalent taxa were also tolerant. The relationship between taxon occupancy and their sensitivity to a stressor was therefore most influential when sample sizes, and hence our power to detect rare taxa, were low (SI Appendix 2, Fig. S2.6).

Discussion

The Peace-Athabasca Delta represents one of Canada's national biodiversity treasures. Yet multiple external pressures, including the development of oil sands, hydroelectric power, wildfires and climate change are potentially affecting biodiversity through modification of natural physical processes in the area and threaten its World Heritage listing (17). Our study demonstrates that the PAD is an immensely rich habitat, including over 25% and 20% of all aquatic macroinvertebrate families and genera recorded by the CABIN national biomonitoring program (32). This total may still underestimate the total diversity present, and we demonstrate the importance sampling errors can have for modelling this community. Communities exhibited near-random patterns of spatial and temporal turnover, a property rarely observed in freshwater systems (33). As a result, impacts on the wetland macroinvertebrate community are difficult to establish at local scales because occurrence is weakly related to environmental factors and site

5

composition can fluctuate rapidly over time (SI Appendix 1, Fig. S1.7). Properties of the metacommunity must therefore be aggregated across sites, and directional shifts can only be inferred when dissimilarities are unlikely to be explained by stochastic differences in our null baseline model (34). Our analysis shows that detecting a decline in metacommunity condition would depend on both sample size and stressor type, and that further changes to sampling design may be required to detect change earlier or at specific locations of concern.

The most significant finding of this study was the value added to biomonitoring data generated by DNA metabarcoding of bulk community samples. Our analysis supports previous studies that have shown the breadth and resolution of taxonomic information achievable with metabarcoding (e.g. 14, 35). Clear differences in occupancy and detectability profiles with metabarcoding (Fig. 3) influence our description of baseline reference conditions (36, 37). Further differences in estimates of occupancy with increasing taxonomic resolution (SI Appendix 1, Fig. S1.12) may also indicate differential environmental responses (38, 39). We did not find evidence to suggest the count data ('relative abundance') in CABIN samples were necessary to detect changes in ecological structure. In fact, only the presence-absence DNA metabarcoding models identified significant relationships with the major environmental covariates of this region (16). These effects could be estimated precisely because detectability, and thereby sampling efficiency, was higher for so many taxa using DNA metabarcoding (13). We also used the occupancy model framework to compare detectability of each taxon with different primers, a more robust measure of their complementarity than lists of taxa observed. Quantifying detectability is vital to making the results of this study comparable to others with varying protocols, and this approach could be used to refine and select complementary primers (28, 31). Crucially, DNA metabarcoding, particularly at genus-level, substantially improved our power to detect ecosystem-scale changes compared to traditional CABIN sampling (Fig. 4, SI Appendix 2). Extending this approach to the species-level could improve overall power further still, as long as a sufficient number of species have a similar probability of detection as their parent genera.

A second significant outcome was the importance of explicitly considering imperfect detection. Practitioners are well aware of sampling differences (e.g. 40), but have typically focused on how those errors propagated to aggregated metrics, rather than explicitly quantifying the sources of uncertainty (23). Hierarchical occupancy models accommodate irregularly sampled data, estimate community properties (that extend inference to rare taxa), and allow straightforward biological interpretations of those parameter estimates (41). There have been few examples of hierarchical models accounting for detectability in freshwater ecology (e.g. 42, 43), despite studies showing it can bias our interpretation of taxonomic, functional, and phylogenetic diversity at the community-level (e.g. 44). Given the high prevalence of false absences it is not surprising occupancy models are becoming commonplace for analysing eDNA data (45), although it appears multi-species models are still rare (46). Importantly, what these and other studies have shown is that the time and expense of adding replicate samples may be the most efficient way to improve the statistical power of a study (24, 47). Inferences about the number of taxa missed in the metacommunity naturally carry some uncertainty (48), but by acknowledging imperfect detection, risks can be quantified, and decision makers' overall efficiency can be improved. This analysis minimises the likelihood of management agencies responding to a false signal of degradation (Type 1 error), and identifies how to optimize survey design to ensure we have the necessary power to detect a desired degree of change (Type 2 error) (27, 49).

