qPCR-Based Prediction of Low-Level Microcystin-LR Using *mcy*E and Passive Sampling Across Multiple Lakes Over Three Years

David J Redden1\*, Clarke Brown1, Morgan Harasymchuk1, Saksham Bafna1, Justin Laforest2, Nicole Taylor2, Lindsay Anderson1,3, and Graham A. Gagnon1

1 Centre for Water Resources Studies, Department of Civil & Resource Engineering, Dalhousie University, 1360 Barrington St. Halifax, Nova Scotia, Canada

2 City of Moncton, 655 Main St., Moncton, New Brunswick, Canada

3 Halifax Water, 450 Cowie Hill Road, Halifax, Nova Scotia, Canada

\* Corresponding author

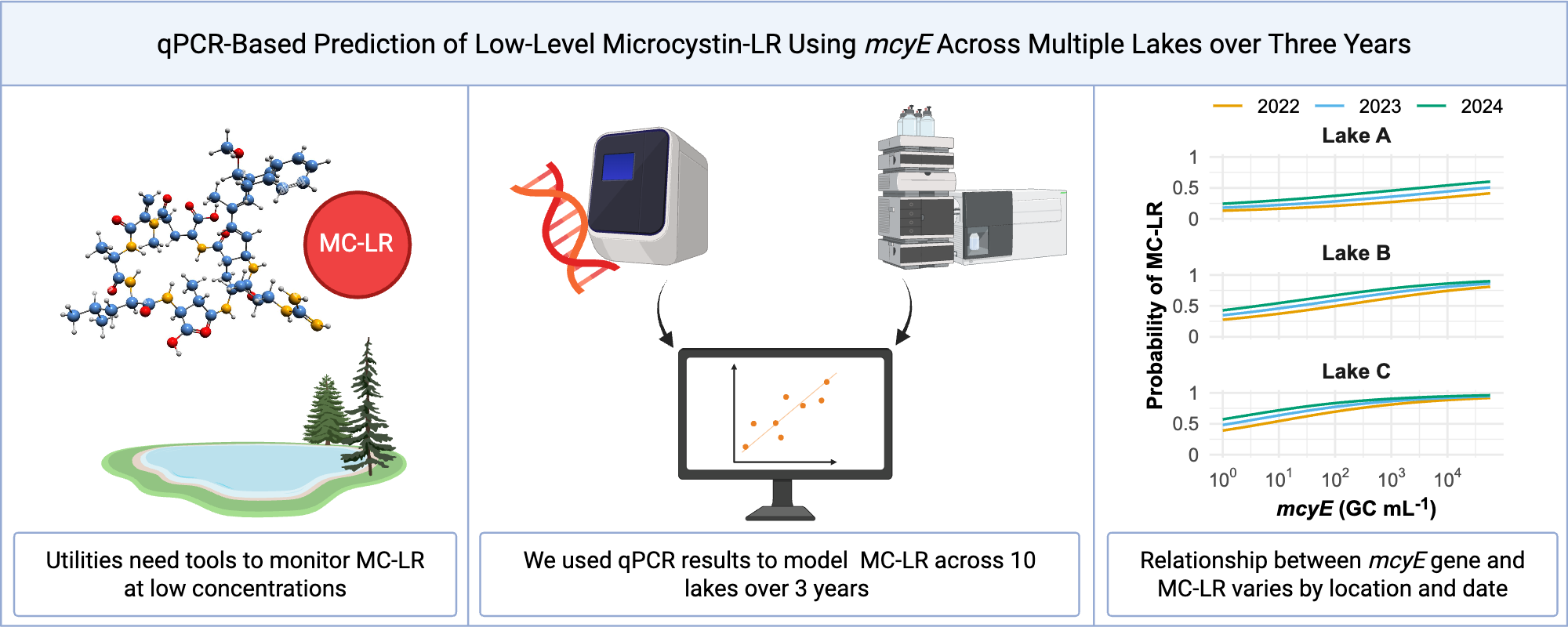
Manuscript submitted for possible publication in *Water Research X*

Date: August XX, 2025

# Highlights

* Passive sampling and qPCR targeting the *mcy*E gene were used to monitor low concentrations (< 1 µg L-1) of MC-LR over three years in ten lakes.
* Differences in MC-LR detection between lakes were not explained by differences in bulk nutrients concentrations or NOM content.
* Random forest and Bayesian hierarchical logistic regression identified *mcy*E as the best predictor of MC-LR presence.
* Interpretation of *mcy*E gene concentrations must be tailored to specific locations and times when assessing risk of low-level MC-LR

# Graphical Abstract



Created in https://BioRender.com

# Abstract

Microcystin-LR (MC-LR) is a cyanobacterial hepatotoxin that poses health risks even at low concentrations. Because quantitative analysis of MC-LR is costly and time-consuming, water managers rely on early warning tools to determine when confirmatory testing is warranted. Quantitative PCR (qPCR) targeting the *mcy* genes has emerged as one such tool, but its reliability across lakes and seasons — particularly at low toxin concentrations — remains unclear. In this study, we used passive sampling to detect low concentrations (< 1 µg L-1) of MC-LR, and paired this with qPCR monitoring of *mcy*E to assess the utility of qPCR for monitoring low-level MC-LR presence over three years across ten lakes (total of n = 893 distinct samples). We developed location- and year-specific hierarchical Bayesian models to estimate the probability of MC-LR detection from *mcy*E concentrations and environmental covariates. Although *mcy*E was the strongest overall predictor, its relationship with MC-LR varied substantially by lake and year, and these hierarchical models were essential in capturing this variability. These findings highlight the promise of *mcy*E-based early warning systems for low-concentrations of MC-LR, but emphasize that interpretation must be tailored to local ecological and seasonal conditions.

Keywords: Cyanotoxin monitoring, microcystins, qPCR, *mcy*E,

# 1. Introduction

Microcystins are a group of cyanotoxins produced by several cyanobacterial genera, including *Microcystis*, *Planktothrix*, and *Dolichospermum*, and are considered the most widely distributed cyanotoxins in freshwater (Rastogi et al., 2014; Wei et al., 2024). Microcystins are well-known hepatotoxins with established health risks at low concentrations (e.g., USEPA (2015) 10-day health advisory level of 0.3 µg L-1 total microcystins for children under six years of age), with recent work suggesting that even lower thresholds (0.01 to 0.03 µg L-1) might be warranted (Miller et al., 2017).

Passive sampling is a valuable tool for detecting microcystins at low concentrations. When using grab sampling, large water volumes may need to be concentrated (e.g., solid phase extraction) to recover sufficient toxin mass for quantification, particularly when concentrations are low. Additionally, cyanobacteria are mobile within the water column, subject to wind- and current-driven transport, and capable of episodic toxin production (Chorus and Welker, 2021), so discrete grab samples may fail to capture the full extent of microcystin presence. Passive samplers, by contrast, integrate exposure over time providing a more representative assessment of toxin occurrence (Loaiza-González et al., 2024). Several studies have shown that passive sampling results correlate with grab sample results (Jaša et al., 2019; Wang et al., 2022) or detect microcystins that are not detected using grab samples (Brophy et al., 2019; Kohoutek et al., 2010; Kudela, 2011; Wiltsie et al., 2018).

Regardless of the sampling method used, quantification of microcystins relies on expensive and time-consuming analytical techniques — typically liquid chromatography with tandem mass spectrometry (LC-MS/MS) and enzyme-linked immunosorbent assays (ELISA). In addition to their cost and relatively long analytical turnaround times, these methods are inherently retrospective. For these reasons, monitoring programs typically adopt a tiered approach, using other less-expensive and more-rapid tools to provide early warning of potential toxin events, and reserving LC-MS/MS or ELISA for confirmation when microcystin concentrations are likely to approach regulatory thresholds (Almuhtaram et al., 2021; Chorus and Welker, 2021; Kibuye et al., 2021; Schürmann et al., 2024).

