DETECTION AND QUANTIFICATION OF TASTE AND ODOR PRODUCING BACTERIA IN EAGLE CREEK RESERVOIR

by

Ioanna Koltsidou

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Biological Sciences

Indianapolis, Indiana

August 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Christine J. Picard, Chair

School of Science, Department of Biology

Dr. Gregory G. Anderson

School of Science, Department of Biology

Dr. Gregory K. Druschel

School of Science, Department of Earth Sciences

Approved by:

Dr. Theodore R. Cummins

Head of the Graduate Program

I would like to dedicate this work to my parents and siblings who have always been supportive of my choices. Also, to Tasos, Christina and Dimitra, people who will always have a special place in my heart.

ACKNOWLEDGMENTS

I would first like to express my gratidute to my advisor Dr. Christine Picard, for the continuous support of my M.S research, guidance and motivation. Also, to Dr. Gregory Druschel who offered me the opportunity to work on this research project and has shared his immense knowledge and enthusiasm. My sincere thanks also goes to Dr. Gregory Anderson, for generously offering his time to serve on my committee.

I would also like to thank my fellow lab mates: Anne Andere, Laura Doll, Sarah Lewis and Charity Owings, for the valuable advice, moral support and all the fun we have had in the last two years. A special thanks to Chase Howard, Igor Ogashawara and Sarah Denny who assisted me with field work and sample collection. Particularly helpful to me during this time were Martin Kurek and John Shuckle who were always willing to provide their help and constructive feedback.

Finally, this research would not have been possible without the financial support of Citizens Water Indianapolis, a branch of the Citizens Energy Group.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ECR: Eagle Creek Reservoir

MEP: 2-methyl-erythritol-4-phosphate

MIB: 2-methyl-isoborneol

MVA: Mevalonate

OTU: Operational Taxonomic Unit

T&O: Taste-and-Odor

ABSTRACT

Author: Koltsidou, Ioanna. MS
Institution: Purdue University
Degree Received: August 2019
Title: Detection and Quantification of Taste and Odor Producing Bacteria in Eagle Creek Reservoir.
Committee Chair: Christine Picard

The accelerated growth of algal blooms in water bodies has caused the increased occurrence of taste and odor (T&O) episodes worldwide. Even though T&O compounds have not been associated with adverse health effects, their presence can have extensive socio-economic impacts in contaminated waters. Eagle Creek Reservoir, a eutrophic water body, which supplies about 80% of Indianapolis drinking water, experiences frequent and sometimes severe odorous outbreaks. The terpenoid bacterial metabolites, 2-methylisoborneol (2-MIB) and geosmin, have been identified as the main compounds contributing to those T&O problems, which occur seasonally when the reservoir receives most of its water and nutrient loads from discharge events. In this study, ECR's microbial community composition was assessed by a 16S next generation sequencing approach, confirming the presence of the major bacterial phyla of Cyanobacteria, Proteobacteria, Actinobacteria and Bacteroidetes, which are commonly found in freshwater environments. The relative abundance of Cyanobacteria, which are regarded as the main T&O producers in freshwater, followed the fluctuation of 2-MIB and geosmin concentrations closely. Mapping sequence analysis of a metagenomic dataset, successfully recovered the genes responsible for the synthesis of geosmin and 2-MIB, demonstrating the microbial ability for odorous compound production in ECR. Quantification of the *geoA* and *MIBS* genes in Cyanobacteria was achieved by the development and application of qPCR assays on water samples collected from the reservoir. A statistically significant positive correlation was found between *MIBS* gene quantity and MIB concentration for all sampling locations, implying that this assay could potentially be used as a tool for the early prediction of upcoming T&O episodes. The *geoA* gene detection assay, did not correlate well with geosmin concentrations, suggesting that even though the gene might be present, this does not necessarily mean that it is metabolically active.

CHAPTER 1. INTRODUCTION

1.1 Global Rise of Taste and Odor Episodes

Water is one of the most important substances and an essential component for the preservation of life on this planet. Since water covers 70% of the earth, we often think of it as being inexhaustible and plentiful, however, less than 1% of freshwater is available making its proper conservation necessary for sustainable use [*Lee et al.,* 2017]. In direct threats to water conservation is the continued rise of urbanization and intensive industrial and agricultural practices, leading to eutrophication of freshwater ecosystems, a process that favors the proliferation and dominance of bacterial blooms [*Gkelis et al.,* 2014; *Paerl and Paul,* 2012]. Contained with those blooms are a group of bacteria known as cyanobacteria, bacteria that harness the suns energy and produce oxygen photosynthetically. Although cyanobacteria have existed on earth on Earth since ancient times and blooms are a natural phenomenon, reports and studies show that they are currently increasing worldwide and becoming an issue of a great concern [*Huisman et al.,* 2018; *Paerl and Paul,* 2012*; Winter et al.* 2011]. Alarmingly, a recent study developed a model predicting the effect of climate change and anthropogenic activities on cyanobacteria, and found that in the USA the mean number of days with Cyanobacterial blooms will increase from 7 days per year per waterbody under current conditions to 18-39 days in 2090 [*Chapra et al.* 2017; *Huisman et al.,* 2018]. This accelerated growth of algal blooms in water bodies has a secondary effect – the production of metabolites, which cause the occurrence of taste and odor (T&O) episodes in diverse waters such as lakes, fishponds and reservoirs [*Lee et al.,* 2017]. Even though taste and odor compounds have not been associated with any serious biological or pathological effects, consumers perceive water as unsafe to drink [*Giglio et al.,* 2010]. Therefore, water management utilities are called to deal with this problem and develop methods for the early detection and prediction of emerging issues in order to avoid economic losses and gain consumer trust.

1.2 Eagle Creek Reservoir

Eagle Creek Reservoir (ECR;86°18′13.07″W, 39°51′09.84″N; A = 5.0 km2; Z = 4.2 m), was constructed in 1967 in the northwest corner of Marion County, Indiana, to initially provide

flood control and subsequently drinking water when the T.W. Moses Drinking Water Plant came online in 1976 [*Harris et al.,* 2016; *Li et al.,* 2006]. Currently, ECR is part of a major drinking and recreational water system, along with Geist and Morse Reservoirs, which supplies potable water for over 900,000 residents of the Indianapolis Metropolitan region [*Song et al.,* 2012]. The land use in Eagle Creek watershed is primarily agricultural (about 60%), resulting in high nutrient loads and contributing to the classification of the reservoir as being mesotrophic to eutrophic [*Song et al.,* 2012]. In the last few years, drinking water managers have documented algal blooms of taste and odor causing microorganisms, with the problem starting to become more frequent and sometimes severe in the early 2000's [*Li et al.,* 2006*; Pascual and Tedesco,* 2006]. The odorous events are usually observed during the Spring, when the reservoir receives most of its water and is fully mixed and turbid, and at the beginning of Fall. In order to identify the sources of the problem and develop efficient management strategies, Citizens Water Indianapolis, the water division of Citizens Energy Group, started a long-term research partnership with the Center for Earth and Environmental Science at IUPUI.

1.3 Geosmin and 2-MIB

Most of the taste and odor episodes in drinking water are caused by the terpenoid bacterial metabolites 2-MIB (2-methylisoborneol) and geosmin (trans-1,10-dimethyl-trans-9-decalol) which produce a mildewed and earthy flavor respectively [*Lee et al.,*2017; *Jüttner and Watson,* 2007; *Watson et al.,* 2008]. Each compound exists as (+) and (-) enantiomers (Figure 1.1), but odor outbreaks are caused by the biological production of the (-) stereoisomer which is 10 times more potent than the (+) molecules [*Jüttner and Watson,* 2007]. Their hydrophobic properties help them penetrate and accumulate in fish flesh, reducing the quality of freshwater aquaculture [*Guttman and Rijn,* 2009; *Klausen et al.,* 2005]. Since geosmin and 2-MIB were first isolated from actinomycetes and identified by *Gerber et al.* [1965] and *Medsker et al.* [1969], much research has revolved around them [*Giglio et al.,* 2008; *Komatsu et al.,* 2008]. One of the main reasons being that due to their strong resistance to oxidation, a process routinely applied to water purification, and their low detection threshold by human senses (2-10 ng/L), they tend to be the most common substances in odor incidents [*Auffret et al.,* 2011; *Suurnäkki et al.,* 2015; *Wang et al.,* 2016].

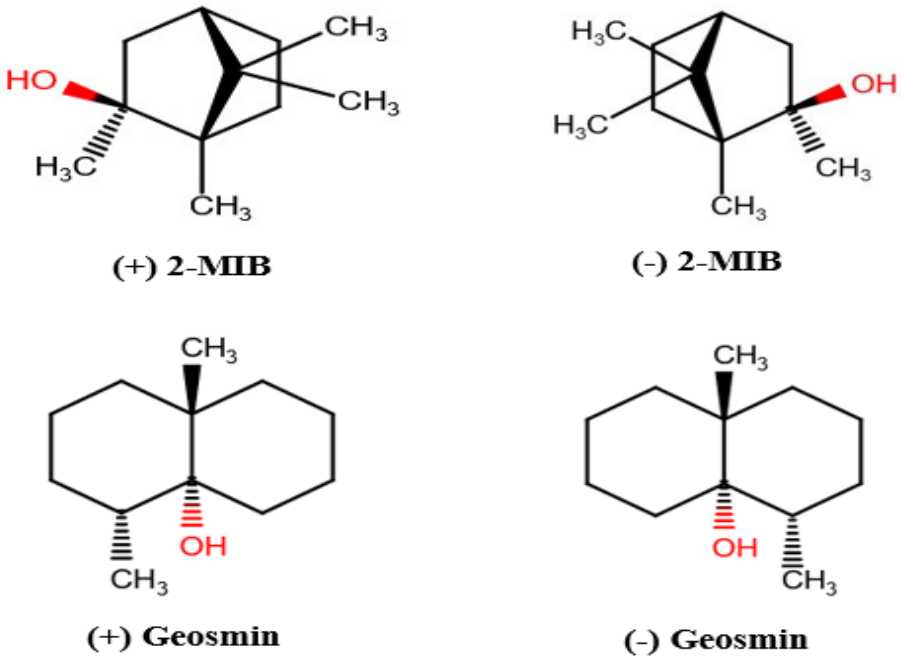


Figure 1.1 Chemical structures of 2-MIB and geosmin enantiomers [*Watson,* 2003].

1.4 Biological Sources

The secondary metabolites geosmin and 2-MIB are produced by a wide range of microorganisms, including eukaryotes such as fungi, amoeba, liverworts and beets, but mainly prokaryotes belonging to cyanobacteria, actinomycetes, proteobacteria and myxobacteria [*Giglio et al.,* 2008; *Jüttner and Watson,* 2007]. In freshwater environments, planktonic and benthic cyanobacteria are considered the main producers of T&O compounds, since oftentimes strong correlations have been reported between microscopic cell counts and concentrations of geosmin and 2-MIB [*Asquith et al.,* 2013; *Chiu et al.,* 2016; *Tsao et al.,* 2014; *Watson et al.,* 2008]. Cyanobacteria are oxygen producing bacteria that use sunlight as an energy source to convert carbon dioxide into biomass and can rapidly increase, forming blooms that are frequently associated with the production of odorous metabolites [*Huisman et al.,* 2018; *Su et al.,* 2013; *Tsao et al.,* 2014]. More than 40 species have been confirmed as MIB and/or geosmin producers, including strains of *Anabaena, Lyngbya, Planktothrix, Aphanizomenon, Oscillatoria, Pseudanabaena* [*Asquith et al.,* 2013; *Chiu et al.,* 2016; *Jüttner and Watson,* 2007; *Lee et al.,* 2017; *Zhang et al.,* 2009]. The synthesis, production and secretion of those metabolites depends on

environmental factors and microbial population dynamics, but most of the intracellular pools get released during senescence, lysis and biodegradation of the cells [*Lee et al.*, 2017; *Watson*, 2003; *Watson et al.*, 2008]. Cyanobacteria belong to a morphologically diverse group and greatly differ in their metabolite chemistry and production dynamics [*John et al.*, 2018; *Watson et al.*, 2003; *Watson et al.*, 2008]. Consequently, even closely related species and different stains of the same species, may produce varying amounts of 2-MIB and geosmin, making it difficult to anticipate and predict emerging T&O episodes. Perhaps more importantly, it is increasingly difficult to trace their origin, as many times microbial biomass and subsequent metabolites do not follow closely one another.