Although our analysis provided evidence of environmental filtering, the distribution of beta diversity was equivalent to that expected from random assembly, suggesting that the metacommunity was operating in a quasi-neutral manner at the scale of our analysis (50). Quasi-neutral dynamics are expected to be commonly observed in taxon-rich communities, but given the high degree of landscape connectivity, we would expect mass effects, rather than dispersal limitation, to underlie the low habitat specificity of the community. Our dataset was insufficient to identify which mechanisms underlie metacommunity assembly, because the same patterns of

turnover may be the result of different assembly processes (51). While further studies could reduce this uncertainty, currently models of coexistence that combine stochasticity with niche theory may be the most suitable option to explain the structure and dynamics of aquatic invertebrate communities in the PAD, without relying on the fragile premise of ecological equivalence in neutral theory (50, 52). Although many ecologists have acknowledged stochastic processes are likely to have a role in understanding community composition (53, 54), we are unaware of any biomonitoring programs that incorporate, or even acknowledge community assembly mechanisms other than environmental filtering (e.g. 37). Our results firmly challenge that traditional perspective, and if we wish to understand the resilience of the PAD, we must adopt a metacommunity perspective (55). More broadly, a metacommunity perspective of the PAD could indicate which assembly processes are absent from more managed landscapes, therefore providing critical insights into the mitigation of biodiversity loss at the landscape scale.

The isolation of wilderness areas like the Peace-Athabasca Delta implies a pristine nature, but that isolation has also hindered our appreciation of the sheer magnitude of diversity which occurs there, and has until now precluded a basic description of how community structure changes over space and time. Near-random patterns of assembly and substantial sampling error pose a challenge to detecting ecosystem change. Without evaluating data quality and statistical power at the start, many monitoring programmes are unable to confidently reject a false null hypothesis, undermining project goals and providing a misleading sense of achievement (27, 56). Despite the high turnover, we show the statistical power of data generated by DNA metabarcoding was superior to traditional biomonitoring approaches for the detection of large-scale ecosystem change. Although macroinvertebrate composition provides a wealth of information, the power to detect and draw inference from taxonomic changes will be improved by further refining the list of taxa that respond to particular threats (e.g. oil sands contaminants), particularly by linking metabarcoding to trait databases (57), and this remains a major focus of our ongoing research.

Materials and Methods

Field surveys
Field survey methods followed the CABIN wetland macroinvertebrate protocol (14, 19). Briefly, aquatic invertebrates were sampled by sweeping submerged and emergent aquatic vegetation at wetland edges for 2 minutes. A sterile 400µm mesh net was steadily moved in a zig-zag pattern, from the surface of the sediment to the water surface, to capture disturbed organisms and minimise the amount of sediment collected. Excess vegetation was carefully rinsed and removed, and samples with excess sediment were sieved. Material was placed in sterile 1L polyethylene sample jars, filled no more than half full, and immediately preserved in 95% ethanol in the field. Samples were stored in a cooler with ice in the field and transferred to a freezer at the field station before shipment. Nets were disinfected between each new site, and field crews wore nitrile gloves to collect and handle samples, minimizing the risk of cross-site contamination.

Sample Processing
In total, 126 and 138 samples were collected from 72 separate site visits for the CABIN and DNA metabarcoding datasets, respectively (SI Appendix1, Table S1.2). Samples identified using morphological characteristics were processed and identified in accordance with the CABIN laboratory manual (19). Briefly, material from each 2-minute sweep was subsampled using a 100-cell Marchant box. Successive cells were processed until at least 300 individuals were identified and a minimum of five cells were processed. Most taxa were identified to family level, although for some groups only class- or order-level identification was recovered, and given the importance and diversity of Chironomidae, we retained four sub-family divisions that could be reliably identified (SI Appendix 1).

The lab protocol for processing samples for DNA metabarcoding followed the same procedure as outlined in Gibson, et al. (14). This targeted the CO1 amplicon using two complementary primers, BE/BR5 and F230R (30, 58). All DNA samples were analysed using BE or BR5 that target the same COI region, and F230R was introduced in 2012. While field and lab protocols have remained consistent since the study began, there have been a number of advances made in bioinformatic tools, as well as expansion of the reference sequence libraries supporting the identification of taxa (59). The bioinformatic pipeline used to process all samples in this study, as well as the CO1 classifier that allocates sequences to the most likely taxa, is described in SI Appendix 1 and available on Github (60). The sequences generated have been deposited in the NCBI Sequence Read Archive (SRA), project PRJNA603969.