Quantitative polymerase chain reaction (qPCR) has become a common tool for assessing the potential for cyanotoxins, and microcystins in particular (Feist and Lance, 2021; Pacheco et al., 2016). Microcystin biosynthesis is regulated by the *mcy* gene cluster, (Dittmann et al., 1997; Kaebernick et al., 2002; Nishizawa et al., 2000; Tillett et al., 2000) a conserved set of genes found across microcystin producing genera (Christiansen et al., 2003; Hisbergues et al., 2003; Rouhiainen et al., 2004; Wei et al., 2024). While the absence of this cluster precludes microcystin production (Dittmann et al., 1997; Tillett et al., 2000), its presence does not guarantee toxin synthesis. Regulation of these genes is not fully understood, but depends on a variety of factors including light, temperature, nutrients (primarily phosphorus and nitrogen), trace metals, and predation (Dai et al., 2016; Neilan et al., 2013; Wei et al., 2024). Further, partial deletions, insertions, and point mutations may result in cyanobacteria with *mcy* genes that are incapable of toxin synthesis (Christiansen et al., 2008, 2006; Kaebernick et al., 2001; Noguchi et al., 2009).

Correlations between *mcy* gene concentrations and microcystin levels depend on the specific *mcy* gene targeted (Pacheco et al., 2016). Studies that have developed qPCR assays based on the *mcy*E gene have found that by targeting different regions of that gene, it is possible to detect all major microcystin producers (Al-Tebrineh et al., 2012; Al-Tebrineh et al., 2011; Jungblut and Neilan, 2006), as well as specific microcystin-producing genera (Sipari et al., 2010; Vaitomaa et al., 2003). Consequently, *mcy*E has become a widely used target for monitoring microcystins in freshwater ecosytems (Duan et al., 2022; Gonzales Ferraz et al., 2024; Lu et al., 2020; Madany et al., 2025; Padovan et al., 2023; Panksep et al., 2020). While many studies have found correlations between *mcy* genes and microcystin concentrations, few have explored how qPCR results can inform public health and safety decisions for water utilities and resource managers.

While previous studies have demonstrated that *mcy*E concentrations can serve as early warning indicators of microcystins in a single-lake system with moderate (> 8 µg L-1) microcystin concentrations (Duan et al., 2022; Lu et al., 2020), it remains unclear whether this approach can be generalized across lakes with different physicochemical conditions and cyanobacterial communities, and specifically in systems where MC-LR concentrations are low. In this work, we evaluated the utility of *mcy*E as a predictor of infrequent, low (< 1 µg L-1) MC-LR concentrations over three years, across multiple lakes. Using passive samplers to detect low-level MC-LR events often missed by grab samples, we developed location-specific, time-varying probabilistic models for MC-LR detection based on *mcy*E concentrations. We also assessed whether incorporating additional environmental or climate predictors could improve predictive performance beyond *mcy*E alone.

# 2. Methods

## 2.1 Sample locations and collection methodology

The eight lakes included in the monitoring program ranged in size from 0.6 to 160 ha and the maximum depths ranged from 2 to 30 m (Table S3). Turtle Creek Reservoir (TCR), Tower Road Reservoir (TRR), and Irishtown Nature Park (INP) are current or former water supply lakes situated in areas that are mostly forested. Lake Fletcher is situated in a suburban residential area and a large portion of its shoreline and watershed is forested. The remaining lakes are popular recreational sites in urban areas with a mix of commercial and residential development. Sample locations were chosen to either assess risks to recreational users, or were in areas where suspected cyanobacterial blooms were observed in the past.

At TRR and TCR, grab samples were collected from the side of a boat; at all other locations, grab samples were collected from the shoreline. All grab samples were taken within the top 1 m of the water column. Passive samples — comprising 1000 mg of Oasis HLB adsorbent (Waters Limited, Stamford Ave., UK) in a heat-sealed nylon mesh bag with 25 µm pore size — were held in plastic sampling cages and affixed at the sampling location. The sampling cages were modified versions of those used by Hayes et al. (2021). When grab samples were collected, the corresponding passive sample was removed from the sampling cage and placed in a falcon tube. A new bag of adsorbent was then placed in the sampling cage and left in the water until the next sampling event — for a total sampling duration of 7 days for each passive sample. We collected samples weekly from late-May until mid-November on the same day of the week and at approximately the same time of day at each site. Field blanks were included during each sampling event. Further details regarding sample bottle preparation and sampling procedure are provided in Supplementary Methods.

## 2.2 Physicochemical and climate data

A hand held sonde (YSI Ohio, USA) was used to measure in-situ temperature, dissolved oxygen, conductivity, total dissolved solids, and pH. Turbidity, total organic carbon (TOC), total nitrogen (TN), and elemental composition — including total phosphorus — were measured using unfiltered lake water. Samples were filtered through 0.45 µm polysulfone membrane filters for measurement of dissolved organic carbon (DOC), dissolved anions (chloride, nitrate, nitrite, sulfate, and orthophosphate), and characterization of true color and . Details regarding these analyses are provided in Supplementary Methods.

Historical precipitation and temperature data for the weather stations nearest the study lakes — with complete data over the monitoring period — was obtained from Environment Canada (Environment and Climate Change Canada, 2025). For INP, TCR, and TRR data was obtained from the Moncton Romeo Leblanc International Airport Station; for all other lakes, data was obtained from the Shearwater RCS station. Moving averages (3- and 7-day) for a variety of temperature and precipitation measures were computed and used as predictor variables in random forest modelling.

## 2.3 DNA extraction

Two DNA extraction methods were used during this monitoring program. In 2022, samples were concentrated onto 25 mm, 0.8 µm pore size Versapore® acrylic copolymer membrane filters (Pall Corporation, Port Washington, NY) and extracted by bead-beating in lysis buffer. In 2023 and 2024, water samples were concentrated onto 47 mm, 0.22 µm polyvinylidene fluoride (PVDF) membrane filters (MilliporeSigma, Burlington, MA) and extracted using the DNeasy PowerWater Kit (Qiagen, Venlo, The Netherlands). Filtered water volumes varied between samples and years, with lower volumes obtained in 2022 due to the smaller filter size. Across the program, water was filtered until the membranes clogged, with filtered volumes ranging from 45 to 525 mL (median: 350 mL). In 2024, a subset of samples was extracted using both methods to compare the resulting qPCR gene copy concentrations. The DNeasy PowerWater method yielded higher gene copy concentrations, particularly at lower concentrations (Figure S1). Further details regarding DNA extraction and quality control are provided in Supplementary Methods.

## 2.4 qPCR

Phytoxigene™ CyanoDTec Total Cyanobacteria and Toxin Gene kits (Phytoxigene™, Inc., Akron, OH) were used to quantify cyanobacteria-specific 16S rRNA and cyanotoxin synthesis genes. The Toxin Gene kit targets the *mcy*E*/nda*F, *cyr*A, and *sxt*A genes required for production of microcystins/nodularins, cylindrospermopsins, and saxitoxins, respectively. Note that *mcy*E *and nda*F are orthologous gene that encode for an aminotransferase domain required for the production of microcystins and nodularins, respectively (Jungblut and Neilan, 2006). Herein, we simply refer to this gene target as *mcy*E. We chose to use these qPCR assays because they are currently used by the Ohio EPA for cyanobacterial monitoring (Ohio EPA, 2018) and given their commercial availability, they are likely to be used by other water utilities or resource managers that do not have internal capacity for qPCR assay development.

Development of the assays, including coverage and primer/probe sequences is provided in Al-Tebrineh et al. (2010) and Al-Tebrineh et al. (2012). Additional details regarding qPCR methodology, quality control, and quantification of gene concentrations are provided in Supplementary Methods.