Although cyanobacteria are regarded as the primary producers of geosmin and 2-MIB, the presence of those compounds during periods of low cyanobacterial biomass indicates that there are other bacteria also contributing to the T&O problem [*Asquith et al.*, 2013]. Actinobacteria, which belong to a class of Gram-positive heterotrophic bacteria with DNA high in GC content, also have the ability to produce geosmin and 2-MIB in aquatic environments [*Asquith et al.*, 2013; *Klausen et al.*, 2005]. They are typically characterized as soil bacteria, but they can be found in aerobic, organic-rich environments, such as the water-sediment interface, where they usually get introduced through terrestrial runoff, and in some cases constitute more than 60% of the bacterial community population [*Jüttner and Watson*, 2007; *Klausen et al.*, 2005; *Lyllof et al.*, 2012]. Initially, since the two compounds were first identified and isolated from *Streptomyces*, these species were regarded as the main producers. However, further research revealed other actinobacterial odor producers such as species of *Nocardia*, *Arthrobacter* and *Micromonospora* [*Jüttner and Watson*, 2007; *Lindholm-Lehto and Vielma*, 2018]. In some cases, species can produce either geosmin or 2-MIB, and some produce both [*Zaitlin et al.*, 2003]. In the case of *Streptomyces*, which produces both, following sporulation and during mycelial growth, demonstration of a direct relationship with odor production was not always straightforward, since spores that are not metabolically active cannot be distinguished from vegetative forms when using traditional cell count techniques [*Asquith et al.*, 2013]. Furthermore, Actinobacteria display diverse morphologies and have variable metabolic and cellular capacities, resulting in inconsistent odor production by isolates that have identical 16S rRNA sequences. This means, the detection of the origin of the problem becomes more difficult and T&O episodes are not well predicted [*Asquith et al.*, 2013; *Klausen et al.*, 2005].

1.5 Biosynthetic Pathways and Production Mechanism

Although geosmin and 2-MIB were identified in the late 1960's, it was not until much later that their biosynthetic pathways were elucidated. *Bentley and Meganathan* [1981], first demonstrated through labeling experiments on *Streptomyces,* that the compounds are terpenes, derived from precursors of the isoprenoid pathway. They concluded that 2-MIB is a methylated monoterpene and geosmin is most likely derived from a sesquiterpenoid precursor. There are three major biosynthetic pathways of isoprenoid synthesis that may lead to the production of geosmin and 2-MIB by different microorganisms (Figure 1.2) [*Jüttner and Watson,* 2007]. Bacterial taxa primarily use the 2-methylerytthritol-4-phosphate (MEP) pathway but some organisms have the ability to preferentially choose between the MEP and mevalonic acid (MVA) pathway, depending on their growth stage [*Jüttner and Watson,* 2007; *Rodríguez-Concepción and Boronat,* 2002]. Both pathways lead to the production of isopentenyl diphosphate (IPP), a 5-carbon precursor for geranyl diphosphate and farnesyl diphosphate, ultimately leading to 2-MIB and geosmin, respectively [*Asquith et al.,* 2013]. The MEP pathway synthesizes IPP in plastids, whereas the MVA pathway produces cytosolic IPP [*Rodríguez-Concepción and Boronat,* 2002].

Significant progress in the understanding of genes and enzymes involved in the production mechanism of geosmin has been recently made. *Cane et al.* [2006], showed that the geosmin synthase gene (*geoA)* is essential for geosmin biosynthesis, by demonstrating the inability of *Streptomyces avermilitis* mutants for the *geoA* gene to produce either germacradienol or geosmin. The same was reported in cyanobacteria, based on the presence of a single gene encoding the geosmin synthase enzyme [*Giglio et al.,* 2008]. The production of geosmin happens through the conversion of farnesyl diphosphate, the universal C_{15} sesquiterpene precursor, to geosmin catalyzed by geosmin synthase encoded by the *geoA* gene (Figure 1.3) [*Giglio et al.,* 2008; *Giglio et al.,* 2010].

Labeling experiments have shown that 2-MIB is a methylated monoterpene alcohol, and the additional methyl group is derived from S-adenosyl-L-methionine (SAM) [*Bentley and Meganathan,* 1981; *Komastu et al.,* 2008; *Wang et al.,* 2011]. In combination with the feeding experiments conducted by *Dickschat et al.,* [2007], which demonstrated that methylation of geranyl diphosphate (GPP) produces the substrate for 2-MIB cyclase and then subsequent formation of 2-MIB occurs, it was suggested that there are two steps in the synthesis of 2-MIB

[*Wang et al.*, 2011]. First, we have the S-adenosylmethionine-dependent methylation of GPP to 2-methyl-GPP and then its subsequent cyclization to 2-MIB (Figure 1.4) [*Asquith et al.*, 2013; *Wang et al.*, 2011]. The genes involved in the biosynthesis of 2-MIB, were identified by *Komatsu et al.*, [2008], via bioinformatic tools to screen 2-MIB producing actinomycetes and found two genes forming an MIB synthesis operon. Those genes are GPP-2-methyltranferase (GPPMT) which catalyzes the first step of the reaction and a monoterpene cyclase (2-MIB synthase, MIBS) that is involved in the cyclization step. *Giglio et al.*, [2010], revealed that the same two reactions are present in 2-MIB synthesis in cyanobacteria and demonstrated the successful activity of GPPMT and MIBS genes in *Pseudanabaena* through incubation experiments with GPP and SAM. The 2-MIB synthesis genes in cyanobacteria are homologous to actinomycetes genes, suggesting ancestral origins for this biosynthetic pathway [*Wang et al.*, 2011].

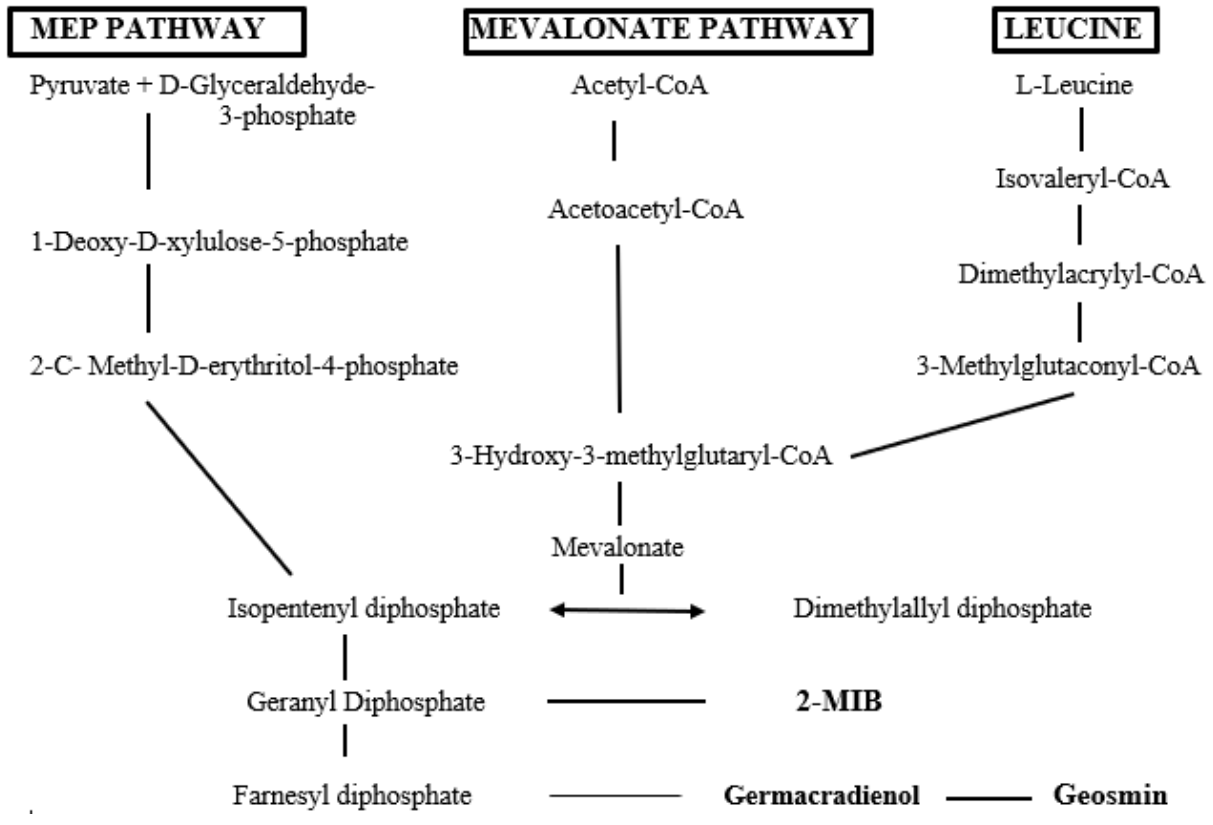


Figure 1.2 Simplified biosynthetic pathways for the formation of 2-MIB and geosmin [*Jüttner and Watson*, 2007].

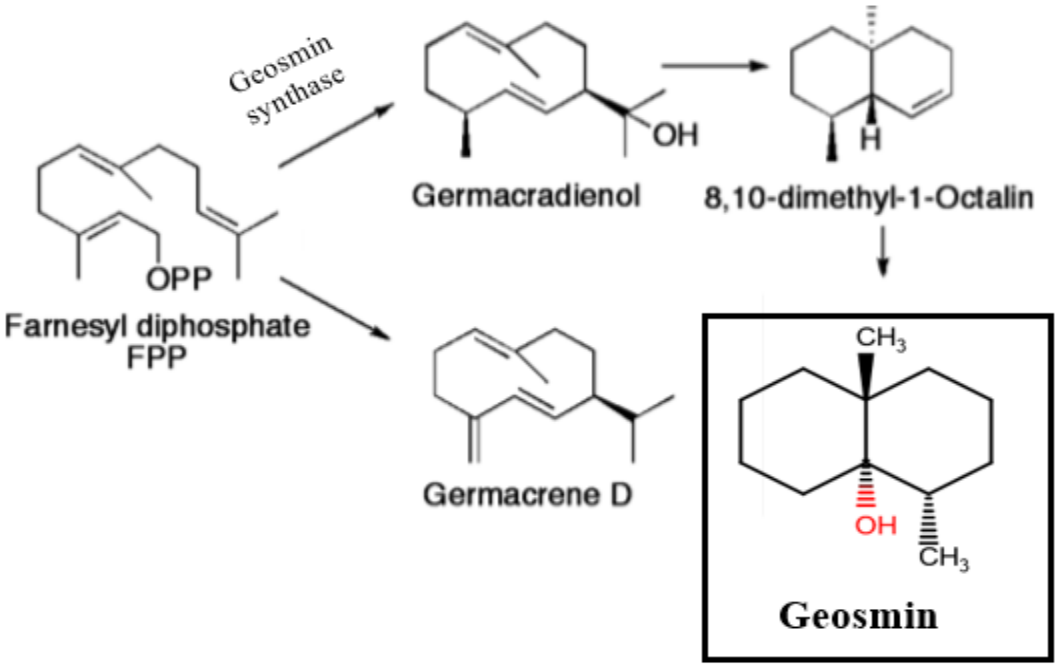


Figure 1.3 Production mechanism of geosmin [*Giglio et al.*, 2008].

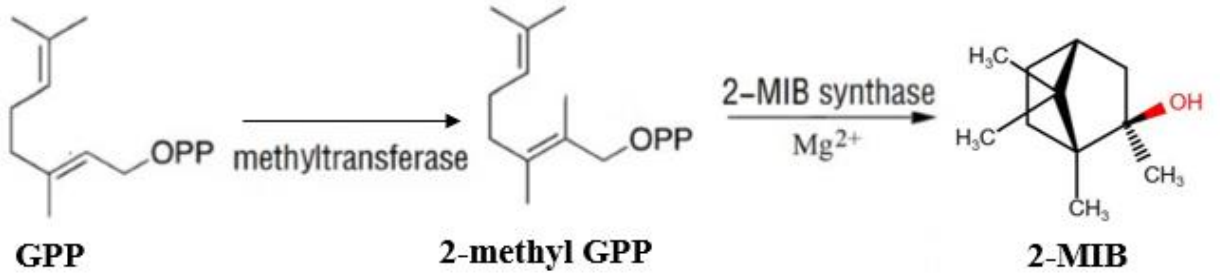


Figure 1.4 Production mechanism of 2-MIB [*Wang et al.*, 2011].

1.6 Purpose of Biosynthesis

Although 2-MIB and geosmin are well known secondary terpenoid metabolites, their biological function and ecological roles have not been elucidated. In general, secondary metabolites are not directly involved in the growth, development or reproduction of the growing bacteria, but they impact microbial community dynamics and interactions [*Tyc et al.*, 2017]. Volatile organic compounds, a subcategory of secondary metabolites that includes 2-MIB and

geosmin, are small molecules that can have a wide range of influence since they can easily travel long distances via aqueous and atmospheric diffusion [*Tyc et al.*, 2017; *Wheatley*, 2002]. They can act in two ways affecting microbial interactions, either as infochemicals contributing to communication and signaling or antimicrobial agents which serve as defensive mechanisms [*Tyc et al.*, 2017; *Watson et al.*, 2003; *Wheatley*, 2002]. Secondary metabolite production in *Streptomyces* for example, is regarded as a response to stress factors that can ensure survivability of the next generation of geminating spores through antibacterial activity [*Asquith et al.*, 2013]. In addition to their possible role as biological weapons in microbial competing interactions, some compounds such as geosmin can serve as a communication mechanism between them when facing stress [*Liato and Aïder*, 2017]. The fact that a freshwater environment is a dynamic system and microbial interactions are responsive to environmental factors, along with the possibility of biased result interpretation under laboratory conditions, makes the clarification of the biosynthetic purpose of 2-MIB and geosmin harder.