Hierarchical MSOM

Multi-species occupancy models (MSOMs) employ a flexible hierarchical framework that allows for imperfect detection to predict species' occurrence (25). The hierarchy consists of an underlying state-model that describes the probability of species' occurrence, and a second observation-model to describe the probability of detecting that species when it is present (informed by detection across replicates). The fitted state-model is thereby updated to account for the probability of false-negatives. MSOMs extend this single-species approach by assuming species' coefficients are related, and can be treated as random effects, drawn from a common distribution (hyper-parameters). Data augmentation extends the community approach a step further by using the hyper-parameters for occupancy and detectability to estimate the possibility additional taxa may have been present, but by chance were never observed. Our analysis adapted the notation and code provided by (41) as the basis for this study (see SI Appendix 1 for model code):

1. Data augmentation process: $w_k \sim Bernoulli(\Omega)$
2. State process: $z_{ik} \sim Bernoulli(w_k\psi_k)$
3. Observation process: $y_{ijk}|z_{ik} \sim Bernoulli(z_{ik}p_{ijk})$
4. Models of taxon heterogeneity: $logit(\psi_k) \sim lpsi_k + betalpsi_k \times \text{covariate}_i + ...$
$logit(p_{ijk}) \sim lp_k + betalp_k \times \text{covariate}_{ij} + ...$

Given: $lpsi_k \sim Normal(\mu_{lpsi}, \sigma^2{}_{lpsi})$
$betalpsi_k \sim Normal(\mu_{betalpsi}, \sigma^2{}_{betalpsi})$
$lp_k \sim Normal(\mu_{lp}, \sigma^2{}_{lp})$
$betalp_k \sim Normal(\mu_{betalp}, \sigma^2{}_{betalp})$

The observed data y_{ijk} describe the detection or non-detection of taxon k at site i in replicate sample j. Replicate observations, in our case simultaneous independent samples (21), allowed the model to discriminate between processes that determine the system's state (occupancy) and the observation process (detectability). The occupancy of each taxon at each site z_{ik} is described by a Bernoulli trial with probability ψ_{ik}, and the likelihood of detecting the respective taxa in each replicate sample is described by another set of Bernoulli processes with probability p_{ijk}. Seven water temperature and flood regime variables were tested as covariates within a multiple logistic regression for occupancy, and measures of sample processing effort were tested for detectability (sequencing depth and the number of individuals identified). Individual intercepts and slopes represented species-specific random-effects, governed by a common prior distribution whose mean and variance were estimated as a community-level hyperparameter.

The statistical distributions of the parameters governing occupancy and detectability shared by the community were used to consider the possibility of other taxa in the metacommunity that were not recorded in any visit to any site, a process known as data augmentation (48, 61). Given a sufficiently large total pool of M potential taxa, a set of binary indicators w_k, governed by the parameter Ω, represent the probability each taxon is part of the community. The total number of taxa in the metacommunity (γ-diversity) is therefore simply the sum of w_k.

The occupancy model above was suitable for presence/absence observations of taxa, but CABIN samples also included information on the relative abundances of taxa. To utilize all information available, we constructed a community-level N-mixture model that estimates the latent abundance N_{ik} of each taxon, rather than their occurrence (z_{ik}), and modelled counts as a function of a Poisson distribution.

2. State process: $\qquad$ $N_{ik} \sim Poisson(w_k \lambda_k)$
3. Observation process: $\qquad$ $y_{ijk}|N_{ik} \sim Binomial(N_{ik}, p_{ijk})$
4. Models of taxon heterogeneity: $\qquad$ $log(\psi_k) \sim lpsi_k + betalpsi_k \times covariate_i + ...$

Finally, model selection for covariates of taxon heterogeneity in both the occupancy and N-mixture models was determined by a set of binary indicator variables Vx1-xn, one for each of the *n* predictor variables used (62). Using Vx ~ Bernoulli(0.5) as standard priors, variables had an equal likelihood of being included or excluded from likelihood estimates, and model selection was therefore based on which combination had the highest joint posterior probability *p*(Vx1-xn=1). Note convergence of the Vx indicators was very slow, particularly in the most complex models, and a "slab and spike" approach did not improve mixing (see 7.6.2 in 41).

Analyses were conducted using the R package jagsUI (63). We assessed model convergence of all monitored parameters across chains by visual inspection of trace plots and by using the Gelman-Rubin statistic (64), with the diagnostic value <1.1. As over-dispersion cannot be estimated from the binary responses in occupancy models (41), plots of Dunn-Smyth residuals for fitted estimates of occupancy and detectability were used to evaluate the fit of separate taxa (65). Although plots suggest the models were well fit in most cases, the pattern of residuals suggested there may have been other covariates, or non-linear effects, missing from the models influencing the occupancy of some taxa.