## 2.5 MC-LR analysis

MC-LR was sampled using grab and passive samples. Grab samples were analyzed following three cycles of freeze-thaw to lyse the cells and thus measure total (intra- and extracellular) MC-LR in the water. Passive samples were analyzed by eluting adsorbed MC-LR from the Oasis HLB adsorbent (Waters Limited, Stamford Ave., UK) and the nylon bag used to hold it. It is assumed that passive samplers predominantly adsorb extracellular MC-LR and do so over the 7-day deployment period (Kudela, 2011; Lance et al., 2021).

Concentrations obtained from analysis of the passive samplers are given in units of µg MC-LR per gram of adsorbent (µg g-1). Converting this to an aqueous concentration would require determining the sampling rate constant of the passive sampler at each location (Harman et al., 2012; Wang et al., 2022). Throughout this work, we treated the detection of MC-LR on passive samplers as a binary indicator of the presence of extracellular MC-LR in the water. Further details regarding processing and analysis of grab and passive samples is provided in Supplementary Methods.

## 2.6 Data processing and analysis

All statistical analyses were performed in R (R Core Team, 2024) using the R Studio environment (Posit team, 2025). The ‘dunn.test’ package (Dinno, 2024) was used for post-hoc pairwise comparisons (with Benjamini-Hochberg correction) of physicochemical water parameters and the ‘zoo’ package version 1.8-12 (Zeileis and Grothendieck, 2005) was used to compute moving averages of climate data. Other R packages used for specific tasks are detailed below. Data manipulation and figure generation made use of the following packages:‘tidyverse’ v. 2.0.0 (Wickham et al., 2019), ‘tidymodels’ v. 1.2.0 (Kuhn and Wickham, 2020), ‘ggtext’ v. 0.1.2 (Wilke and Wiernik, 2022), ‘tidybayes’ v. 3.0.7 (Kay, 2024), ‘glue’ v. 1.8.0 (Hester and Bryan, 2024), ‘ggsci’ v. 3.2.0 (Xiao, 2024). All data and code required to reproduce the analyses are available in a GitHub repository (Redden, 2025).

The qPCR data contained left-censored (below LOD) observations, and the qPCR, MC-LR, and water chemistry data all contained missing observations. Many statistical methods cannot account for missing or censored data and instead rely on substitution and/or analysis using only complete cases, both of which lead to biased results and errors (Helsel, 2011). In this work we replaced missing data in two ways: 1) through multiple imputation using random forests and 2) Bayesian modelling where missing observations are treated as parameters in the model and estimated during model fitting. Bayesian estimation was also used account for censored predictors the hierarchical logistic regression modelling. Any variable with more than 40% missing observations was excluded from all analyses.

### 2.6.1 Random forest modelling

The ‘ranger’ package (Wright and Ziegler, 2017) was used to fit random forest (RF) models to classify the presence (i.e.  LOD) or absence of MC-LR in grab and passive samples, using all available physicochemical, climate, and qPCR variables as predictors. Prior to model fitting, missing predictor values were multiply imputed using the ‘missRanger’ package (Mayer, 2024). A total of 25 imputed datasets were generated, and RF models were fitted separately to each dataset for prediction of MC-LR presence in grab and passive samples. To verify the appropriateness of imputed values, the distributions of variables were compared before and after imputation (Figures S2 to S6). Variable importance (measured by mean decrease in impurity) was extracted from each RF model.

To evaluate predictive performance for passive sample models, two testing strategies were used: (1) holding out all data from 2024, and (2) holding out one lake at a time. Models were trained on the remaining data and used to predict MC-LR presence in the held-out set. For grab samples, holding out 2024 left too few detections for model training, so only the one-lake-out cross-validation approach was used. RF hyperparameters (*mtry* and *min\_n*) were tuned during training and upsampling (via the *sample.fraction* argument in ‘ranger’) was used during training to address class imbalance (i.e., more non-detect than detect observations).

Predictive performance was summarized across imputations using receiver operating characteristic (ROC) curves, the area under the ROC curve (AUC), and variable importance metrics, with results reported as the median and interquartile range (IQR) across the 25 imputed datasets. For each sampling method, ROC curves are only presented for locations at which there were at least three MC-LR detections.

### 2.6.2 Hierarchical logistic regression

Hierarchical logistic regression models were used to predict the probability of detectable MC-LR in passive samplers, using the variables identified as important in the random forest models. These Bayesian models were fit using the ‘brms’ package (Bürkner, 2017) with custom modifications implemented via the ‘bgamcar1’ package (Trueman, 2024). This approach allows each missing or censored predictor variable to be included as a model parameter that is estimated during fitting, with censored values constrained below their detection limit. Functions from ‘bgamcar1’ were adapted to support the multiple left-censoring limits of qPCR variables. Missing or censored variables were predicted using correlated covariates (Figure S7).

We fit a series of models incorporating different combinations of predictor variables, smooth functions, and varying intercepts and slopes (i.e., random effects). These models were compared using Pareto Smoothed Importance Sampling (PSIS) as an approximation of leave-one-out cross-validation (LOO-CV; Vehtari et al. (2017)), implemented via the ‘loo’ package (Vehtari et al., 2024). Final model selection was based on a combination of LOO information criterion (LOOIC), model stability (Pareto-*k* diagnostics), and ecological plausibility.

Model fit was evaluated using posterior predictive checks via the ‘bayesplot’ package (Gabry and Mahr, 2024), which compare the observed data to simulations drawn from the posterior distribution (Gabry et al., 2019). We also compared predicted probabilities to the observed proportion of MC-LR detections to confirm that the model adequately captured empirical variation.

All predictor variables were scaled by their standard deviation and mean-centered prior to model fitting. qPCR data were transformed before scaling and centering. All models were run with no divergent transitions and values 1.01, indicating good Markov Chain Monte Carlo (MCMC) convergence.

# 3. Results and discussion

## 3.1 Water chemistry differs between lakes and years

Significant differences (Kruskal-Wallis test, *p* < 0.05) in key water chemistry parameters were observed across lakes ([Figure 1](#fig-nutrients)), with post-hoc pairwise comparisons (Dunn test, adjusted *p* < 0.05) confirming inter-lake variability (Figure S8). Among the most notable differences were chloride concentrations and measures of natural organic matter (NOM) including TOC/DOC and true colour. At Penhorn Lake and both sites in Lake Banook, chloride concentrations were an order of magnitude higher than at the other lakes, and measures of NOM were generally lower. These two lakes are subject to winter road salting and urban runoff (Bermarija et al., 2023). In contrast, INP, TCR, and TRR — lakes located in predominantly forested regions with minimal anthropogenic influence — had low chloride concentrations. INP had the highest concentrations of total phosphorus (TP) and total nitrogen (TN), along with elevated measures of NOM.

|  |
| --- |
| Figure 1: Select water chemistry values measured at the study locations over the duration of the three year monitoring program. The colour of the points denotes the year of sampling. Differences in each water chemistry parameter between locations were tested using the Kruskal-Wallis test, as noted on each panel. |

A major rainfall event in July 2023 caused flash flooding in catchments of several study lakes, leading to elevated concentrations of TOC/DOC, true colour, and TN. Increases in NOM were particularly pronounced in INP, Kearney Lake, and Lake Fletcher, while spikes in TN were most apparent at Lake Banook. These shifts in water chemistry may have influenced MC-LR production in 2023 (See [Section 3.2](#sec-results-mclr)).

## 3.2 Low, infrequent microcystin-LR concentrations highlight utility of passive sampling

Grab sample MC-LR concentrations were low and relatively infrequent, with all detections below 1.0 µg L-1 and most occurring between mid-July and late-September ([Figure 2](#fig-all-toxins) and Figure S9). During 2023, MC-LR was not detected in grab samples at any location; in 2024, detections were more frequent and concentrations were generally higher than in 2022.