1.7 Environmental Drivers

Over the past several decades, anthropogenic activities including urbanization, and intensified industrial and agricultural practices have caused eutrophication in many freshwater environments [*Huisman et al.*, 2018; *Paerl and Paul*, 2012]. This nutrient enrichment favors the proliferation of more frequent and severe bacterial blooms which are linked to the production of taste and odor compounds [*Srinivasan and Sorial*, 2011]. When increased amounts of phosphorus (P) and nitrogen (N) are present in a water body, cyanobacteria tend to dominate in the microbial community and outcompete other microorganisms [*O'Neil et al.*, 2012]. One reason that those nutrients provide a competing advantage is the fact that cyanobacteria can utilize atmospheric nitrogen (N_2) as a nitrogen source via fixation in order to support their growth [*O'Neil et al.*, 2012; *Paerl and Paul*, 2012]. Some of the N_2-fixing genera that have also been confirmed as geosmin and/or 2-MIB producers include *Anabaena*, *Aphanizomenon*, *Lyngbya* and *Oscillatoria* [*Giglio et al.*, 2008; *Giglio et al.*, 2010; *Paerl and Paul*, 2012; *Zhang et al.*, 2009]. In general, cyanobacteria can utilize various forms of dissolved and particulate organic N and P, with iron (Fe) also having an important role as a micronutrient since it is required for the enzymatic activity of nitrogenase [*O'Neil et al.*, 2012]. Several studies have indicated that in freshwater environments, dominance of cyanobacteria and T&O concentrations increased when TN:TP ratios are relatively low and

nitrogen availability is limited relative to phosphorus [*Harris et al.*, 2016; *Lee et al.,* 2017; *Zhang et al.*, 2016]. This is also the case for Eagle Creek Reservoir, where elevated abundances of cyanobacteria and actinobacteria have been observed when TN:TP ratios are low. However, it is important to note that cell death and decay can result in a large release of N and P, and become a source of internal nutrient loading, obscuring the relationships between nutrient input and microbial abundance [*Ma et al.*, 2013].

Climate change is another factor that can promote conditions that favor the formation of blooms, including increased water temperature, reduced water column mixing, severe storms and a short ice-covered season [*Huisman et al.,* 2018; *Winter et al.,* 2011]. Cyanobacteria are widely tolerant, but cold temperatures tend to inhibit their metabolism as they have higher temperature requirements for their growth [*Lee et al.,* 2017; *Winter et al.,* 2011]. *Zhang et al.* [2016] demonstrated in their study that the cyanobacterial genera of *Anabaena* and *Aphanizomenon* were most abundant in Lake Chaohu (China) at a temperature range of 13–30°C and 13–25°C respectively. Furthermore, the species of *Lyngbya kuetzingii* has been confirmed to release more intracellular geosmin into the medium at higher temperatures and as its growth progresses [*Zhang et al.,* 2009]. In many cases however, different strains of the same species have varying optimal T&O production temperatures and may not be a good representation of the entire freshwater system as they are isolates studied under laboratory conditions [*Anuar et al.,* 2017].

Shorter periods of ice cover and warming of the water surface leads to stratification, another condition that favors the proliferation of cyanobacteria [*Huisman et al.,* 2018]. Many planktonic genera can regulate their buoyancy under conditions of reduced water column mixing, and either float to the surface where they have better access to light or migrate deeper to find nutrient concentrated waters [*Huisman et al.,* 2018; *O'Neil et al.,* 2012; *Paerl and Paul,* 2012; *Winter et al.,* 2011]. In addition, climate change has contributed to more frequent and extreme storms, resulting in increased amounts of rainfall and the mobilization of nutrients through terrestrial runoff, that accumulate faster in freshwater systems [*Huisman et al.,* 2018; *Paerl and Paul,* 2012]. Although some of the parameters described above might impact to a greater extent the proliferation of cyanobacterial blooms, there is not a single factor triggering the onset of T&O events, as they all act in synergy and create a dynamic and mobile system.

1.8 Detection Methods

Current approaches for direct detection of odorants in water are mainly based on gas chromatography/mass spectrometry (GC/MS), a sensitive analytical method that can measure metabolites at levels as low as parts per trillion (ppt) [*Liato and Aïder,* 2017; *Tsao et al.,* 2014; *Wang et al.,* 2016]. Typically, GC/MS is coupled with enrichment and extraction techniques, solid phase microextraction (SPME) which has been widely used in the food/beverage industry and utilizes a fused silica fiber for extraction of the contaminants from the sample headspace [*Srinivasan and Sorial,* 2011; *Wang et al.,* 2016; *Watson et al.,* 2000]. Although this method is sensitive, it has several drawbacks including the technical expertise required by the operator, the expensive and largely immobile instrumentation that cannot be used for on-site monitoring, and most importantly, the fact that it does not identify the source of the problem [*John et al.,* 2018; *Su et al.,* 2013]. In order to investigate the origin of T&O episodes, microscopy and traditional plate count techniques are employed for identification and enumeration of odor-producing bacteria. However, those methods are also problematic, as most bacteria in nature are difficult to culture and require selective media [*Jüttner and Watson,* 2007; *Nielsen et al.,* 2006]. Microorganisms such as Actinomycetes which fragment when plated, are challenging to enumerate and other bacteria that are more visible such as cyanobacteria, require taxonomical expertise for their classification since they often lack distinct morphological characteristics [*Jüttner and Watson,* 2007; *Su et al.,* 2013]. Furthermore, microscopic techniques are time consuming and do not distinguish between T&O producers and non-producers [*Jüttner and Watson,* 2007; *Su et al.,* 2013]. Culture independent techniques for identification of T&O producing bacteria such as fluorescence in situ hybridization (FISH) have been successfully used, demonstrating higher sensitivity for quantification of active microorganisms but lacking the ability to distinguish between bacteria that can and those that can not synthesize geosmin and 2-MIB [*Auffret et al.,* 2011; *Nielsen et al.,* 2006].

Perhaps a more comprehensive approach for investigation of the microbial T&O producers is the use of quantitative real-time polymerase chain reaction (qPCR), which contains an inherent ability to quantitate specific genes or genetic loci of interest. In this method, PCR-based DNA amplification of the target is monitored after each cycle by measuring a fluorescence signal which is proportional to the starting quantity of the target gene [*Antonella and Luca,* 2013]. One of the major advantages using this molecular-based assay is its unique identification, as primers and

probes are designed to specifically bind to the gene of interest. In addition, it can detect very low cell numbers, making it a highly sensitive method that can also be used for rapid on-site monitoring of emerging T&O episodes in a relatively inexpensive way [*John et al.,* 2018; *Lyllofff et al.,* 2012]. In the last few years, qPCR assays have been successfully used for the quantification of genes related to toxin production in freshwater systems [*Rasmussen et al*., 2008; *Marbun et al.,* 2012; *Tsao et al.,* 2014]. Similar approaches have also been applied for the detection and quantification of odor producing microorganisms based on genes responsible for geosmin and 2-MIB synthesis. *Giglio et al.* [2008], first reported a qPCR assay to evaluate the presence of geosmin synthase genes in cyanobacterial strains. Screening of seventeen cyanobacterial isolates, coupled with GC-MS analysis, revealed the correlation between the *geoA* gene presence and geosmin production. Thereafter, several research groups have demonstrated the positive correlation between the *geoA* and 2-MIB synthesis genes and their respective concentrations [*Chiu et al.,* 2016; *Su et al.,* 2013; *Wang et al.,* 2016]. Those results indicate, that with further optimization, qPCR assays could serve as a valuable tool for the monitoring and early prediction of emerging T&O problems in freshwater systems.

1.9 Taste and Odor Control

Since both geosmin and 2-MIB have low detection odor threshold by human senses, their control and treatment by water utilities is necessary. Because of their stability and resistance to oxidation, conventional water treatment methods, which usually include the steps of coagulation-sedimentation-filtration-chlorination, do not effectively eliminate them [*Jüttner and Watson,* 2007; *Lee et al.,* 2017; *Lindholm-Lehto and Vielma,* 2018]. Chlorine (Cl_2), which is widely used as an oxidant and disinfectant, does not destroy or remove geosmin and 2-MIB, but merely masks the musty/earthy odor in water [*Liato and Aïder,* 2017; *Nerenberg et al.,* 2000]. Application of ozone (O_3), a stronger oxidant, has greater success in eliminating 2-MIB and geosmin but it can also react with natural organic substances to produce by-products such as low molecular weight ketones and aldehydes, tainting the water with a fruity odor [*Lee et al.,* 2017; *Nerenberg et al.,* 2000]. Currently the most widely used technology for the treatment of odors in water is absorption by powdered activated carbon (PAC), despite the fact that its effectiveness for geosmin and 2-MIB is lower than that for other T&O producing organics and it can have high operational costs [*Liato and Aïder,* 2017; *Nerenberg et al.,* 2000; *Srinivasan and Sorial,* 2011]. Algaecides treatments, in

the form of copper-based chemicals, also control the growth of bloom forming microorganisms by inhibiting both respiration and photosynthesis in algae [*Boyd and Massaut,* 1999]. This is the approach adopted by the Citizens Energy Group, for the control of T&O episodes in ECR, albeit requiring multiple treatments during the same season as episodes become more frequent and severe [*Pascual and Tedesco,* 2006]. However, algaecide treatments should be used with caution, not only because they are a high expense for water utilities, but also because disruption of the cells might release cell-bound material and magnify the problem [*Jüttner and Watson,* 2007; *Lee et al.,* 2017].

1.10 Scope of this Study

Taste and Odor outbreaks pose a significant issue for water utilities, as they tend to lose consumer trust who consider water to be unsafe to drink. Most T&O episodes are not anticipated or linked to their biological origin, making their treatment complicated and not always effective. The research presented in this document, which is fully supported by Citizens Water Indianapolis, the water division of Citizens Energy Group, is an attempt to understand geosmin and 2-MIB microbial production dynamics in Eagle Creek Reservoir, that experiences frequent T&O problems. In the following chapters, we will first investigate and characterize the changing microbial population relative to geosmin and 2-MIB concentrations, with the help of next generation DNA sequencing technology. Then, based on the development of qPCR assays, targeting the synthesis genes of those compounds, we will elucidate the correlation between the genes and the respective concentrations and observe the spatiotemporal patterns of their distribution. Our overall goal is to establish a reliable method that could serve as a tool for the early detection of upcoming T&O events and guide algaecide treatments in a more efficient way.

1.11 References

Antonella, P., & Luca, G. (2013). The quantitative real-time PCR applications in the monitoring of marine harmful algal bloom (HAB) species. *Environmental Science and Pollution Research, 20*(10), 6851-6862.

Anuar, N. S. S., Kassim, A. A., Utsumi, M., Iwamoto, K., Goto, M., Shimizu, K., Othman, N., Zakaria, Z., Sugiura, N., & Hara, H. (2017). Characterization of Musty Odor-Producing Actinomycetes from Tropics and Effects of Temperature on the Production of Musty Odor Compounds. *Microbes and environments, 32*(4), 352-357.

Asquith, E. A., Evans, C. A., Geary, P. M., Dunstan, R. H., & Cole, B. (2013). The role of Actinobacteria in taste and odour episodes involving geosmin and 2-methylisoborneol in aquatic environments. *Journal of Water Supply: Research and Technology-Aqua, 62*(7), 452-467.

Auffret, M., Pilote, A., Proulx, É., Proulx, D., Vandenberg, G., & Villemur, R. (2011). Establishment of a real-time PCR method for quantification of geosmin-producing Streptomyces spp. in recirculating aquaculture systems. *Water research, 45*(20), 6753-6762.

Bentley, R., & Meganathan, R. (1981). Geosmin and methylisoborneol biosynthesis in streptomycetes: Evidence for an isoprenoid pathway and its absence in non-differentiating isolates. *FEBS letters, 125*(2), 220-222.

Boyd, C. E., & Massaut, L. (1999). Risks associated with the use of chemicals in pond aquaculture. *Aquacultural Engineering, 20*(2), 113-132.

Cane, D. E., He, X., Kobayashi, S., Ōmura, S., & Ikeda, H. (2006). Geosmin biosynthesis in Streptomyces avermitilis. Molecular cloning, expression, and mechanistic study of the germacradienol/geosmin synthase. *The Journal of antibiotics, 59*(8), 471.

Chapra, S. C., Boehlert, B., Fant, C., Bierman Jr, V. J., Henderson, J., Mills, D., Mas, D. M. L., Rennels, L., Jantarasami, L., Martinich, J., Strzepek, K. M., & Paerl, H. W. (2017). Climate change impacts on harmful algal blooms in US freshwaters: a screening-level assessment. *Environmental science & technology, 51*(16), 8933-8943.