Simulation and power analysis
The code and process used to simulate communities is described in detail in SI Appendix 2. In summary, a hypothetical presence-absence matrix of the metacommunity was derived from estimates of gamma diversity and occupancy in the DNA *Gpa* occupancy model from which we could manipulate sampling designs. Environmental covariates were varied according to the mean and standard deviation of values observed from the surveys available to us (SI Appendix 2, Fig. S2.1), but the simulation was not spatially explicit. Whilst occupancy covariates drove some temporal turnover (SI Appendix 2, Fig, S2.2, approximately 10-27%), this was insufficient to replicate the turnover observed (SI Appendix 2, Fig. S2.3), so permutation of the presence-absence matrix was used to simulate further stochastic changes in composition (i.e. local extinction/colonization; 66). Replicating observed turnover required the complete redistribution of occurrences (i.e. random assembly patterns). Taxon occupancy (row sums) and site richness (column sums) were held constant during permutation. The metacommunity was modified by successively removing occurrences of taxa based on a hypothetical distribution of tolerances, which were themselves generated to covary with the distribution of occupancy. Sampling error was applied by a binomial function weighted by the taxon's probability of detection, and the 'detected' composition of reference and modified metacommunities were then compared using mvabund (67). Power of DNA *Gpa* was compared to DNA *Fpa* and CABIN *Fpa* approaches by aggregating genera to family-level, and subsequently applying the family-level detection probabilities.

Acknowledgments

References

1. IPBES (2019) Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science- Policy Platform on Biodiversity and Ecosystem Services. eds Brondizio ES, Settele J, Díaz S, & Ngo HT (IPBES Secretariat, Bonn, Germany).
2. Pereira HM*, et al.* (2013) Essential Biodiversity Variables. *Science* 339(6117):277-278.
3. Ormerod SJ, Dobson M, Hildrew AG, & Townsend CR (2010) Multiple stressors in freshwater ecosystems. *Freshwater Biology* 55(SUPPL. 1):1-4.
4. Gray C*, et al.* (2014) FORUM: Ecological networks: the missing links in biomonitoring science. *Journal of Applied Ecology* 51(5):1444-1449.
5. Friberg N*, et al.* (2011) Biomonitoring of Human Impacts in Freshwater Ecosystems: The Good, the Bad and the Ugly. *Advances in Ecological Research*, ed Guy W (Academic Press), Vol Volume 44, pp 1-68.
6. Bailey RC, Linke S, & Yates AG (2014) Bioassessment of freshwater ecosystems using the Reference Condition Approach: comparing established and new methods with common data sets. *Freshwater Science* 33(4):1204-1211.
7. Taberlet P, Coissac E, Hajibabaei M, & Rieseberg LH (2012) Environmental DNA. *Molecular Ecology* 21(8):1789-1793.
8. Yu DW*, et al.* (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* 3(4):613–623.
9. Baird D & Hajibabaei M (2012) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular ecology* 21(8):2039-2044.
10. Jones FC (2008) Taxonomic sufficiency: The influence of taxonomic resolution on freshwater bioassessments using benthic macroinvertebrates. *Environmental Reviews* 16:45-69.
11. Strachan SA & Reynoldson TB (2014) Performance of the standard CABIN method: comparison of BEAST models and error rates to detect simulated degradation from multiple data sets. *Freshwater Science* 33(4):1225-1237.
12. Chariton AA*, et al.* (2015) Emergent technologies and analytical approaches for understanding the effects of multiple stressors in aquatic environments. *Marine and Freshwater Research* 67(4):414–428.
13. Bush A*, et al.* (2019) Studying Ecosystems With DNA Metabarcoding: Lessons From Biomonitoring of Aquatic Macroinvertebrates." Frontiers in Ecology and Evolution 7(434): 1-12.
14. Gibson JF*, et al.* (2015) Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. *PLoS ONE* 10(10):e0138432.
15. Peters DL, Prowse TD, Pietroniro A, & Leconte R (2006) Flood hydrology of the Peace-Athabasca Delta, northern Canada. *Hydrological Processes* 20(19):4073-4096.
16. Peters DL, Caissie D, Monk WA, Rood SB, & St-Hilaire A (2016) An ecological perspective on floods in Canada. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques* 41(1-2):288-306.
17. WBNP (2019) Action Plan to protect the World Heritage Values of Wood Buffalo National Park. (Parks Canada, Fort Smith, NT Canada).
18. Hajibabaei M, Baird DJ, Fahner NA, Beiko R, & Golding GB (2016) A new way to contemplate Darwin's tangled bank: how DNA barcodes are reconnecting biodiversity

science and biomonitoring. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 371(1702).

19. ECCC (2018) CABIN Wetland Macroinvertebrate Protocol. (Environment and Climate Change Canada, Gatineau QC, Canada).

20. Ugland KI, Gray JS, & Ellingsen KE (2003) The species–accumulation curve and estimation of species richness. *Journal of Animal Ecology* 72(5):888-897.