Passive samples detected MC-LR more frequently than grab samples ([Figure 2](#fig-all-toxins)). In many cases there was an initial sampling event where both methods detected MC-LR and then, in the following weeks, MC-LR continued to be detected via passive sampling. This was likely the result of passive sampling detecting extracellular (i.e. dissolved) MC-LR (See [Section 2.5](#sec-methods-mclr)), and microcystins being predominantly intracellular until cell lysis (Buratti et al., 2017). There were also events where grab and passive sample detections did not co-occur and the passive samples did not detect MC-LR in subsequent weeks (e.g. Cunard Pond). In those cases, adsorption of MC-LR on the passive samples may have been inhibited due to physicochemical conditions in the water (Godlewska et al., 2021). Finally, there were many events where MC-LR was only detected using the passive samples. These cases highlight the ability of the passive samples to accumulate MC-LR over time — providing a stronger signal and/or capturing episodic MC-LR missed by grab samples.

As noted in [Section 2.5](#sec-methods-mclr), MC-LR concentrations obtained from passive samples are difficult to interpret (µg g-1) or compare against guideline or regulatory concentrations (µg L-1). Thus, we have used data from passive samplers as binary indicators of the presence of dissolved MC-LR. However, temporal trends in the passive sample concentrations clearly coincide with grab samples concentrations. This indicates the ability of the passive samplers to detect dissolved MC-LR when total MC-LR was not detected in grab samples.

|  |
| --- |
| Figure 2: Concentrations of MC-LR measured in passive (µg g-1) and grab (µg L-1) samples at each location. Grab samples that were measured as below the LOD have not been shown, but transparent points represent MC-LR concentrations from passive samples that were below the LOD. Dates where there are no points for passive samples indicate that passive samples were not collected. Vertical dashed lines highlight passive samples corresponding to grab sample detections. |

## 3.3 Bulk nutrients don’t adequately explain differences in microcystin-LR detection rates between lakes

Correlations between the frequency of MC-LR detection in the passive samples and key physicochemical parameters revealed weak associations with nutrient concentrations ([Figure 3](#fig-nutrient-toxins)), but moderate, positive associations with indicators of NOM — including TOC, DOC (not shown), and true colour. While some cyanobacteria can metabolize organic carbon under low-light or dark conditions, in most lake ecosystems, atmospheric is the primary carbon source (Stebegg et al., 2023). Thus, the observed relationships between NOM and MC-LR detection likely reflect increased cyanobacterial biomass contributing to NOM, or shared drivers such as eutrophication and watershed inputs that elevate NOM and promote cyanobacterial growth.

Concentrations of total phosphorus and total nitrogen differed between locations ([Figure 1](#fig-nutrients)); however, the magnitudes of these differences were typically small (Figure S8). For this reason, the weak association between nutrients and MC-LR detections is not unexpected. Furthermore, total cyanobacterial abundance (total cyanobacterial 16S rRNA gene) was similar across most locations ([Section 3.4](#sec-results-qpcr)). This indicates that conditions required for cyanobacterial growth are present at most locations, but the factors driving MC-LR production likely involve additional biological or environmental variables beyond bulk nutrient or NOM concentrations.

|  |
| --- |
| Figure 3: Concentrations of select physicochemical parameters at each location. Samples in which MC-LR was detected in passive samples are indicated by red points; non-detects are shown in gray. Locations are ordered top to bottom by the proportion of samples with MC-LR detections, indicated in parentheses (detections / total samples). Spearman correlations coefficients, , were calculated between MC-LR detection rates and the median concentration of each parameter across all samples at a location. These values are shown in the top-right of each panel. |

## 3.4 qPCR results reveal different cyanobacterial communities across lakes

qPCR results for total cyanobacterial 16S rRNA and *mcy*E gene concentrations from 2022 to 2024 revealed clear spatial and temporal variation across sites. Total cyanobacterial 16S rRNA gene concentrations typically ranged between 104 and 106 GC mL-1, with no consistent seasonal trends ([Figure 4](#fig-qpcr)). In contrast, *mcy*E concentrations varied more widely. Sites with infrequent MC-LR detections (e.g., Lake Banook 1 and 2, Penhorn Lake, TRR, TCR) generally exhibited lower *mcy*E concentrations. Seasonal peaks in *mcy*E often occurred in August or September, aligning with periods of increased MC-LR detections in both grab and passive samples ([Figure 2](#fig-all-toxins)). However, these peaks were not always well-defined — in some locations and years, *mcy*E concentrations remained low or fluctuated modestly, making it difficult to distinguish meaningful increases from baseline variation. Across all locations, the frequency of MC-LR detection in passive samples was moderately and positively correlated with *mcy*E concentrations, and with the *mcy*E:16S ratio, but only weakly correlated with total cyanobacteria (Figure S10). On their own, these gene concentrations do not reliably identify locations with more frequent MC-LR detections.

Cyanobacterial genomes typically contain one to four copies of the 16S rRNA gene (Engene and Gerwick, 2011; Schirrmeister et al., 2012), so changes in its concentration may reflect changes in total cyanobacterial abundance and/or taxonomic composition. In contrast, *mcy*E is a single copy gene (Kaebernick et al., 2002; Kaneko et al., 2007; Tanabe et al., 2004), making it a more direct proxy for the abundance of microcystin-producing taxa. Thus, changes in *mcy*E concentrations that occur without corresponding changes in 16S rRNA gene concentrations likely indicate shifts in community composition toward or away from microcystin producers. Additionally, differences in the *mcy*E:16S ratio suggest distinct cyanobacterial communities across locations.

These patterns highlight the limitations of using gene concentrations alone to predict toxin presence, particularly in lakes with low or infrequent MC-LR detections. Instead, the results point to site-specific relationships between toxin-producing taxa and local environmental conditions. This variability reinforces the need for a probabilistic, lake-specific modeling approach, as developed in the remainder of this study.

|  |
| --- |
| Figure 4: Concentrations of total cyanobacteria 16S rRNA (blue) and *mcy*E genes (yellow) at each study location over three years of monitoring. Transparent points represent values that were below their respective limits of detection. Values clustered along the x-axis are samples for which there was no signal detected during qPCR. A pseudocount (10-3) was added to these observations (for plotting only) to distinguish them from missing samples that were not analyzed. |

## 3.5 Random forest models identify *mcy*E as the most important predictor

We trained random forest models to identify the most important qPCR, physicochemical, and climate variables for predicting the presence of MC-LR in grab and passive samples. Models were initially trained on data from 2022 and 2023 and tested by predicting the presence of MC-LR in 2024 (Figure S11). The model predicting MC-LR in grab samples showed good discriminatory performance (AUC = 0.759), while the model for passive samples was more limited (AUC = 0.653). Across both models, important predictors included *mcy*E gene concentrations, their 7-day lagged values, 16S rRNA gene concentrations, their lagged values, and a variety of physicochemical variables.

It was anticipated that *mcy*E (and/or lagged *mcy*E ) would be the most important variable for prediction of MC-LR. This was the case for grab samples, where the concentration of *mcy*E from the previous week was more important than any other variable. However, for prediction of MC-LR in passive samples, TDS, *mcy*E, and cyanobacterial 16S rRNA were of nearly equal importance. TDS and other physicochemical variables (e.g., iron, aluminum, and measures of NOM like TOC/DOC, and colour) differed widely (Kruskall-Wallis *p* < 0.05) between the years as a result of the flash flooding in 2023. In addition to their effects on physicochemical characteristics, extreme events like this are known to perturb aquatic microbial communities (Shabarova et al., 2021) — including cyanobacteria (Walker et al., 2022). Associations between MC-LR and physicochemical variables identified by the models might be the result of that shared influence, and the predictive ability of TDS and other physicochemical variables may not be generalizable.