Chiu, Y. T., Yen, H. K., & Lin, T. F. (2016). An alternative method to quantify 2-MIB producing cyanobacteria in drinking water reservoirs: Method development and field applications. *Environmental research*, *151*, 618-627.

Dickschat, J. S., Nawrath, T., Thiel, V., Kunze, B., Müller, R., & Schulz, S. (2007). Biosynthesis of the Off-flavor 2-Methylisoborneol by the Myxobacterium Nannocystis exedens. *Angewandte Chemie International Edition*, *46*(43), 8287-8290.

Gerber, N. N., & Lechevalier, H. A. (1965). Geosmin, an earthy-smelling substance isolated from actinomycetes. *Appl. Environ. Microbiol.*, *13*(6), 935-938.

Giglio, S., Jiang, J., Saint, C. P., Cane, D. E., & Monis, P. T. (2008). Isolation and characterization of the gene associated with geosmin production in cyanobacteria. *Environmental science & technology*, *42*(21), 8027-8032.

Giglio, S., Chou, W. K. W., Ikeda, H., Cane, D. E., & Monis, P. T. (2010). Biosynthesis of 2-methylisoborneol in cyanobacteria. *Environmental science & technology*, *45*(3), 992-998.

Gkelis, S., Papadimitriou, T., Zaoutsos, N., & Leonardos, I. (2014). Anthropogenic and climate-induced change favors toxic cyanobacteria blooms: Evidence from monitoring a highly eutrophic, urban Mediterranean lake. *Harmful Algae*, *39*, 322-333.

Guttman, L., & van Rijn, J. (2009). 2-Methylisoborneol and geosmin uptake by organic sludge derived from a recirculating aquaculture system. *Water research*, *43*(2), 474-480.

Harris, T. D., Smith, V. H., Graham, J. L., Van de Waal, D. B., Tedesco, L. P., & Clercin, N. (2016). Combined effects of nitrogen to phosphorus and nitrate to ammonia ratios on cyanobacterial metabolite concentrations in eutrophic Midwestern USA reservoirs. *Inland Waters*, *6*(2), 199-210.

Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M., & Visser, P. M. (2018). Cyanobacterial blooms. *Nature Reviews Microbiology*, *16*(8), 471.

John, N., Koehler, A. V., Ansell, B. R., Baker, L., Crosbie, N. D., & Jex, A. R. (2018). An improved method for PCR-based detection and routine monitoring of geosmin-producing cyanobacterial blooms. *Water research*, *136*, 34-40.

Jüttner, F., & Watson, S. B. (2007). Biochemical and ecological control of geosmin and 2-methylisoborneol in source waters. *Appl. Environ. Microbiol.*, *73*(14), 4395-4406.

Klausen, C., Nicolaisen, M. H., Strobel, B. W., Warnecke, F., Nielsen, J. L., & Jørgensen, N. O. (2005). Abundance of actinobacteria and production of geosmin and 2-methylisoborneol in Danish streams and fish ponds. *FEMS microbiology ecology*, *52*(2), 265-278.

Komatsu, M., Tsuda, M., Ōmura, S., Oikawa, H., & Ikeda, H. (2008). Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol. *Proceedings of the National Academy of Sciences*, *105*(21), 7422-7427.

Lee, J., Rai, P. K., Jeon, Y. J., Kim, K. H., & Kwon, E. E. (2017). The role of algae and cyanobacteria in the production and release of odorants in water. *Environmental pollution*, *227*, 252-262.

Li, L., Pascual, D. L., Tedesco, L. P., Randolph, K. L., Sengpiel, R. E., Hall, B. E., & Wilson, J. S. (2006). Developing a survey tool for the rapid assessment of blue-green algae in central Indiana's reservoirs. *IUPUI. Indianapolis: Central Indiana Resources Partnership/CEES*, *7*(11).

Liato, V., & Aïder, M. (2017). Geosmin as a source of the earthy-musty smell in fruits, vegetables and water: Origins, impact on foods and water, and review of the removing techniques. *Chemosphere*, *181*, 9-18.

Lindholm-Lehto, P. C., & Vielma, J. (2019). Controlling of geosmin and 2-methylisoborneol induced off-flavours in recirculating aquaculture system farmed fish—A review. *Aquaculture Research*, *50*(1), 9-28.

Lylloff, J. E., Mogensen, M. H., Burford, M. A., Schlüter, L., & Jørgensen, N. O. (2012). Detection of aquatic streptomycetes by quantitative PCR for prediction of taste-and-odour episodes in water reservoirs. *Journal of Water Supply: Research and Technology-Aqua*, *61*(5), 272-282.

Ma, Z., Niu, Y., Xie, P., Chen, J., Tao, M., & Deng, X. (2013). Off-flavor compounds from decaying cyanobacterial blooms of Lake Taihu. *Journal of Environmental Sciences*, *25*(3), 495-501.

Marbun, Y. R., Yen, H-K., Lin, T-F., Lin, H-L., & Michinaka, A. (2012). Rapidon-site monitoring of cylindrospermopsin-producers in reservoirs usingquantitative PCR. Sustain Environ Res22:143–151.

Medsker, L. L., Jenkins, D., Thomas, J. F., & Koch, C. (1969). Odorous compounds in natural waters. 2-Exo-hydroxy-2-methylbornane, the major odorous compound produced by several actinomycetes. *Environmental Science & Technology*, *3*(5), 476-477.

Nerenberg, R., Rittmann, B. E., & Soucie, W. J. (2000). Ozone/biofiltration for removing MIB and geosmin. *Journal-American Water Works Association*, *92*(12), 85-95.

Nielsen, J. L., Klausen, C., Nielsen, P. H., Burford, M., & Jørgensen, N. O. (2006). Detection of activity among uncultured Actinobacteria in a drinking water reservoir. *FEMS microbiology ecology*, *55*(3), 432-438.

O'Neil, J. M., Davis, T. W., Burford, M. A., & Gobler, C. J. (2012). The rise of harmful cyanobacteria blooms: the potential roles of eutrophication and climate change. *Harmful algae*, *14*, 313-334.

Paerl, H. W., & Paul, V. J. (2012). Climate change: links to global expansion of harmful cyanobacteria. *Water research*, *46*(5), 1349-1363.

Pascual, D. L., & Tedesco, L. P. (2006). Eagle Creek Reservoir: Responses to Algaecide Treatment. *Indiana University–Purdue University, Indianapolis, Department of Geology. Indianapolis, IN. 41pp. Available online at:< http://www. cees. iupui. edu/Research/Water_Resources/CIWRP/Publications/Reports/CEES-2004-02_2003-Algaecide-Study. pdf> Accessed*, *12*.

Rasmussen, J. P., Giglio, S., Monis, P. T., Campbell, R. J., & Saint, C. P. (2008). Development and field testing of a real-time PCR assay for cylindrospermopsin-producing cyanobacteria. *Journal of Applied Microbiology*, *104*(5), 1503-1515.

Rodrıguez-Concepción, M., & Boronat, A. (2002). Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant physiology*, *130*(3), 1079-1089.

Song, K., Li, L., Li, S., Tedesco, L., Hall, B., & Li, L. (2012). Hyperspectral remote sensing of total phosphorus (TP) in three central Indiana water supply reservoirs. *Water, Air, & Soil Pollution*, *223*(4), 1481-1502.

Srinivasan, R., & Sorial, G. A. (2011). Treatment of taste and odor causing compounds 2-methyl isoborneol and geosmin in drinking water: A critical review. *Journal of Environmental Sciences*, *23*(1), 1-13.

Su, M., Gaget, V., Giglio, S., Burch, M., An, W., & Yang, M. (2013). Establishment of quantitative PCR methods for the quantification of geosmin-producing potential and Anabaena sp. in freshwater systems. *Water research*, *47*(10), 3444-3454.

Suurnäkki, S., Gomez-Saez, G. V., Rantala-Ylinen, A., Jokela, J., Fewer, D. P., & Sivonen, K. (2015). Identification of geosmin and 2-methylisoborneol in cyanobacteria and molecular detection methods for the producers of these compounds. *Water research*, *68*, 56-66.

Tsao, H. W., Michinaka, A., Yen, H. K., Giglio, S., Hobson, P., Monis, P., & Lin, T. F. (2014). Monitoring of geosmin producing Anabaena circinalis using quantitative PCR. *Water research*, *49*, 416-425.

Tyc, O., Song, C., Dickschat, J. S., Vos, M., & Garbeva, P. (2017). The ecological role of volatile and soluble secondary metabolites produced by soil bacteria. *Trends in microbiology*, *25*(4), 280-292.

Wang, Z., Xu, Y., Shao, J., Wang, J., & Li, R. (2011). Genes associated with 2-methylisoborneol biosynthesis in cyanobacteria: isolation, characterization, and expression in response to light. *PLoS One*, *6*(4), e18665.

Wang, Z., Song, G., Shao, J., Tan, W., Li, Y., & Li, R. (2016). Establishment and field applications of real-time PCR methods for the quantification of potential MIB-producing cyanobacteria in aquatic systems. *Journal of applied phycology*, *28*(1), 325-333.

Watson, S. B., Brownlee, B., Satchwill, T., & Hargesheimer, E. E. (2000). Quantitative analysis of trace levels of geosmin and MIB in source and drinking water using headspace SPME. *Water Research*, *34*(10), 2818-2828.

Watson, S. B. (2003). Cyanobacterial and eukaryotic algal odour compounds: signals or by-products? A review of their biological activity. *Phycologia*, *42*(4), 332-350.

Watson, S. B., Ridal, J., & Boyer, G. L. (2008). Taste and odour and cyanobacterial toxins: impairment, prediction, and management in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, *65*(8), 1779-1796.

Wheatley, R. E. (2002). The consequences of volatile organic compound mediated bacterial and fungal interactions. *Antonie Van Leeuwenhoek*, *81*(1-4), 357-364.

Winter, J. G., DeSellas, A. M., Fletcher, R., Heintsch, L., Morley, A., Nakamoto, L., & Utsumi, K. (2011). Algal blooms in Ontario, Canada: increases in reports since 1994. *Lake and Reservoir Management*, *27*(2), 107-114.

Zaitlin, B., Watson, S. B., Ridal, J., Satchwill, T., & Parkinson, D. (2003). Actinomycetes in Lake Ontario: habitats and geosmin and MIB production. *Journal-American Water Works Association*, *95*(2), 113-118.

Zhang, T., Li, L., Song, L., & Chen, W. (2009). Effects of temperature and light on the growth and geosmin production of Lyngbya kuetzingii (Cyanophyta). *Journal of Applied Phycology*, *21*(3), 279-285.

Zhang, M., Zhang, Y., Yang, Z., Wei, L., Yang, W., Chen, C., & Kong, F. (2016). Spatial and seasonal shifts in bloom-forming cyanobacteria in L ake C haohu: Patterns and driving factors. *Phycological research*, *64*(1), 44-55.

CHAPTER 2. COMPOSITION OF MICROBIAL COMMUNITY ASSOCIATED WITH TASTE AND ODOR EPISODES IN EAGLE CREEK RESERVOIR

2.1 Introduction

Microbial populations in freshwater and marine environments play key roles in global biogeochemical cycles and primary productivity, as they regulate the flux of elements through mobilization of particulate matter derived from terrestrial runoff [*Eiler and Bertilsson,* 2004; *Kennedy et al.,* 2010]. Unfortunately, a big part of their structure and biological functions remains unexplored, since most bacteria lack distinctive morphological characteristics and are hard to isolate using traditional culturing techniques [*Kennedy et al.,* 2010; *Pope and Patel,* 2008; *Zwart et al.,* 2002]. The relatively recent development of DNA sequencing technologies enable us to overcome cultivation difficulties, explore the population dynamics and genetic diversity of microbial communities and characterize novel organisms using taxonomically informative genomic markers [*Sharpton,* 2014]. In the case of archaea and bacteria, the most widely used marker gene is the small subunit rRNA gene (16S rDNA) which is taxonomically and phylogenetically conserved among species [*Sharpton,* 2014; *Zwart et al.,* 2002]. The major limitation of this approach is that only microbes for which phylogenetic information is known can be detected and characterized in a sample. Shotgun metagenomic sequencing is an alternative method that can partly overcome this problem since DNA from all cells is extracted and independently sequenced instead of targeting a specific gene [*Sharpton,* 2014]. In the last few years the number of studies using metagenomic approaches has increased because of the advancements in next-generation sequencing technologies that enable researchers to explore microbial biodiversity and function in a great range of environments, including soil, freshwater and human biological samples [*Gomez-Alvarez et al.,* 2012; *McVeigh et al.,* 1996; *Pope and Patel,* 2008; *Seashols-Williams et al.,* 2018].