21. Guillera-Arroita G (2017) Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography* 40(2):281-295.

22. Cao Y, Hawkins Charles P, Larsen David P, & Van Sickle J (2007) Effects of Sample Standardization on Mean Species Detectabilities and Estimates of Relative Differences in Species Richness among Assemblages. *The American Naturalist* 170(3):381-395.

23. Clarke R (2009) Uncertainty in WFD assessments for rivers based on macroinvertebrates and RIVPACS. in *Integrated catchment science programme Science report: SC060044/SR4* (Bristol, UK), pp 1-87.

24. Mackenzie DL & Royle JA (2005) Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology* 42(6):1105-1114.

25. Dorazio RM & Royle JA (2005) Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association* 100:389-398.

26. Dorazio RM, Royle JA, Söderström B, & Glimskär A (2006) Estimating species richness and accummulation by modelling species occurrence and detectability. *Ecology* 87(4):842–854.

27. Legg CJ & Nagy L (2006) Why most conservation monitoring is, but need not be, a waste of time. *Journal of Environmental Management* 78(2):194-199.

28. Braukmann TWA*, et al.* (2019) Metabarcoding a diverse arthropod mock community. *Molecular Ecology Resources* 19(3):711-727.

29. McGill BJ (2010) Towards a unification of unified theories of biodiversity. *Ecology Letters* 13(5):627-642.

30. Gibson J*, et al.* (2014) Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences* 111(22):8007-8012.

31. Elbrecht V & Leese F (2017) Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science* 5(11).

32. Curry CJ, Gibson JF, Shokralla S, Hajibabaei M, & Baird DJ (2018) Identifying North American freshwater invertebrates using DNA barcodes: are existing COI sequence libraries fit for purpose? *Freshwater Science* 37(1):178-189.

33. Gregoire Taillefer A & Wheeler TA (2018) Tracking wetland community evolution using Diptera taxonomic, functional and phylogenetic structure. *Insect Conservation and Diversity* 11(3):276-293.

34. Hillebrand H*, et al.* (2018) Biodiversity change is uncoupled from species richness trends: Consequences for conservation and monitoring. *Journal of Applied Ecology* 55(1):169-184.

35. Leese F*, et al.* (2018) Chapter Two - Why We Need Sustainable Networks Bridging Countries, Disciplines, Cultures and Generations for Aquatic Biomonitoring 2.0: A

Perspective Derived From the DNAqua-Net COST Action. *Advances in Ecological Research*, eds Bohan DA, Dumbrell AJ, Woodward G, & Jackson M (Academic Press), Vol 58, pp 63-99.

36. Schmidt-Kloiber A & Nijboer RC (2004) The effect of taxonomic resolution on the assessment of ecological water quality classes. *Hydrobiologia* 516(1):269-283.

37. Hawkins CP, Norris RH, Hogue JN, & Feminella JW (2000) Development and Evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10(5):1456-1477.

38. Macher JN*, et al.* (2016) Multiple-stressor effects on stream invertebrates: DNA barcoding reveals contrasting responses of cryptic mayfly species. *Ecological Indicators* 61(Part 2):159-169.

39. Beermann AJ, Zizka VMA, Elbrecht V, Baranov V, & Leese F (2018) DNA metabarcoding reveals the complex and hidden responses of chironomids to multiple stressors. *Environmental Sciences Europe* 30(1):26.

40. Ramos-Merchante A & Prenda J (2017) Macroinvertebrate taxa richness uncertainty and kick sampling in the establishment of Mediterranean rivers ecological status. *Ecological Indicators* 72:1-12.

41. Kéry M & Royle AJ (2015) *Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and Static Models* (Academic Print).

42. Reid SM & Haxton TJ (2017) Backpack electrofishing effort and imperfect detection: Influence on riverine fish inventories and monitoring. *Applied Icthyology* 33(6):1083-1091.

43. Wedderburn SD (2018) Multi-species monitoring of rare wetland fishes should account for imperfect detection of sampling devices. *Wetlands Ecology and Management* 26(6):1107-1120.

44. Xingfeng S*, et al.* (2018) The importance of accounting for imperfect detection when estimating functional and phylogenetic community structure. *Ecology* 99:2103-2112.

45. Ficetola GF*, et al.* (2015) Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources* 15(3):543-556.

46. Doi H*, et al.* (2019) Evaluation of detection probabilities at the water-filtering and initial PCR steps in environmental DNA metabarcoding using a multispecies site occupancy model. *Scientific Reports* 9(1):3581.

47. Guillera-Arroita G & Lahoz-Monfort JJ (2012) Designing studies to detect differences in species occupancy: power analysis under imperfect detection. *Methods in Ecology and Evolution* 3(5):860-869.