To ensure generalizability of our findings, we retrained the models using data from all three years, holding out one location at a time as the test set. With this approach, the discriminatory performance of the models when predicting MC-LR in grab samples was good ([Figure 5](#fig-roc) a), while performance for passive samples was more variable ([Figure 5](#fig-roc) c). Some passive sample models had excellent discriminatory performance (e.g., models tested on Lake Charles 1: AUC = 0.96; Shubenacadie Canal (SC) and Lake Charles 2: AUC = 0.9), whereas others performed poorly (e.g., models tested on TCR: AUC = 0.59; and INP: AUC = 0.54). In both cases, *mcy*E and lagged *mcy*E consistently ranked as the most important predictors ([Figure 5](#fig-roc) b and d). For grab samples, other important predictors included several climate variables, while for passive samples, physicochemical variables played a larger role. As expected, *mcy*E remained the strongest overall predictor of MC-LR, with no other variables in the dataset approaching its importance. However, its utility as a predictor varied greatly between locations, as seen in the discriminatory performance of these random forest models.

|  |
| --- |
| Figure 5: Receiver operating characteristic (ROC) curves for random forest models trained on three years of data, using leave-one-location-out cross-validation, to predict MC-LR presence in grab (**a**) and passive (**c**) samples. Each curve represents model performance on a held-out test location, with the area under the curve (AUC) indicated in the legend. Panels (**b**) and (**d**) show the median relative importance of the 15 most important predictors for grab and passive MC-LR prediction, respectively, with error bars representing the interquartile range across test folds. Shading indicates whether a predictor appeared in the top 15 most important variables for both grab and passive MC-LR models (black) or only for one (grey). |

## 3.6 Lake-and-year specific *mcy*E models offer best fit to observed data

We fit five hierarchical logistic regression models to estimate the probability of detecting extracellular MC-LR in passive samples, and compared them using leave-one-out cross-validation (LOOIC). The first was a null model with only an intercept and random effects for location and year. The remaining four models included a fixed effect for year and location-year-specific smooth functions of sampling week, in addition to: (1) a fixed effect for *mcy*E with location-specific intercepts, (2) *mcy*E with varying intercepts and slopes by location, (3) the top predictors identified by both random forest models (*mcy*E, lagged *mcy*E, TDS, conductivity, chloride, , true colour, and 7-d total precipitation), and (4) the top five predictors from the passive sample random forest model (*mcy*E, lagged *mcy*E, TDS, conductivity, and chloride). The shared and passive predictor models included both fixed and group-specific slopes and intercepts. Missing and censored values in predictors were imputed within a joint modeling framework using Student’s *t*-distribution submodels. Full model specifications are provided in the Supplementary Methods.

The two models that predicted MC-LR as a function of *mcy*E concentration, year, and location-year-specific smooth terms for week of sampling had the lowest LOOIC values, with no meaningful difference between them (Table S4). Adding other predictors did not improve model performance and reduced model stability (large Pareto-*k*). Although the model without varying slopes for *mcy*E was slightly more stable, we selected the model with varying slopes because it most accurately reflects the data generating process. The relationship between the *mcy*E concentration and the probability of MC-LR detection is unlikely to be consistent across lakes. The *mcy*E gene is present in multiple genera with different MC-LR production rates (Beversdorf et al., 2015) and environmental conditions unique to each lake may further influence toxin production (Neilan et al., 2013). The selected model predicts the presence of extracellular MC-LR in passive sample, , at location , as:

Where represents a location- () and year- () specific smooth function of the week of sampling, and and are the random intercepts and slopes, respectively, for the location. This model accurately captured the proportion of MC-LR in passive samples over the entire data set and at each location ([Figure 6](#fig-model-fits)), despite large differences in proportions of MC-LR detections across the sample locations. Posterior predictive checks also confirmed that the predicted standard deviations matched the observed standard deviations (Figure S12).

|  |
| --- |
| Figure 6: Assessment of the logistic regression model used to predict the probability of detecting MC-LR in passive samplers. The predicted probabilities derived from the model closely matched the observed proportion of MC-LR detection in the data (**a**) and posterior draws from the model accurately capture the proportion of MC-LR detections at each sample location (**b**). |

## 3.7 Probabilistic risk associated with *mcy*E varies by location, date, and year

Posterior predictive draws from this model were used to estimate the probability of MC-LR detection over the range of observed *mcy*E values and sampling weeks for each combination of year and location. We used these probabilities to visualize how the *mcy*E gene concentrations associated with 50% ([Figure 7](#fig-0.5prob)), 75% and 95% (Figures S13 and S14) probabilities of detectable MC-LR change over time at each location in each year. These results illustrate that the probabilistic relationship between *mcy*E and detection of low concentrations of MC-LR varies considerably between locations — and over time at each location.

For example, at INP there was a 50% probability of MC-LR at low *mcy*E concentrations early in the season (June through August), whereas at most other locations the *mcy*E concentrations required to reach 50% probability remain high until later in the summer or early fall, indicating that toxin production at INP becomes possible earlier in the season.

Lake Fletcher stands out as the only location where the model predicted a 95% probability of MC-LR detection (Figure S12). At this location and at Shubenacadie Canal (SC) the model also predicted >50% probability of MC-LR for any concentration of *mcy*E. This occurred in all three years at Lake Fletcher but only in 2024 at SC (note that SC was not sampled during 2022), and is in-line with the observed MC-LR detections. At Lake Fletcher we detected MC-LR in nearly every passive sample collected from early-August to October in each of the three years, while at SC every sample was positive for MC-LR from mid-August to November in 2024, but detections were rare during 2023 ([Figure 2](#fig-all-toxins)). These result suggests that Lake Fletcher may be a particularly suitable site for implementing targeted analysis to ensure MC-LR concentrations remain below regulatory guidelines and also suggests that there are external factors at SC that vary year-to-year and play an important role in toxin production.

At most locations, the *mcy*E concentrations required to a reach a 50% probability of MC-LR detection were notably higher in 2023 than in 2022 and 2024 and the seasonal trends in predicted risk were shifted. This was most apparent at Kearney Lake, Lake Fletcher, Lake Charles 1, and SC — locations most impacted by the 2023 flooding event. Altered water chemistry, nutrient loads, and microbial community structure may have delayed and/or suppressed toxin production during that year. This highlights how extreme weather events can introduce variability in gene-toxin relationships, underscoring the importance of considering the broader environmental context when interpreting *mcy*E concentrations.

This approach to evaluating the probability of detectable MC-LR emphasizes the need for tailored, location- and time-specific interpretation of qPCR results. A given *mcy*E concentration cannot be interpreted in isolation — the same gene copy concentration may imply elevated risk at one lake and negligible risk at another, depending on local conditions and seasonal timing — particularly when toxin concentrations are low. These findings reinforce the importance of context-specific interpretation of qPCR data and provide a framework for incorporating molecular tools into early warning systems for cyanotoxins.

Although we observed consistent differences in *mcy*E–MC-LR relationships across locations and years, we did not characterize cyanobacterial community composition directly. It is plausible that variation in dominant genera (e.g., *Microcystis* vs. *Planktothrix*) and their respective *mcy* gene regulation could explain differences in probabilities. Additionally, unmeasured covariates such as light availability or grazing pressure could influence both *mcy*E expression and MC-LR release, introducing latent confounding into model predictions. Future work integrating 16S amplicon or metagenomic data could clarify these community-level dynamics.

|  |
| --- |
| Figure 7: *mcy*E gene concentrations associated with a 50% predicted probability of detectable MC-LR from the hierarchical Bayesian logistic regression model. The lines and ribbons represent the median and interquartile range of the predicted probability across 4000 posterior draws. |

# 4. Conclusion

Our results demonstrate that qPCR targeting the *mcy*E gene, when paired with passive sampling and appropriate modeling, can serve as an effective tool for understanding the probability of low concentrations of MC-LR in freshwater systems. However, we also show that the utility of *mcy*E as a predictor varies substantially across lakes and over time — even within a lake, the relationship between *mcy*E and MC-LR can differ substantially year-to-year.