Common Bacteria detected by next generation sequencing techniques and metagenomic analysis in freshwater environments include the phyla of Proteobacteria, Cyanobacteria, Actinobacteria, Bacteroidetes and Verrucomicrobia [Eiler and Bertilsson, 2004; Zwart et al., 2002]. Among those, Cyanobacteria play an important role shaping the microbial structure of the system, by producing substances such as organic carbon, oxygen and fixed nitrogen which constitute

highly available substrates for heterotrophic bacteria and affect their interactions [Eiler and Bertilsson, 2004; Huisman et al., 2018]. In addition, their presence in freshwater is of a great concern due to the ability of some species to produce secondary metabolites such as toxins, impacting human and animal health, as well as odorous compounds that deteriorate drinking water quality [Pope and Patel, 2008; Winter et al., 2011]. Although non-toxic, the cyanobacterial blooms occurring in Eagle Creek Reservoir, the major drinking water supply of Indianapolis, are associated with the production of taste and odor compounds. The odorous events are usually observed during the Spring, when the reservoir receives most of its water and is fully mixed and turbid, and at the beginning of Fall. Those problems are mainly caused by the terpenoid bacterial metabolites 2-MIB (2-methylisoborneol) and geosmin (trans-1,10-dimethyl-trans-9-decalol) which produce a mildewed and earthy flavor, respectively [Jüttner and Watson, 2007; Wang et al., 2016]. Each compound exists as (+) and (-) enantiomers, but odor outbreaks are caused by the biological production of the (-) stereoisomer which is 10 times more potent than the (+) molecules [Jüttner and Watson, 2007]. Both exhibit strong resistance to oxidation, which is a process routinely applied in water purification, and along with their low detection threshold by human senses (2-10 ng/L), they tend to be the most persistent substances in odor incidents [Auffret et al., 2011; Suurnäkki et al., 2015]. A wide variety of microorganisms can synthesize and secrete MIB and geosmin, but cyanobacteria and actinobacteria species are regarded as the main producers in freshwater ecosystems [Auffret et al., 2011; Wang et al., 2016]. The production mechanism of the terpenoid metabolites in cyanobacteria and actinomycetes is similar, with homologous genes involved in the synthesis [Giglio et al., 2010; Wang et al., 2016]. Geosmin is synthesized through the conversion of farnesyl diphosphate to geosmin catalyzed by geosmin synthase encoded by the *geoA* gene [Giglio et al., 2010; Suurnäkki et al., 2015]. MIB synthesis consists of 2 key reactions: 1) methylation of geranyl diphosphate (GPP) to 2-methyl-GPP, and, 2) further cyclization of 2-methyl-GPP to MIB catalyzed by MIB synthase [Giglio et al., 2010].

Since taste and odor episodes tend to become frequent and sometimes severe in ECR, it is necessary to trace the origin of the problem in order to find effective management solutions. To better understand microbial population dynamics in the system, this study aimed to answer the main questions of “who is there?” and “what are they doing?”. Using next generation sequencing techniques, targeting the 16S rRNA gene, the purpose of the present work was to characterize bacterial community composition, in relation to geosmin and 2-MIB concentrations, and evaluate

its response to the algaecide treatment applied in the reservoir. Furthermore, we attempted to retrieve the genes responsible for the synthesis of geosmin and 2-MIB, using a previously analyzed metagenomic dataset [*Clercin,* 2018], in which the *geoA* and *MIB synthase* genes were not successfully recovered.

2.2 Materials and Methods

2.2.1 Study site

Eagle Creek Reservoir (ECR;86°18′13.07″W, 39°51′09.84″N; A = 5.0 km2; Z = 4.2 m), was constructed in 1967 in the northwest corner of Marion County to initially provide flood control and then drinking water for the city of Indianapolis when T.W. Moses Drinking Water Plant became available in 1976 [*Harris et al.,* 2016; *Li et al.,* 2006]. The depth ranges from 4 to 13 meters, with the deepest areas located in the southern basin near the dam. The reservoir is classified as being mesotrophic to eutrophic [*Song et al.,* 2012], and its water column is mixed and turbid in April/May when it receives most of its water from snow melt and intense rainfall and then in October after the seasonal thermal stratification is over.

2.2.2 Sample collection and processing

For the metagenomic analysis, water samples were collected in May, July and October 2013 near the dam where the strongest water column stratifications occur (Figure 2.1(A)). Discrete water samples were collected with a vertical Van Dorn sampler at four different depths corresponding to sub-surface (0 m), epilimnion (3 m), metalimnion (6 m) and hypolimnion (9-10 m), i.e. 1 meter above the water-sediment interface. A total of 11 samples was collected as the sub-surface sample from October did not recover enough genetic materials to be processed. After collection near the dam all the samples were put on ice in autoclaved 1-L HDPE brown bottles and filtered in the lab through 0.22 µm mesh size pores on a sterile glass filtration unit, then frozen for storage in 15-mL Falcon tubes.

For the 16S rRNA gene sequencing analysis, water samples were collected with a vertical Van Dorn sampler at two locations, near the dam and north of intake (Figure 2.1(B)). After that, 50ml of water were filtered through a Sterivex filter unit while being on the boat and 3ml of RNA*later* stabilization solution was used for preservation purposes. Filters were stored in a -80 $^{\circ}C$ freezer before further processing.

A)

B)

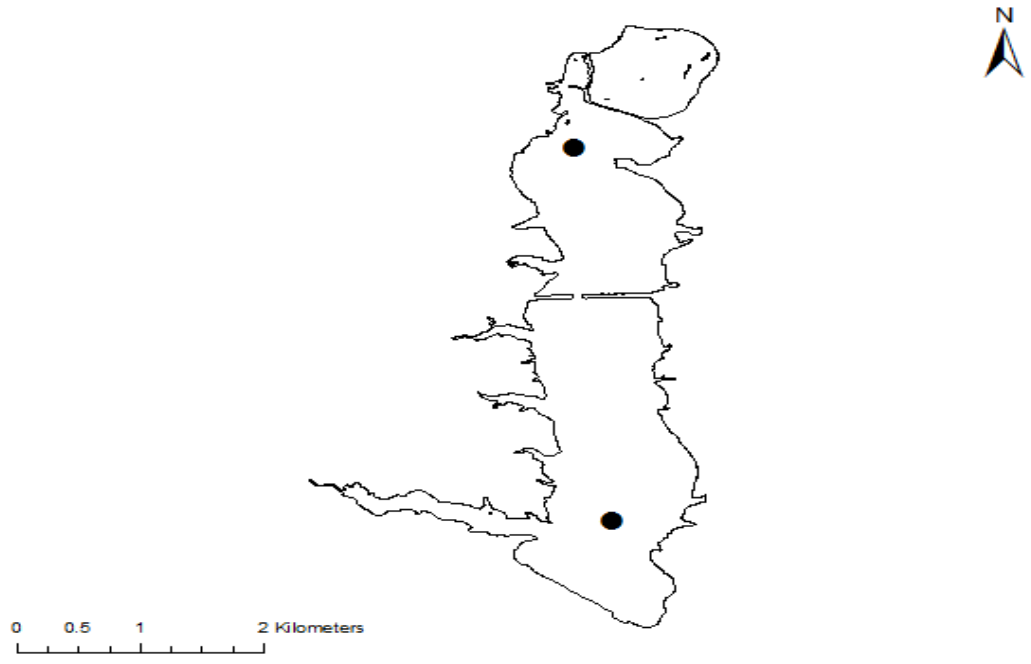


Figure 2.1 Eagle Creek Reservoir sampling locations. A) 2013 Sampling location used for metagenomic analysis; B) 2017 Sampling locations used for 16S sequencing analysis.

2.2.3 DNA extraction

The genomic DNA of the water samples was extracted using the DNeasy PowerWater Sterivex Kit (Qiagen). Lysis buffer was added to the Sterivex filter units, which were then mixed. Subsequently, the lysate was removed for additional lysis using a 5 ml bead beating tube. After the removal of proteins and inhibitors, total genomic DNA was captured on an MB Spin Column. The column was washed, and purified DNA was eluted in 100 μL of EB solution (sterile elution buffer) and then stored at −20 °C. The concentration and purity of the DNA was measured with a NanoDrop 2000/2000c Spectrophotometer (Thermo Scientific Inc.).

2.2.4 Sequencing technologies

The samples collected during the 2013 sampling season were shipped to Illumina, Inc., San Diego, CA for analysis on frozen filters to determine the phylogenetic structure of the BCC by next-generation sequencing (NGS) Shotgun Metagenomics method [*Clercin,* 2018]. Sequencing-ready libraries were prepared using the Prep Kits for MiSeq v.3 Nextera XT, Illumina. The samples collected during the 2017 sampling season were shipped to MR DNA lab, Shallowater, TX for 16S sequencing analysis. Primers targeted the 16S rRNA gene V4 variable region and along with HotStarTaq Plus Master Mix Kit (Qiagen, USA) they were used in a single-step 30 cycle PCR under the following conditions: 94°C for 3 minutes, followed by 30 cycles (5 cycle used on PCR products) of 94°C for 30 seconds, 53°C for 40 seconds and 72°C for 1 minute, after which a final elongation step at 72°C for 5 minutes was performed. The subsequent sequencing was performed on an Ion Torrent PGM system.

2.2.5 Sequence processing

The sequence analysis for both datasets (2013 & 2017) was performed using the CLC Genomics Workbench software (Qiagen Bioinformatics). Reference sequences were retrieved from the NCBI database and their similarity was assessed by the Clustal Omega-Multiple Sequence Alignment program [*McWilliam et al.,* 2013]. Read mapping for the metagenomic dataset was performed using CLC's default parameters except from a length fraction of 0.5 and a similarity fraction of 0.6. Raw mapping results were converted to percentages and then rescaled on a 0 - 1 range both on a per gene/reference sequence basis and a "global" basis. Heat Maps for the scaled percentage of mapped reads were generated using the R software for statistical and

graphical analysis. Sequence identity was confirmed by generating a consensus sequence for the sample with the highest number of mapped reads for each reference sequence and then doing a BLASTn search on the NCBI (National Center for Biotechnology Information) sequence database. The first five sequences producing significant alignments were selected to be listed in this report. Reference based OTU clustering was performed using CLC's default parameters and the Greengenes database [*DeSantis et al.*, 2006] as a reference with a similarity score of 97%. Alpha diversity was estimated based on phylogenetic diversity distance calculations. Estimation of beta diversity was based on the *Bray-Curtis* equation.

2.2.6 T&O analysis

2-Methylisoborneol and geosmin concentrations were quantified by a Head-Space Solid-Phase Micro-Extraction (HS-SPME) combined with a Gas Chromatography-Mass Spectrometry (GC-MS) at the Citizens Energy lab.

2.3 Results

2.3.1 Presence of T&O synthesis genes in the metagenomic dataset

In order to investigate whether the bacterial population in ECR has the potential of producing geosmin and 2-MIB, and recover the genes responsible for the synthesis of those compounds, read mapping was performed on the 2013 metagenomic samples by aligning their reads to known *geoA* and *MIBS* containing reference sequences retrieved from the NCBI database. Cyanobacterial reference sequences had good alignment and high similarity scores between them, ranging from 81.33 to 99.77% for the *geoA* gene and from 84.37 to 91.42% for the *MIBS* gene (Table 2.1). This enabled us to trim those reference sequences so that they can have the exact same length. In contrast, the actinobacterial reference sequences for the *geoA* gene had low similarity scores between them (Table 2.1) resulting in a poor alignment that did not allow for a length adjustment and implying that the gene is highly variable among different actinobacterial species. Only one *MIBS* Actinobacterial reference sequence was used due to limited information on the database.

Relative to each sample's total number of reads, raw results for both *geoA* (Table 2.2) and *MIBS* (Table 2.3) genes show a low number of mapped reads to the reference sequences. Generally,

regarding both genes, more reads were successfully mapped to Actinobacterial compared to Cyanobacterial reference sequences. Percentages of the mapped reads were calculated, and the range was adjusted on a scale of 0 to 1 both on a per gene/reference sequence basis and a "global" basis. Heat maps (Figure 2.2 A & B), where comparisons for the same reference sequence across different samples can only be made, indicate that for each reference sequence, regarding both the *geoA* and *MIBS* genes, the May-10m sample has the highest level of mapped reads. The "global" scaled heat map for the *geoA* gene (Figure 2.3(A)), shows that mapping of the May-10m sample to *Streptomyces fradiae* has the highest percentage of mapped reads across all reference sequences and samples. In addition, more reads mapped to the *geoA* sequences coming from Actinobacterial species compared to Cyanobacterial, and higher levels of mapped reads were detected in May. Looking at the "global" scaled heat map regarding the *MIBS* gene (Figure 2.3 (B)), the May-10m sample mapped to *Streptomyces lividans* generated the highest percentage of mapped reads. Once again, the May samples had the highest number of mapped reads among the three sampling months and *Streptomyces lividans* used as a reference sequence resulted in more mapped reads compared to Cyanobacterial reference sequences.