48. Guillera-Arroita G, Kéry M, & Lahoz-Monfort JJ (2019) Inferring species richness using multispecies occupancy modeling: Estimation performance and interpretation. *Ecology and Evolution* 9(2):780-792.

49. Keizer-Vlek HE, Verdonschot PFM, Verdonschot RCM, & Goedhart PW (2012) Quantifying spatial and temporal variability of macroinvertebrate metrics. *Ecological Indicators* 23:384-393.

50. Souza AF, Bezerra AD, & Longhi SJ (2016) Quasi-neutral community assembly: Evidence from niche overlap, phylogenetic, and trait distribution analyses of a subtropical forest in South America. *Perspectives in Plant Ecology, Evolution and Systematics* 23:1-10.

51.   Leibold M & Chase J (2018) *Metacommunity Ecology* (Princeton University Press, Princeton and Oxford).
52.   O'Sullivan JD, Knell RJ, & Rossberg AG (2019) Metacommunity-scale biodiversity regulation and the self-organised emergence of macroecological patterns. *Ecology Letters* 0(0).
53.   Heino J (2011) A macroecological perspective of diversity patterns in the freshwater realm. *Freshwater Biology* 56(9):1703-1722.
54.   Thompson R & Townsend C (2006) A truce with neutral theory: local deterministic factors, species traits and dispersal limitation together determine patterns of diversity in stream invertebrates. *Journal of Animal Ecology* 75(2):476-484.
55.   Adler PB*, et al.* (2005) Evidence for a general species-time-area relationship. *Ecology* 86(8):2032-2039.
56.   Lindenmayer DB & Likens GE (2010) The science and application of ecological monitoring. *Biological Conservation* 143(6):1317-1328.
57.   Compson ZG*, et al.* (2018) Linking DNA Metabarcoding and Text Mining to Create Network-Based Biomonitoring Tools: A Case Study on Boreal Wetland Macroinvertebrate Communities. *Advances in Ecological Research*, (Academic Press), Vol 59, pp 33-74.
58.   Folmer O, Black M, Hoeh W, Lutz R, & Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Marine Biol. Biotech.* 3:294-299.
59.   Porter TM & Hajibabaei M (2018) Over 2.5 million COI sequences in GenBank and growing. 13(9):e0200177.
60.   Porter TM & Hajibabaei M (2018) Automated high throuput animal CO1 metabarcode classification. *Scientific Reports* 8(1):4226.
61.   Royle AJ, Dorazio RM, & Link WA (2007) Analysis of Multinomial Models With Unknown Index Using Data Augmentation. *Journal of Computational and Graphical Statistics* 16(1):67-85.
62.   Kuo L & Mallick B (1998) Variable selection for regression models. *Sankhya: The Indian Journal of Statistics, Series B (1960–2002)* 60(65-81).
63.   Kellner K (2017) jagsUI: A Wrapper Around 'rjags' to Streamline 'JAGS' Analyses. R package version 1.4.9.).
64.   Gelman A & Rubin DB (1992) Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* 7:457-511.
65.   Warton DI, Stoklosa J, Guillera-Arroita G, MacKenzie DI, & Welsh AH (2017) Graphical diagnostics for occupancy models with imperfect detection. *Methods in Ecology and Evolution* 8(4):408-419.
66.   Gibert C & Escarguel G (2018) PER-SIMPER—A new tool for inferring community assembly processes from taxon occurrences. *Global Ecology and Biogeography* 0(0).
67.   Wang Y, Naumann U, Eddelbuettel D, Wilshire J, & Warton D (2019) mvabund: Statistical Methods for Analysing Multivariate Abundance Data. R package version 4.0.1. https://CRAN.R-project.org/package=mvabund).
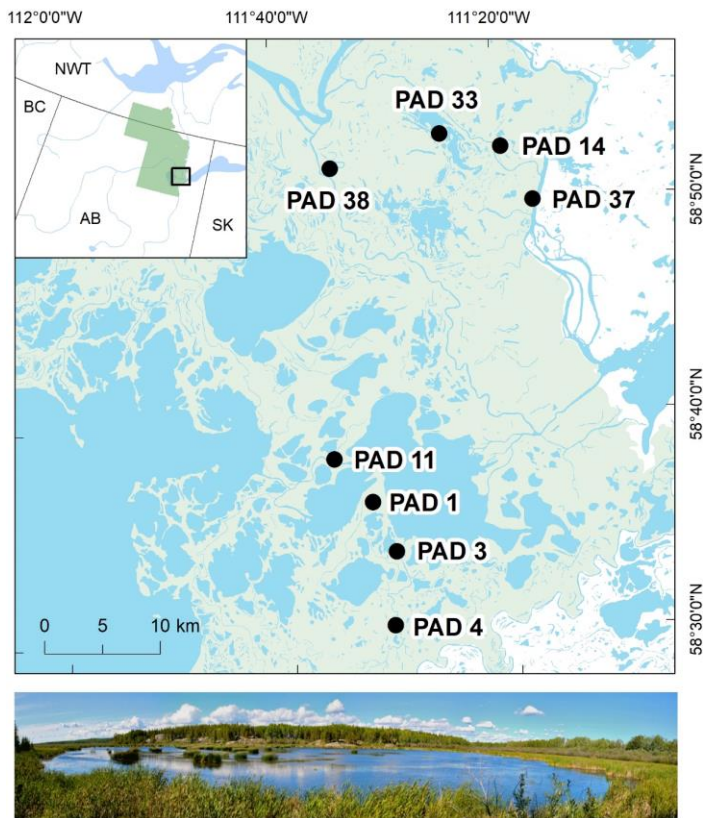
Figures