A given *mcy*E concentration may be associated with increased risk of MC-LR early in the summer at some locations — such as INP — but carry little implication until later in the season at others. Similarly, while *mcy*E was the strongest predictor overall, only two study lakes reached the 95% probability threshold for MC-LR detection, suggesting that predictive strength varies not just in timing but also in magnitude. These findings suggest that universal thresholds are likely not appropriate for identifying the presence of low concentrations of MC-LR — as observed in these study sites — and support the development of tailored, location-specific models that reflect the unique microbial and environmental conditions in each lake.

Together, these insights reinforce the value of integrating molecular tools into toxin surveillance programs, but also emphasize the need for adaptive, context-sensitive interpretation. By modeling the probability of MC-LR detection in relation to gene concentration and time, we offer a flexible, data-informed approach for interpreting qPCR results — an approach that should improve management decision-making across diverse freshwater systems.

# 5. CRediT author statement

**David J Redden** Conceptualization, Methodology, Software, Formal Analysis, Investigation, Writing - Original Draft preparation, Visualization. **Clarke Brown** Conceptualization, Methodology, Investigation, Writing - Review and Editing. **Morgan Harasymchuk** Investigation, Writing - Review and Editing. **Saksham Bafna** Investigation, Writing - Review and Editing. **Justin Laforest** - Investigation, Resources. **Nicole Taylor** - Conceptualization, Resources, Project Administration. **Lindsay Anderson** - Conceptualization, Resources. **Graham A. Gagnon** Conceptualization, Resources, Writing - Review and Editing, Supervision, Funding acquisition.

# 6. Acknowledgements

This study was funded through support from the NSERC Alliance program (ALLRP 568507 - 21) and its member partners, which include LuminUltra Technologies Ltd, Mantech Inc., AGAT Laboratories, Halifax Water, AquiSense Technologies, CBCL Limited, City of Moncton, and Cape Breton Regional Municipality. The authors would also like to extend thanks to Heather Daurie from the Centre for Water Resources Studies at Dalhousie University for her technical support, and to Elizabeth Montgomery and Emma Wattie from Halifax Regional Municipality for their support throughout the study.

# References

Almuhtaram, H., Kibuye, F.A., Ajjampur, S., Glover, C.M., Hofmann, R., Gaget, V., Owen, C., Wert, E.C., Zamyadi, A., 2021. State of knowledge on early warning tools for cyanobacteria detection. Ecological Indicators 133, 108442. <https://doi.org/10.1016/j.ecolind.2021.108442>

Al-Tebrineh, J., Gehringer, M.M., Akcaalan, R., Neilan, B.A., 2011. A new quantitative PCR assay for the detection of hepatotoxigenic cyanobacteria. Toxicon 57, 546–554. <https://doi.org/10.1016/j.toxicon.2010.12.018>

Al-Tebrineh, J., Mihali, T.K., Pomati, F., Neilan, B.A., 2010. Detection of Saxitoxin-Producing Cyanobacteria and *Anabaena circinalis* in Environmental Water Blooms by Quantitative PCR. Applied and Environmental Microbiology 76, 7836–7842. <https://doi.org/10.1128/AEM.00174-10>

Al-Tebrineh, J., Pearson, L.A., Yasar, S.A., Neilan, B.A., 2012. A multiplex qPCR targeting hepato- and neurotoxigenic cyanobacteria of global significance. Harmful Algae 15, 19–25. <https://doi.org/10.1016/j.hal.2011.11.001>

Bermarija, T., Johnston, L., Greene, C., Kurylyk, B., Jamieson, R., 2023. Assessing and predicting lake chloride concentrations in the lake-rich urbanizing halifax region, canada. Journal of Hydrology: Regional Studies 47, 101377. https://doi.org/<https://doi.org/10.1016/j.ejrh.2023.101377>

Beversdorf, L.J., Chaston, S.D., Miller, T.R., McMahon, K.D., 2015. Microcystin mcyA and mcyE Gene Abundances Are Not Appropriate Indicators of Microcystin Concentrations in Lakes. PLOS ONE 10, e0125353. <https://doi.org/10.1371/journal.pone.0125353>

Brophy, M.J., Mackie, A.L., Park, Y., Gagnon, G.A., 2019. Exploring the detection of microcystin-LR using polar organic chemical integrative samplers (POCIS). Environmental Science: Processes & Impacts 21, 659–666. <https://doi.org/10.1039/C8EM00569A>

Buratti, F.M., Manganelli, M., Vichi, S., Stefanelli, M., Scardala, S., Testai, E., Funari, E., 2017. Cyanotoxins: Producing organisms, occurrence, toxicity, mechanism of action and human health toxicological risk evaluation. Archives of Toxicology 91, 1049–1130. <https://doi.org/10.1007/s00204-016-1913-6>

Bürkner, P.-C., 2017. brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software 80, 1–28. <https://doi.org/10.18637/jss.v080.i01>

Chorus, I., Welker, M., 2021. Toxic Cyanobacteria in Water; A Guide to Their Public Health Consequences, Monitoring and Management; Second Edition, 2nd ed. CRC Press on behalf of the World Health Organization, Boca Raton.

Christiansen, G., Fastner, J., Erhard, M., Börner, T., Dittmann, E., 2003. Microcystin Biosynthesis in Planktothrix: Genes, Evolution, and Manipulation. Journal of Bacteriology 185, 564–572. <https://doi.org/10.1128/jb.185.2.564-572.2003>

Christiansen, G., Kurmayer, R., Liu, Q., Börner, T., 2006. Transposons inactivate biosynthesis of the nonribosomal peptide microcystin in naturally occurring planktothrix spp. Applied and Environmental Microbiology 72, 117–123. <https://doi.org/10.1128/AEM.72.1.117-123.2006>

Christiansen, G., Molitor, C., Philmus, B., Kurmayer, R., 2008. Nontoxic strains of cyanobacteria are the result of major gene deletion events induced by a transposable element. Molecular Biology and Evolution 25, 1695–1704. <https://doi.org/10.1093/molbev/msn120>

Dai, R., Wang, P., Jia, P., Zhang, Y., Chu, X., Wang, Y., 2016. A review on factors affecting microcystins production by algae in aquatic environments. World Journal of Microbiology and Biotechnology 32, 51. <https://doi.org/10.1007/s11274-015-2003-2>

Dinno, A., 2024. [Dunn.test: Dunn’s test of multiple comparisons using rank sums](https://CRAN.R-project.org/package=dunn.test).

Dittmann, E., Neilan, B.A., Erhard, M., Von Döhren, H., Börner, T., 1997. Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium Microcystis aeruginosa PCC 7806. Molecular Microbiology 26, 779–787. <https://doi.org/10.1046/j.1365-2958.1997.6131982.x>

Duan, X., Zhang, C., Struewing, I., Li, X., Allen, J., Lu, J., 2022. Cyanotoxin-encoding genes as powerful predictors of cyanotoxin production during harmful cyanobacterial blooms in an inland freshwater lake: Evaluating a novel early-warning system. Science of The Total Environment 830, 154568. <https://doi.org/10.1016/j.scitotenv.2022.154568>

Engene, N., Gerwick, W.H., 2011. Intra-genomic 16S rRNA gene heterogeneity in cyanobacterial genomes. Fottea 11, 17–24. <https://doi.org/10.5507/fot.2011.003>

Environment and Climate Change Canada, 2025. [Historical climate data](https://climate.weather.gc.ca/).

Feist, S.M., Lance, R.F., 2021. Genetic detection of freshwater harmful algal blooms: A review focused on the use of environmental DNA (eDNA) in Microcystis aeruginosa and Prymnesium parvum. Harmful Algae 110, 102124. <https://doi.org/10.1016/j.hal.2021.102124>

Gabry, J., Mahr, T., 2024. [Bayesplot: Plotting for bayesian models](https://mc-stan.org/bayesplot/).