Sequence identity was confirmed by generating a consensus sequence for the sample with the highest number of mapped reads for each reference sequence and then doing a BLASTn search on the NCBI sequence database. The expected value (E) is 0.0 (Table 2.4, Table 2.5, Table 2.6, Table 2.7) for all of the BLASTn results, indicating the presence of a "significant" match. Consensus sequences derived from sample mappings to Actinobacterial reference sequences, when used on the BLASTn search, generated significant alignments with species belonging to the same genus as the reference sequence on the mapping analysis. In contrast, consensus sequences derived from sample mappings to Cyanobacterial reference sequences, when used on the BLASTn search, generated significant alignments with species not necessarily belonging to the same genus as the reference sequence on the mapping analysis. No significant similarity was found on the BLASTn search for the consensus sequence derived from the mapping of the May-6m sample to *Mycobacterium talmoniae*.

Table 2.1 Read mapping reference sequences A) Cyanobacterial reference sequences-*geoA* gene. B) Actinobacterial reference sequences-*geoA* gene. C) Cyanobacterial reference sequences-*MIBS* gene. D) Actinobacterial reference sequence – *MIBS* gene.

A)

		Percent Identity Matrix			
Reference Sequences (geoA-Cyanobacteria)	**Length (bp)**	**KJ658367.1**	**KJ658368.1**	**KJ658376.1**	**KJ658375.1**
KJ658368.1 Nostoc sp. ATCC 53789	884	81.45	**100.00**	82.13	81.9
KJ658376.1 Planktothrix sp. 18	884	81.56	82.13	**100.00**	99.77
KJ658375.1 Oscillatoria sp. PCC 9240	884	81.33	81.9	99.77	**100.00**
KJ658367.1 Aphanizomenon sp. PMC9501	884	**100.00**	81.45	81.56	81.33

B)

		Percent Identity Matrix		
Reference Sequences (geoA-Actinobacteria)	**Length (bp)**	**JX966093.1**	**PPEA01000211.1**	**CP009111.1**
JX966093.1 Streptomyces fradiae strain HX	1585	**100.00**	45.50	47.66
PPEA01000211.1:c7846-7397 Mycobacterium talmoniae strain ATCC BAA-2683 C1Y40_211	450	45.50	**100.00**	68.24
CP009111.1:c1186434-1185346 Rhodococcus opacus strain 1CP	1089	47.66	68.24	**100.00**

C)

		Percent Identity Matrix		
Reference Sequences (MIBS-Cyanobacteria)	**Adjusted Length (bp)**	**HQ630883.1**	**HQ630885.1**	**KJ658378.1**
HQ630883.1 Pseudanabaena limnetica str. Castaic Lake	723	**100.00**	88.66	84.37
HQ630885.1 Oscillatoria limosa	723	88.66	**100**	91.42
KJ658378.1 Planktothrix sp. 328	723	84.37	91.42	**100**

D)

Reference Sequences (*MIBS*-Actinobacteria)	**Length (bp)**
CP009124.1:c157137-155815 *Streptomyces lividans* TK24	1323

Table 2.2 Read mapping raw results – *geoA* gene.

		Reference Sequences (geoA)						
		Cyanobacteria				Actinobacteria		
		Nostoc sp. ATCC 53789	**Planktothrix sp. 18**	**Oscillatoria sp. PCC 9240**	**Aphanizomenon sp. PMC9501**	**Streptomyces fradiae**	**Mycobacterium talmoniae**	**Rhodococcus opacus**
Sample	**Total Reads**	**Mapped Reads**						
May-0m	33,395,202	42	46	46	46	621	102	459
May-3m	37,789,429	70	66	63	63	669	196	555
May-6m	57,196,837	118	122	125	125	1,642	425	1,083
May-10m	31,003,711	109	75	82	82	1,489	304	943
July-0m	32,673,383	14	8	8	8	230	41	178
July-3m	85,796,130	45	19	21	21	463	113	446
July-6m	52,993,984	46	20	24	24	847	138	538
July-10m	19,306,387	34	35	35	35	386	106	287
October-3m	55,503,302	37	25	27	27	738	131	448
October-6m	69,814,786	41	32	36	36	960	192	587
October-10m	46,838,779	14	5	5	5	297	43	185

Table 2.3 Read mapping raw results - *MIBS* gene.

		Reference Sequences (MIBS)			
		Cyanobacteria			Actinobacteria
		Oscillatoria limosa	***Pseudanabaena limnetica***	***Planktothrix sp. 328***	***Streptomyces lividans***
Sample	**Total Reads**	**Mapped Reads**			
May-0m	33,395,202	120	40	130	357
May-3m	37,789,429	115	85	142	455
May-6m	57,196,837	285	150	272	1,046
May-10m	31,003,711	342	118	296	895
July-0m	32,673,383	31	13	47	149
July-3m	85,796,130	84	26	87	307
July-6m	52,993,984	135	28	119	461
July-10m	19,306,387	77	47	70	222
October-3m	55,503,302	144	44	101	393
October-6m	69,814,786	197	54	161	528
October-10m	46,838,779	48	16	40	173

A)

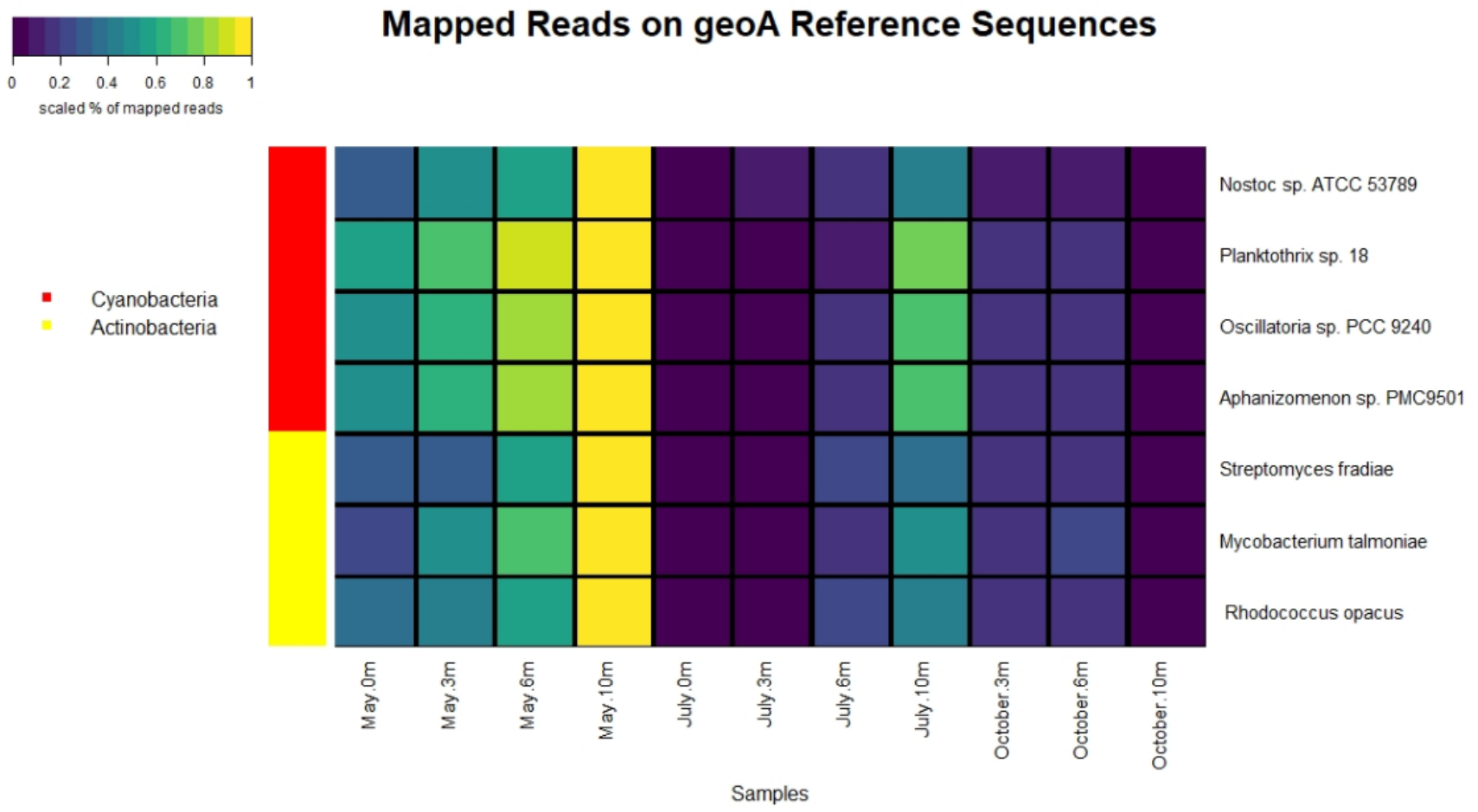
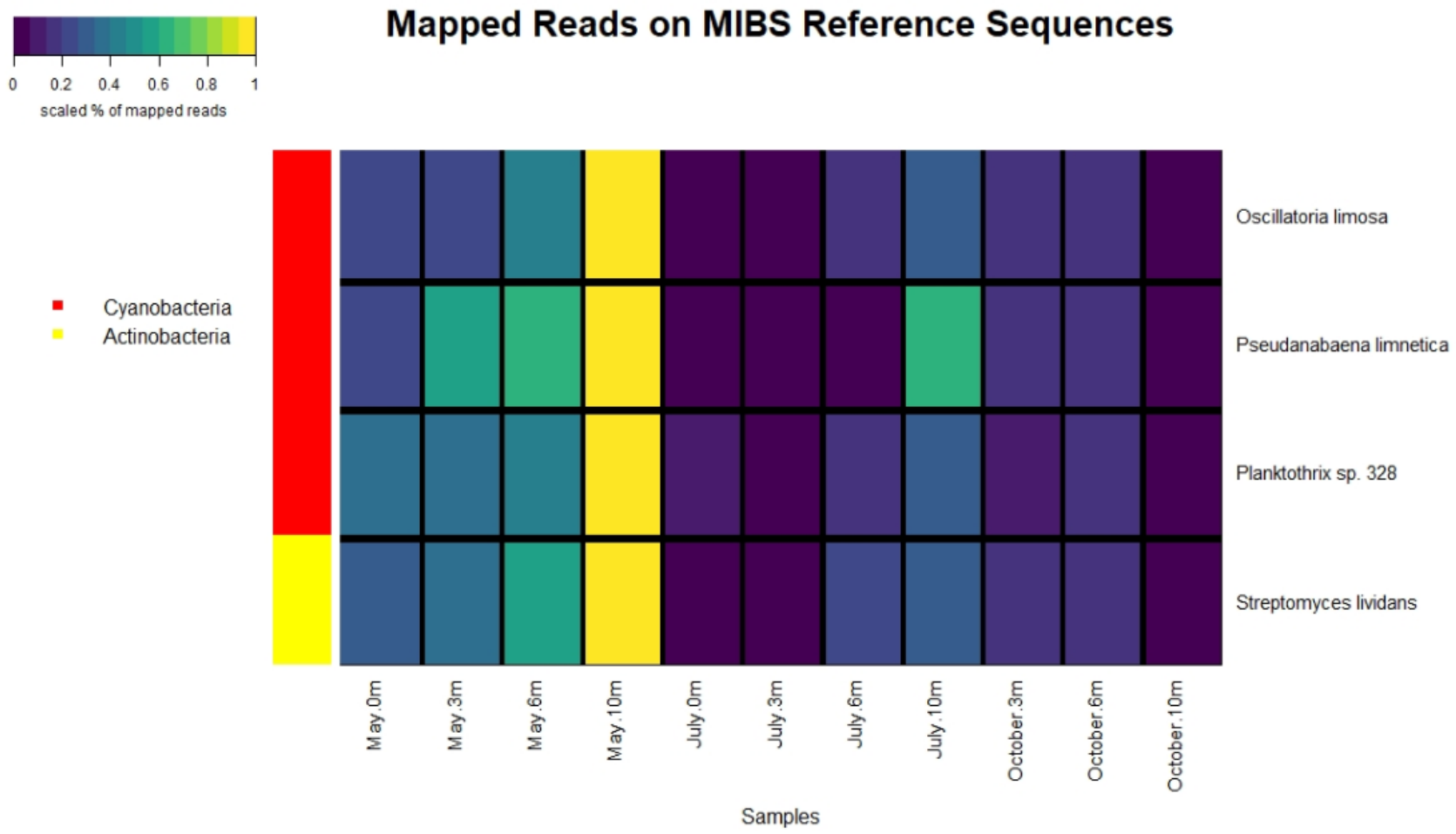


Figure 2.2 Heat Maps representing the percentage of mapped reads scaled on a per gene/reference basis. A) *geoA.* B) *MIBS.*

Figure 2.2 continued

B)



A)