Figure 1. Figure 1: Location of sampling sites in the Peace-Athabasca Delta. The inset shows the full extent of Wood Buffalo National Park in Alberta (AB), and boundaries of neighbouring provinces: British Columbia (BC), Saskatoon (SK), and the Northwest Territories (NWT). Photo taken at Rocher River wetland (PAD 37).
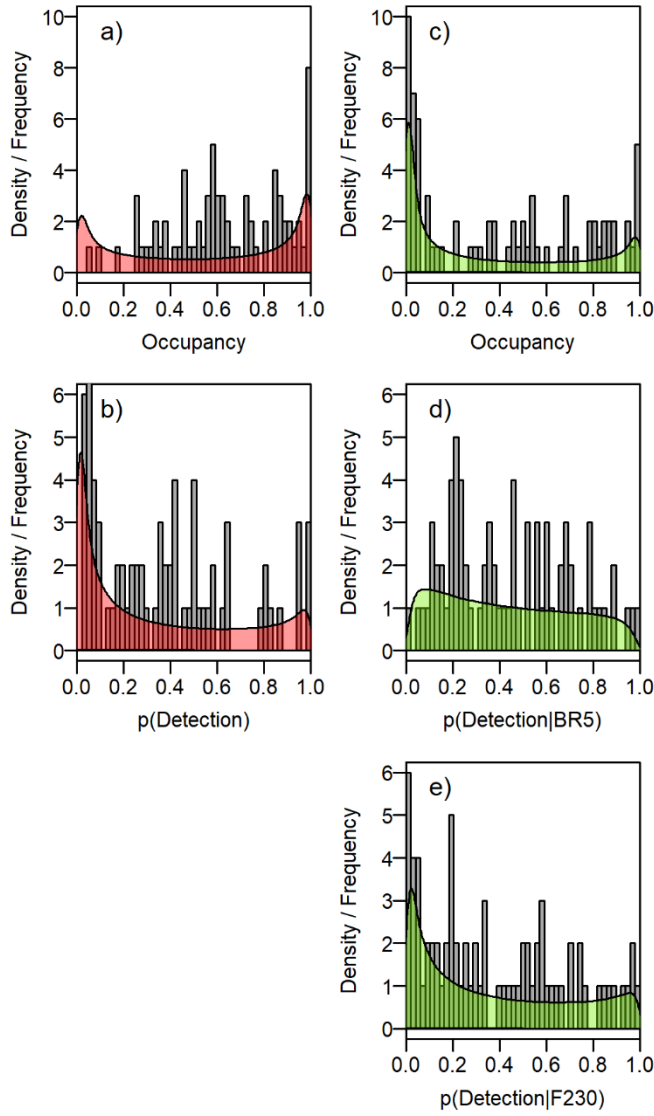
Figure 2. Predicted occupancy (a,c) and detectability (b,d,e) of taxa based on the presence-absence data collected using the CABIN protocol (a,b) and DNA metabarcoding (c-e) at the family-level. Detectability using metabarcoding is further split by primer pair (d,e). The shaded polygons describe the probability density of the community hyperparameters, and the grey bars indicate the underlying frequency of the values estimated for each taxon. See Figure S.10 for the CABIN *Fcount* and DNA *Gpa* model distributions.
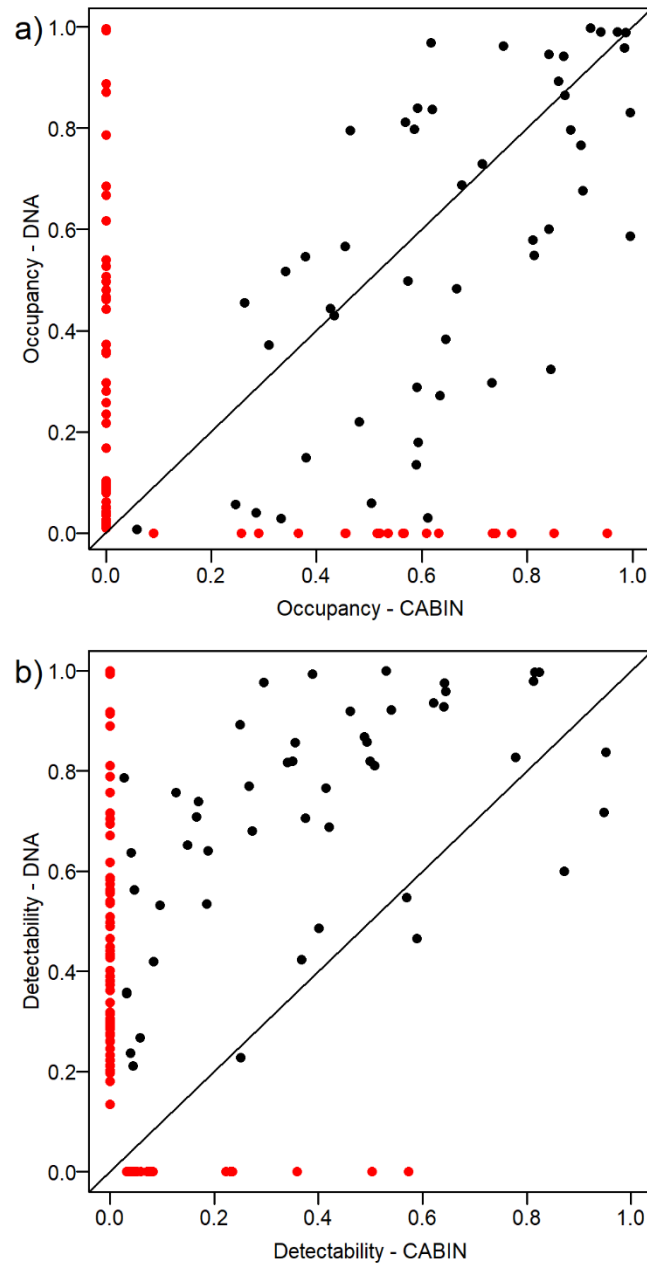