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., Gelman, A., 2019. Visualization in bayesian workflow. J. R. Stat. Soc. A 182, 389–402. <https://doi.org/10.1111/rssa.12378>

Godlewska, K., Stepnowski, P., Paszkiewicz, M., 2021. Pollutant analysis using passive samplers: Principles, sorbents, calibration and applications. A review. Environmental Chemistry Letters 19, 465–520. <https://doi.org/10.1007/s10311-020-01079-6>

Gonzales Ferraz, M.E., Agasild, H., Piirsoo, K., Saat, M., Nõges, T., Panksep, K., 2024. Seasonal dynamics of toxigenic Microcystis in a large, shallow Lake Peipsi (Estonia) using microcystin mcyE gene abundance. Environmental Monitoring and Assessment 196, 747. <https://doi.org/10.1007/s10661-024-12909-3>

Harman, C., Allan, I.J., Vermeirssen, E.L.M., 2012. Calibration and use of the polar organic chemical integrative samplera critical review. Environmental Toxicology and Chemistry 31, 2724–2738. <https://doi.org/10.1002/etc.2011>

Hayes, E.K., Sweeney, C.L., Anderson, L.E., Li, B., Erjavec, G.B., Gouthro, M.T., Krkosek, W.H., Stoddart, A.K., Gagnon, G.A., 2021. A novel passive sampling approach for SARS-CoV-2 in wastewater in a Canadian province with low prevalence of COVID-19. Environmental Science: Water Research & Technology 7, 1576–1586. <https://doi.org/10.1039/D1EW00207D>

Helsel, D.R., 2011. Statistics for censored environmental data using minitab and r. John Wiley & Sons.

Hester, J., Bryan, J., 2024. [Glue: Interpreted string literals](https://CRAN.R-project.org/package=glue).

Hisbergues, M., Christiansen, G., Rouhiainen, L., Sivonen, K., Börner, T., 2003. PCR-based identification of microcystin-producing genotypes of different cyanobacterial genera. Archives of Microbiology 180, 402–410. <https://doi.org/10.1007/s00203-003-0605-9>

Jaša, L., Sadílek, J., Kohoutek, J., Straková, L., Maršálek, B., Babica, P., 2019. Application of passive sampling for sensitive time-integrative monitoring of cyanobacterial toxins microcystins in drinking water treatment plants. Water Research 153, 108–120. <https://doi.org/10.1016/j.watres.2018.12.059>

Jungblut, A.-D., Neilan, B.A., 2006. Molecular identification and evolution of the cyclic peptide hepatotoxins, microcystin and nodularin, synthetase genes in three orders of cyanobacteria. Archives of Microbiology 185, 107–114. <https://doi.org/10.1007/s00203-005-0073-5>

Kaebernick, M., Dittmann, E., Börner, T., Neilan, B.A., 2002. Multiple alternate transcripts direct the biosynthesis of microcystin, a cyanobacterial nonribosomal peptide. Applied and Environmental Microbiology 68, 449–455. <https://doi.org/10.1128/AEM.68.2.449-455.2002>

Kaebernick, M., Rohrlack, T., Christoffersen, K., Neilan, B.A., 2001. A spontaneous mutant of microcystin biosynthesis: Genetic characterization and effect on Daphnia. Environmental Microbiology 3, 669–679. <https://doi.org/10.1046/j.1462-2920.2001.00241.x>

Kaneko, T., Nakajima, N., Okamoto, S., Suzuki, I., Tanabe, Y., Tamaoki, M., Nakamura, Y., Kasai, F., Watanabe, A., Kawashima, K., Kishida, Y., Ono, A., Shimizu, Y., Takahashi, C., Minami, C., Fujishiro, T., Kohara, M., Katoh, M., Nakazaki, N., Nakayama, S., Yamada, M., Tabata, S., Watanabe, M.M., 2007. Complete Genomic Structure of the Bloom-forming Toxic Cyanobacterium Microcystis aeruginosa NIES-843. DNA Research 14, 247–256. <https://doi.org/10.1093/dnares/dsm026>

Kay, M., 2024. tidybayes: Tidy data and geoms for Bayesian models. <https://doi.org/10.5281/zenodo.1308151>

Kibuye, F.A., Almuhtaram, H., Zamyadi, A., Gaget, V., Owen, C., Hofmann, R., Wert, E.C., 2021. Utility practices and perspectives on monitoring and source control of cyanobacterial blooms. AWWA Water Science 3, e1264. <https://doi.org/10.1002/aws2.1264>

Kohoutek, J., Maršálek, B., Bláha, L., 2010. Evaluation of the novel passive sampler for cyanobacterial toxins microcystins under various conditions including field sampling. Analytical and Bioanalytical Chemistry 397, 823–828. <https://doi.org/10.1007/s00216-010-3578-y>

Kudela, R.M., 2011. Characterization and deployment of Solid Phase Adsorption Toxin Tracking (SPATT) resin for monitoring of microcystins in fresh and saltwater. Harmful Algae 11, 117–125. <https://doi.org/10.1016/j.hal.2011.08.006>

Kuhn, M., Wickham, H., 2020. [Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles.](https://www.tidymodels.org)

Lance, E., Lepoutre, A., Savar, V., Robert, E., Bormans, M., Amzil, Z., 2021. *In situ* use of bivalves and passive samplers to reveal water contamination by microcystins along a freshwater-marine continuum in france. Water Research 204, 117620. <https://doi.org/10.1016/j.watres.2021.117620>

Loaiza-González, J.M., Rubio-Clemente, A., Peñuela, G.A., 2024. Cyanotoxin monitoring and detection using passive sampling application. Water, Air, & Soil Pollution 235, 423. https://doi.org/<https://doi.org/10.1007/s11270-024-07195-1>

Lu, J., Struewing, I., Wymer, L., Tettenhorst, D.R., Shoemaker, J., Allen, J., 2020. Use of qPCR and RT-qPCR for monitoring variations of microcystin producers and as an early warning system to predict toxin production in an Ohio inland lake. Water Research 170, 115262. <https://doi.org/10.1016/j.watres.2019.115262>

Madany, P., Olsen, D., Makled, S.O., Cram, E., Page, S., Bonner, H., McLean, J.E., Stevens, D., Li, M., Hou, L., 2025. Innovative multiplex qPCR method for rapid and reliable detection of microcystin-producing genes during harmful algal blooms: Insights from utah reservoirs. Water Research 277, 123322. <https://doi.org/10.1016/j.watres.2025.123322>

Mayer, M., 2024. [missRanger: Fast imputation of missing values](https://CRAN.R-project.org/package=missRanger).

Miller, T.R., Beversdorf, L.J., Weirich, C.A., Bartlett, S.L., 2017. Cyanobacterial Toxins of the Laurentian Great Lakes, Their Toxicological Effects, and Numerical Limits in Drinking Water. Marine Drugs 15, 160. <https://doi.org/10.3390/md15060160>

Neilan, B.A., Pearson, L.A., Muenchhoff, J., Moffitt, M.C., Dittmann, E., 2013. Environmental conditions that influence toxin biosynthesis in cyanobacteria. Environmental Microbiology 15, 1239–1253. <https://doi.org/10.1111/j.1462-2920.2012.02729.x>

Nishizawa, T., Ueda, A., Asayama, M., Fujii, K., Harada, K.-i., Ochi, K., Shirai, M., 2000. Polyketide Synthase Gene Coupled to the Peptide Synthetase Module Involved in the Biosynthesis of the Cyclic Heptapeptide Microcystin. Journal of Biochemistry 127, 779–789. <https://doi.org/10.1093/oxfordjournals.jbchem.a022670>

Noguchi, T., Shinohara, A., Nishizawa, A., Asayama, M., Nakano, T., Hasegawa, M., Harada, K., Nishizawa, T., Shirai, M., 2009. Genetic analysis of the microcystin biosynthesis gene cluster in microcystis strains from four bodies of eutrophic water in japan. The Journal of General and Applied Microbiology 55, 111–123. <https://doi.org/10.2323/jgam.55.111>

Ohio EPA, 2018. [Ohio EPA Quantitative Polymerase Chain Reaction (qPCR) Multi‑Plex Molecular Assay for Determination of Cyanobacteria and Cyanotoxin‑Producing Genes: Analytical Methodology](https://dam.assets.ohio.gov/image/upload/epa.ohio.gov/Portals/28/documents/habs/705.0-qPCR.pdf), Version 1.0. ed. Ohio Environmental Protection Agency, Division of Environmental Services, Columbus, Ohio.