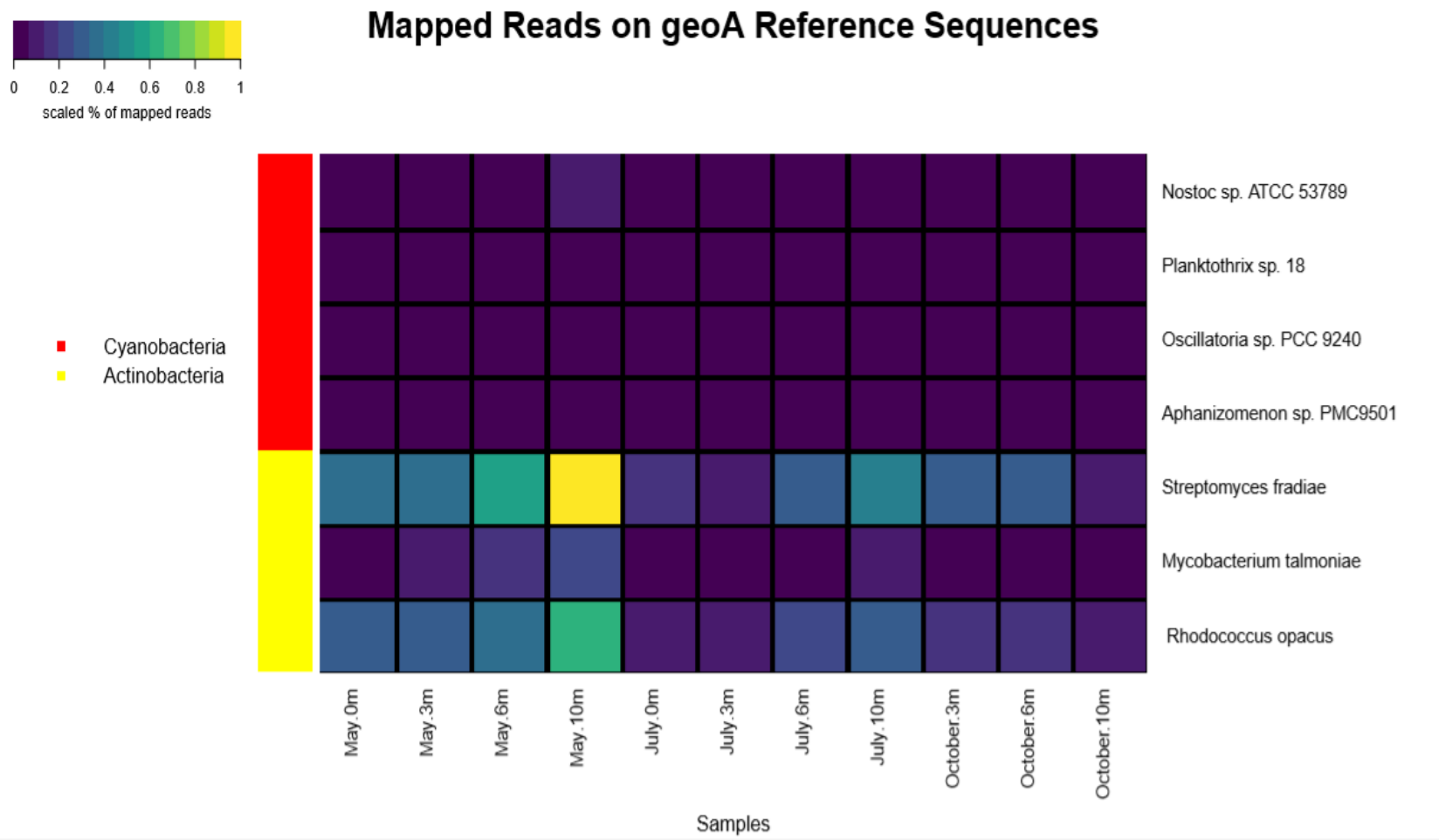


Figure 2.3 Heat Maps representing the percentage of mapped reads scaled on a "global" basis. A) *geo.* B) *MIBS.*

Figure 2.3 continued

B)

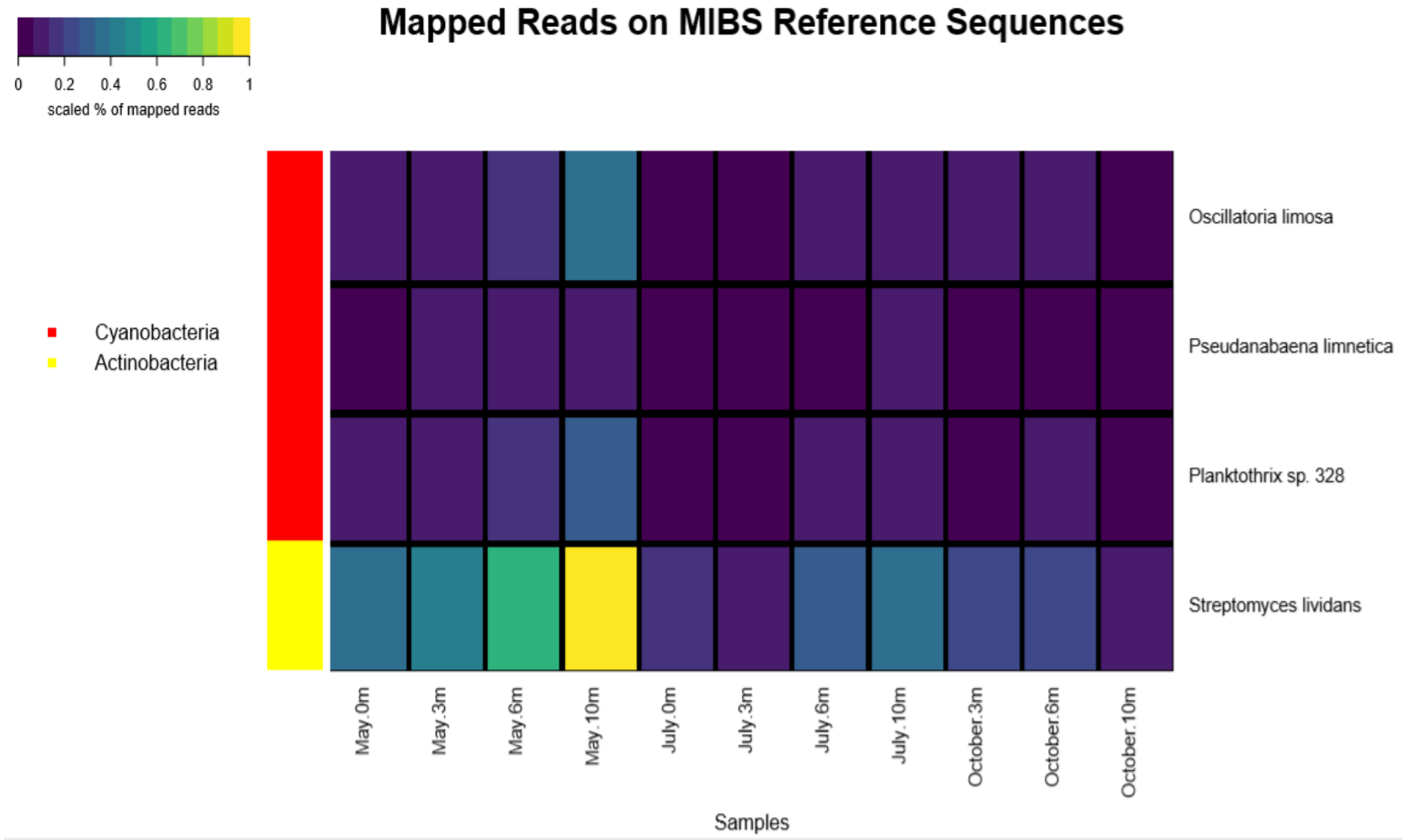


Table 2.4 BLASTn results – *geoA* Cyanobacteria.

Consensus sequence	Sequences Producing Significant Alignments	Max Score	Total Score	Query Cover	E value	Per. Ident.	Accession
May 6m -*Nostoc sp.* ATCC 53789	*Nostoc sp*. ATCC 53789 geosmin synthase (geoA) gene, partial cds	1594	1594	100%	0.0	99.21%	KJ658368.1
	Nostoc punctiforme PCC 73102 NPUNMOD protein gene, complete cds	1417	1417	100%	0.0	95.59%	FJ010203.1
	Nostoc punctiforme PCC 73102 NJ2 protein gene, partial cds	1417	1417	100%	0.0	95.59%	FJ010202.2
	Nostoc punctiforme PCC 73102, complete genome	1417	1417	100%	0.0	95.59%	CP001037.1
	Nostoc sp. UK1 geosmin synthase (geoA) gene, partial cds	1170	1170	99%	0.0	90.61%	KJ658372.1
May 6m- *Planktothrix sp*. 18	*Planktothrix sp*. 328 geosmin synthase (geoA) gene, partial cds	1628	1628	100%	0.0	99.89%	KJ658374.1
	Oscillatoria sp. PCC 9240 geosmin synthase (geoA) gene, partial cds	1616	1616	100%	0.0	99.66%	KJ658375.1
	Calothrix sp. PCC 7507, complete genome	850	850	99%	0.0	84.03%	CP003943.1
	Nostoc punctiforme PCC 73102 NPUNMOD protein gene, complete cds	776	776	99%	0.0	82.56%	FJ010202.1
	Nostoc punctiforme PCC 73102 NJ2 protein gene, partial cds	776	776	99%	0.0	82.56%	FJ010203.2
May 6m- *Oscillatoria sp*. PCC 9240	*Oscillatoria sp*. PCC 9240 geosmin synthase (geoA) gene, partial cds	1628	1628	100%	0.0	99.89%	KJ658375.1
	Planktothrix sp. 328 geosmin synthase (geoA) gene, partial cds	1616	1616	100%	0.0	99.66%	KJ658374.1
	Calothrix sp. PCC 7507, complete genome	839	839	99%	0.0	83.81%	CP003943.1
	Nostoc punctiforme PCC 73102 NPUNMOD protein gene, complete cds	767	767	94%	0.0	83.19%	FJ010203.1
	Nostoc punctiforme PCC 73102 NJ2 protein gene, partial cds	767	767	94%	0.0	83.19%	FJ010202.1
May 6m- Aphanizomenon sp. PMC9501	*Oscillatoria sp*. PCC 9240 geosmin synthase (geoA) gene, partial cds	1628	1628	100%	0.0	99.89%	KJ658375.1
	Planktothrix sp. 328 geosmin synthase (geoA) gene, partial cds	1616	1616	100%	0.0	99.66%	KJ658374.1
	Calothrix sp. PCC 7507, complete genome	839	839	99%	0.0	83.81%	CP003943.1
	Nostoc punctiforme PCC 73102 NPUNMOD protein gene, complete cds	767	767	94%	0.0	83.19%	FJ010203.1
	Nostoc punctiforme PCC 73102 NJ2 protein gene, partial cds	767	767	94%	0.0	83.19%	FJ010202.1

Table 2.5 BLASTn results – *geoA* Actinobacteria.

Consensus sequence	**Sequences Producing Significant Alignments**	**Max Score**	**Total Score**	**Query Cover**	**E value**	**Per. Ident.**	**Accession**
May 6m-*Streptomyces fradiae*	*Streptomyces fradiae* strain HX putative germacradienol synthase (geoA) gene	2822	2822	100%	0.0	98.80%	JX966093.1
	Streptomyces sp. KPB2 chromosome, complete genome	2473	2473	100%	0.0	94.83%	CP034353.1
	Streptomyces sp. CCM_MD2014 chromosome, complete genome	2473	2473	100%	0.0	94.83%	CP009754.1
	Streptomyces sp. CB09001 chromosome, complete genome	2362	2362	100%	0.0	93.58%	CP026730.1
	Streptomyces sp. 2114.2 genome assembly, chromosome: I	2357	2357	100%	0.0	93.52%	LT629768.1
May 6m-*Rhodococcus opacus*	*Rhodococcus opacus* strain 1CP, complete genome	1943	1943	100%	0.0	98.90%	CP009111.1
	Rhodococcus opacus PD630, complete genome	1760	1760	100%	0.0	95.87%	CP003949.1
	Rhodococcus jostii RHA1, complete genome	1572	1572	100%	0.0	92.75%	CP000431.1
	Rhodococcus opacus strain R7 sequence	1539	1539	100%	0.0	92.20%	CP008947.1
	Rhodococcus opacus B4 DNA, complete genome	1356	1356	100%	0.0	89.17%	AP011115.1

Table 2.6 BLASTn results – *MIBS* Cyanobacteria.