Figure 3. Comparison of a) occupancy and b) detectability estimates in models trained by CABIN data and DNA metabarcode data at the family-level (n= 50). Red points indicate taxa not observed by the complementary method i.e. 18 and 59 families were unique to CABIN and metabarcoding respectively. See SI Appendix 1 for further information on the identities of unique taxa.
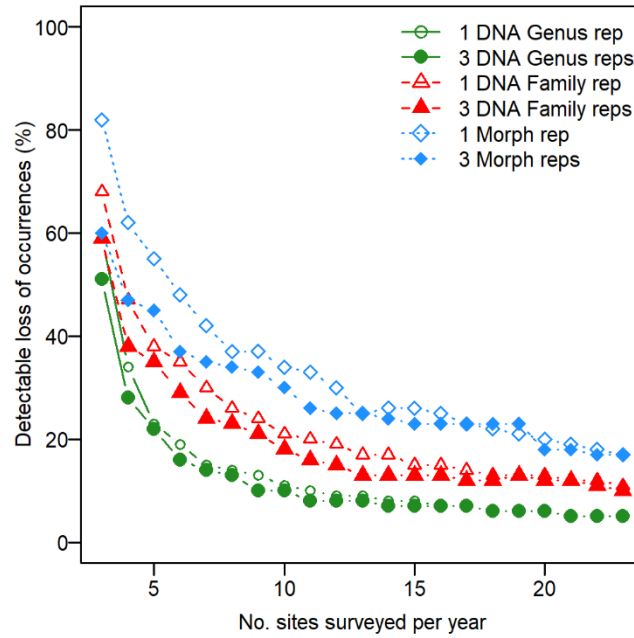
Figure 4. Minimum reduction to community occupancy that is detectable >50% of the time with 95% confidence in response to number of sites surveyed annually. Lines show the average of 100 simulations based on the CABIN *Fpa* (blue), DNA *Fpa* (red) and DNA *Gpa* (green) occupancy-detection models, with either single (open symbol) or triplicate (closed symbol) samples per site. Taxon tolerance was not correlated with occupancy. See SI Appendix 2 for further information.