Pacheco, A.B.F., Guedes, I.A., Azevedo, S.M.F.O., 2016. Is qPCR a Reliable Indicator of Cyanotoxin Risk in Freshwater? Toxins 8, 172. <https://doi.org/10.3390/toxins8060172>

Padovan, A., Kennedy, K., Gibb, K., 2023. A microcystin synthesis mcyE/ndaF gene assay enables early detection of microcystin production in a tropical wastewater pond. Harmful Algae 127, 102476. <https://doi.org/10.1016/j.hal.2023.102476>

Panksep, K., Tamm, M., Mantzouki, E., Rantala-Ylinen, A., Laugaste, R., Sivonen, K., Tammeorg, O., Kisand, V., 2020. Using Microcystin Gene Copies to Determine Potentially-Toxic Blooms, Example from a Shallow Eutrophic Lake Peipsi. Toxins 12, 211. <https://doi.org/10.3390/toxins12040211>

Posit team, 2025. [RStudio: Integrated development environment for r](http://www.posit.co/). Posit Software, PBC, Boston, MA.

R Core Team, 2024. [R: A language and environment for statistical computing](https://www.R-project.org/). R Foundation for Statistical Computing, Vienna, Austria.

Rastogi, R.P., Sinha, R.P., Incharoensakdi, A., 2014. The cyanotoxin-microcystins: Current overview. Reviews in Environmental Science and Bio/Technology 13, 215–249. <https://doi.org/10.1007/s11157-014-9334-6>

Redden, D.J., 2025. Source code for “qPCR-based prediction of low-level microcystin-LR using mcyE and passive sampling across multiple lakes over three years.” Available at <https://github.com/djredden/qPCR-Based-MCLR-Prediction>.

Rouhiainen, L., Vakkilainen, T., Siemer, B.L., Buikema, W., Haselkorn, R., Sivonen, K., 2004. Genes coding for hepatotoxic heptapeptides (microcystins) in the cyanobacterium Anabaena strain 90. Applied and Environmental Microbiology 70, 686–692. <https://doi.org/10.1128/AEM.70.2.686-692.2004>

Schirrmeister, B.E., Dalquen, D.A., Anisimova, M., Bagheri, H.C., 2012. Gene copy number variation and its significance in cyanobacterial phylogeny. BMC Microbiology 12, 177. <https://doi.org/10.1186/1471-2180-12-177>

Schürmann, Q.J.F., Visser, P.M., Sollie, S., Kardinaal, W.E.A., Faassen, E.J., Lokmani, R., Oost, R. van der, Van de Waal, D.B., 2024. Risk assessment of toxic cyanobacterial blooms in recreational waters: A comparative study of monitoring methods. Harmful Algae 138, 102683. <https://doi.org/10.1016/j.hal.2024.102683>

Shabarova, T., Salcher, M.M., Porcal, P., Znachor, P., Nedoma, J., Grossart, H.-P., Seďa, J., Hejzlar, J., Šimek, K., 2021. Recovery of freshwater microbial communities after extreme rain events is mediated by cyclic succession. Nature Microbiology 6, 479–488. <https://doi.org/10.1038/s41564-020-00852-1>

Sipari, H., Rantala-Ylinen, A., Jokela, J., Oksanen, I., Sivonen, K., 2010. Development of a chip assay and quantitative PCR for detecting microcystin synthetase e gene expression. Applied and Environmental Microbiology 76, 3797–3805. <https://doi.org/10.1128/AEM.00452-10>

Stebegg, R., Schmetterer, G., Rompel, A., 2023. Heterotrophy among cyanobacteria. ACS Omega 8, 33098–33114. <https://doi.org/10.1021/acsomega.3c02205>

Tanabe, Y., Kaya, K., Watanabe, M.M., 2004. Evidence for Recombination in the Microcystin Synthetase (mcy) Genes ofToxic Cyanobacteria Microcystis spp. Journal of Molecular Evolution 58, 633–641. <https://doi.org/10.1007/s00239-004-2583-1>

Tillett, D., Dittmann, E., Erhard, M., Döhren, H. von, Börner, T., Neilan, B.A., 2000. Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: An integrated peptide–polyketide synthetase system. Chemistry & Biology 7, 753–764. <https://doi.org/10.1016/S1074-5521(00)00021-1>

Trueman, B., 2024. [bgamcar1: Fit bayesian GAMs with CAR(1) errors to censored data](https://github.com/bentrueman/bgamcar1).

USEPA, 2015. [Health effects support document for the cyanobacterial toxin microcystins](https://www.epa.gov/sites/default/files/2017-06/documents/microcystins-support-report-2015.pdf) (EPA Technical Report No. EPA‑820R15102). United States Environmental Protection Agency, Office of Water, Health and Ecological Criteria Division, Washington, DC.

Vaitomaa, J., Rantala, A., Halinen, K., Rouhiainen, L., Tallberg, P., Mokelke, L., Sivonen, K., 2003. Quantitative real-time PCR for determination of microcystin synthetase e copy numbers for microcystis and anabaena in lakes. Applied and Environmental Microbiology 69, 7289–7297. <https://doi.org/10.1128/AEM.69.12.7289-7297.2003>

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., Gelman, A., 2024. [Loo: Efficient leave-one-out cross-validation and WAIC for bayesian models](https://mc-stan.org/loo/).

Vehtari, A., Gelman, A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27, 1413–1432. <https://doi.org/10.1007/s11222-016-9696-4>

Walker, J.R., Woods, A.C., Pierce, M.K., Steichen, J.L., Quigg, A., Kaiser, K., Labonté, J.M., 2022. Functionally diverse microbial communities show resilience in response to a record-breaking rain event. ISME Communications 2, 81. <https://doi.org/10.1038/s43705-022-00162-z>

Wang, P., Du, B., Smith, J., Lao, W., Wong, C.S., Zeng, E.Y., 2022. Development and field evaluation of the organic-diffusive gradients in thin-films (o-DGT) passive water sampler for microcystins. Chemosphere 287, 132079. <https://doi.org/10.1016/j.chemosphere.2021.132079>

Wei, N., Hu, C., Dittmann, E., Song, L., Gan, N., 2024. The biological functions of microcystins. Water Research 262, 122119. <https://doi.org/10.1016/j.watres.2024.122119>

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. Journal of Open Source Software 4, 1686. <https://doi.org/10.21105/joss.01686>

Wilke, C.O., Wiernik, B.M., 2022. [Ggtext: Improved text rendering support for ’ggplot2’](https://CRAN.R-project.org/package=ggtext).

Wiltsie, D., Schnetzer, A., Green, J., Vander Borgh, M., Fensin, E., 2018. Algal Blooms and Cyanotoxins in Jordan Lake, North Carolina. Toxins 10, 92. <https://doi.org/10.3390/toxins10020092>

Wright, M.N., Ziegler, A., 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. Journal of Statistical Software 77, 1–17. <https://doi.org/10.18637/jss.v077.i01>

Xiao, N., 2024. [Ggsci: Scientific journal and sci-fi themed color palettes for ’ggplot2’](https://CRAN.R-project.org/package=ggsci).

Zeileis, A., Grothendieck, G., 2005. Zoo: S3 infrastructure for regular and irregular time series. Journal of Statistical Software 14, 1–27. <https://doi.org/10.18637/jss.v014.i06>