Consensus sequence	Sequences Producing Significant Alignments	Max Score	Total Score	Query Cover	E value	Per. Ident.	Accession
May 10m-Oscillatoria limosa	*Oscillatoria limosa* LBD 305b MIB synthase gene, partial cds	1260	1260	100%	0.0	98.20%	HQ630885.1
	Planktothricoides raciborskii gene, MIB synthase, partial sequence, strain: SBR1_14	1055	1055	100%	0.0	93.10%	LC157990.1
	Planktothricoides raciborskii gene, MIB synthase, partial sequence, strain: SBR1_8	1055	1055	100%	0.0	93.10%	LC157987.1
	Planktothricoides raciborskii gene, MIB synthase, partial sequence, strain: SBR1_16	1053	1053	99%	0.0	93.09%	LC157992.1
	Planktothricoides raciborskii gene, MIB synthase, partial sequence, strain: SBR1_15	1053	1053	99%	0.0	93.09%	LC157991.1
May 6m-Pseudanabaena limnetica	*Pseudanabaena limnetica* str. Castaic Lake MIB synthase gene, complete cds	1304	1304	100%	0.0	99.17%	HQ630883.1
	Pseudanabaena sp. dqh15 2-methylisoborneol (2-MIB) synthesis associated operon, complete sequence	1127	1127	100%	0.0	94.76%	HQ830028.1
	Pseudanabaena galeata pgmtc gene for monoterpene cyclase, complete cds, strain: NIES-512	1110	1110	100%	0.0	94.34%	AB826230.1
	Pseudanabaena sp. NIVA-CYA 111 MIB synthase gene, partial cds	1046	1046	100%	0.0	92.69%	HQ630887.1
	Oscillatoria limosa LBD 305b MIB synthase gene, partial cds	872	872	100%	0.0	88.45%	HQ630885.1
May 10m-Planktothrix sp. 328	*Planktothrix sp.* 328 2-methylisoborneol synthase gene, partial cds	1271	1271	100%	0.0	98.48%	KJ658378.1
	Oscillatoria sp. 327/2 2-methylisoborneol synthase gene, partial cds	1266	1266	100%	0.0	98.34%	KJ658377.1
	Leptolyngbya sp. A2 MIB cyclase gene, partial cds	1061	1061	100%	0.0	93.23%	KP013063.1
	Oscillatoria limosa LBD 305b MIB synthase gene, partial cds	928	928	100%	0.0	89.92%	HQ630885.1
	Planktothricoides raciborskii gene, MIB synthase, partial sequence, strain: SBR1_14	905	905	100%	0.0	89.39%	LC157990.1

Table 2.7 BLASTn results – *MIBS* Actinobacteria.

Consensus sequence	**Sequences Producing Significant Alignments**	**Max Score**	**Total Score**	**Query Cover**	**E value**	**Per. Ident.**	**Accession**
May 6m-Streptomyces lividans	Streptomyces sp. 2114.2 genome assembly, chromosome: I	2318	2318	100%	0.0	98.34%	LT629768.1
	Streptomyces lividans TK24, complete genome	2318	2318	100%	0.0	98.34%	CP009124.1
	Streptomyces coelicolor A3(2) complete genome; segment 29/29	2318	2318	100%	0.0	98.34%	AL939132.1
	Streptomyces coelicolor orf3, orf1, p52, orf2, orf4, orf5 genes, partial and complete cds	2318	2318	100%	0.0	98.34%	AB035202.1
	Streptomyces sp. S10(2016), complete genome	1421	1421	94%	0.0	87.32%	CP015098.1

2.3.2 Microbial community composition

For the assessment of microbial population composition, a total of 12 samples were selected (Table 2.8) and sequenced, targeting the 16S rRNA gene. Relative abundance of major bacterial phyla for the samples collected north of intake (Figure 2.4) and near the dam (Figure 2.5) showed that Cyanobacteria, Proteobacteria, Bacteroidetes and Actinobacteria were the most prevalent phyla in all samples. At the north of intake location (Figure 2.4), we observed that relative abundance of Cyanobacteria follows the fluctuation of geosmin and 2-MIB concentrations for the first three samples. Specifically, in the beginning of May, when T&O concentrations are relatively low, W26 sample's composition is mainly characterized by Proteobacteria (43%), Cyanobacteria (32%), followed by Bacteroidetes. As geosmin and 2-MIB started to peak in the beginning of June, relative abundance of Cyanobacteria increased to 73% (sample W57) and Proteobacteria dropped from 43% to 11%, with Bacteroidetes also decreasing. A few days after the algaecide treatment was applied to the reservoir, the relative abundance of Cyanobacteria (sample W20) decreased to 5% and Proteobacteria increased to 58% from 11%. The results for the samples collected following that treatment (W14, W11, W4) demonstrated that the relative abundance of Cyanobacteria increased again, even though T&O compound concentrations were low. This implies that even though Cyanobacteria were present in high abundances in those samples, the synthesis genes are either not expressed or possessed by those organisms. Actinobacteria, which are also associated with geosmin and 2-MIB production in freshwater environments [*Jüttner and Watson*, 2007], had a low and relatively stable abundance, ranging from 3% to 6%. As expected, the relative abundance of Cyanobacteria and Actinobacteria did not correlate with geosim and 2-MIB concentrations (Table 2.9). Compared to Cyanobacteria, the relative abundance of Actinobacteria did not decrease as drastically in the sample collected a few days after the algaecide treatment. Regarding the relative abundance of bacterial phyla in samples collected near the dam, the first two samples followed the same patterns as the corresponding samples collected at the north of intake location. However, the relative abundance of Cyanobacteria in sample W18, which was collected after the algaecide treatment, did not decrease as the corresponding sample at the north of intake location, dropping to only 56% compared to 5% in sample W20. As the T&O concentrations decreased, the relative abundance of Cyanobacteria in samples W12 and W9 also decreased, potentially allowing Proteobacteria levels to gradually

increase. On August 15, when geosmin and 2-MIB concentrations were on the rise, Cyanobacteria also increased to 70%. Also, it is worth noting that in sample W18 Actinobacteria's relative abundance increased to 14% even though the algaecide treatment had been applied a few days earlier. Again, the relative abundance of Cyanobacteria and Actinobacteria did not correlate with T&O compound concentration, except from Cyanobacteria which showed a statistically significant positive correlation with geosmin.

Temporal changes in the relative abundance of bacterial taxa at the class level for the samples collected at the locations near the dam and north of intake are shown in Figure 2.6. The major classes of Cyanobacteria were Chloroplast and Synechococcophycideae with Chloroplast predominant in samples W26 and W24 collected on May 10. Synechococcophycideae started to become present in greater proportions from June 7 and after. Proteobacteria in the samples collected north of intake comprised the classes of Alpha, Beta and Gamma Proteobacteria. In addition, the class of Deltaproteobacteria, which is typical of sediment, was present on the samples collected at the dam location [*Teske et al.,* 2011]. Betaproteobacteria appear to be the dominant class in most of the samples except in W20 in which proteobacterial composition shifts to Gammaproteobacteria and Alphaproteobacteria. Flavobacteria, Cytophagia and Saprospirae are the major classes present, belonging to the Bacteroidetes phylum. The heat map (Figure 2.7) constructed from the 25 most unique OTU's across all samples, shows that *Pseudanabaena* and *Synechococcus* are the main Cyanobacterial genera present in most of the samples. The family of *Pseudomonadaceae* and the genus of *Flavobacterium*, both potential 2-MIB degraders belonging to the phylum of Proteobacteria, were present in high numbers in samples W20, W12, W26 and W24 [*Ho et al.,* 2007; *Izaguirre et al.,* 1988]. The family of *Sphingomonadaceae*, which includes genera capable of geosmin biodegradation, were present in samples W20, W14, W12, and W11 [*Ho et al.,* 2007].

The Alpha diversity graph (Figure 2.8), showing the relationship between phylogenetic diversity and sampling depth, indicates that since the rarefaction curves reach the plateau phase, the community was sequenced deeply enough and additional sampling would not lead to increased estimates of diversity [*Sogin et al.,* 2006; *Zhang et al.,* 2012]. Samples W24 and W26 had the lowest phylogenetic diversity while sample W11 had the highest.

Table 2.8 List of samples used for the 16S sequencing analysis.

Sampling Date	Location	Sample ID
5/10/2017	Dam	W24
5/10/2017	North of Intake	W26
6/7/2017	Dam	W56
6/7/2017	North of Intake	W57
6/20/2017	Dam	W18
6/20/2017	North of Intake	W20
7/3/2017	Dam	W12
7/3/2017	North of Intake	W14
7/17/2017	Dam	W9
7/17/2017	North of Intake	W11
8/15/2017	Dam	W3
8/15/2017	North of Intake	W4

Table 2.9 Correlation analysis results between the relative abundance of Cyanobacteria, Actinobacteria and the concentrations of geosmin and 2-MIB.

North of Intake	**Spearman rho**	**p-value**
Cyanobacteria-Geosmin	0.4058397	0.4247
Cyanobacteria-2-MIB	0.1428571	0.8028
Actinobacteria-Geosmin	0	1
Actinobacteria-2-MIB	0.4413674	0.3809
Dam	**Spearman rho**	**p-value**
Cyanobacteria-Geosmin	**0.8116794**	**0.04986**
Cyanobacteria-2-MIB	0.7714286	0.1028
Actinobacteria-Geosmin	0.1739313	0.7417
Actinobacteria-2-MIB	0.7714286	0.1028

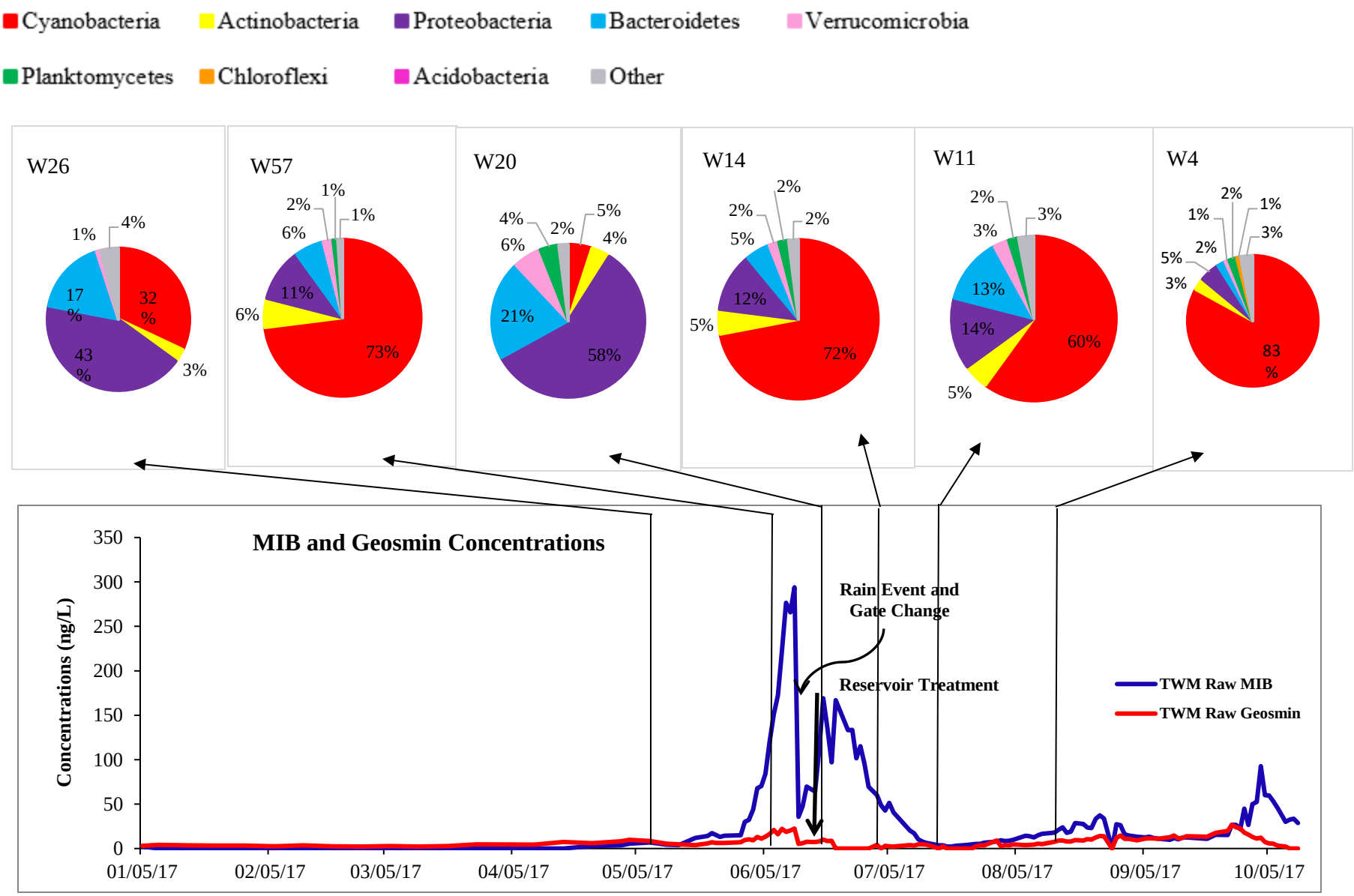


Figure 2.4 Relative abundance of bacterial phyla in comparison to geosmin and 2-MIB concentrations – North of Intake.

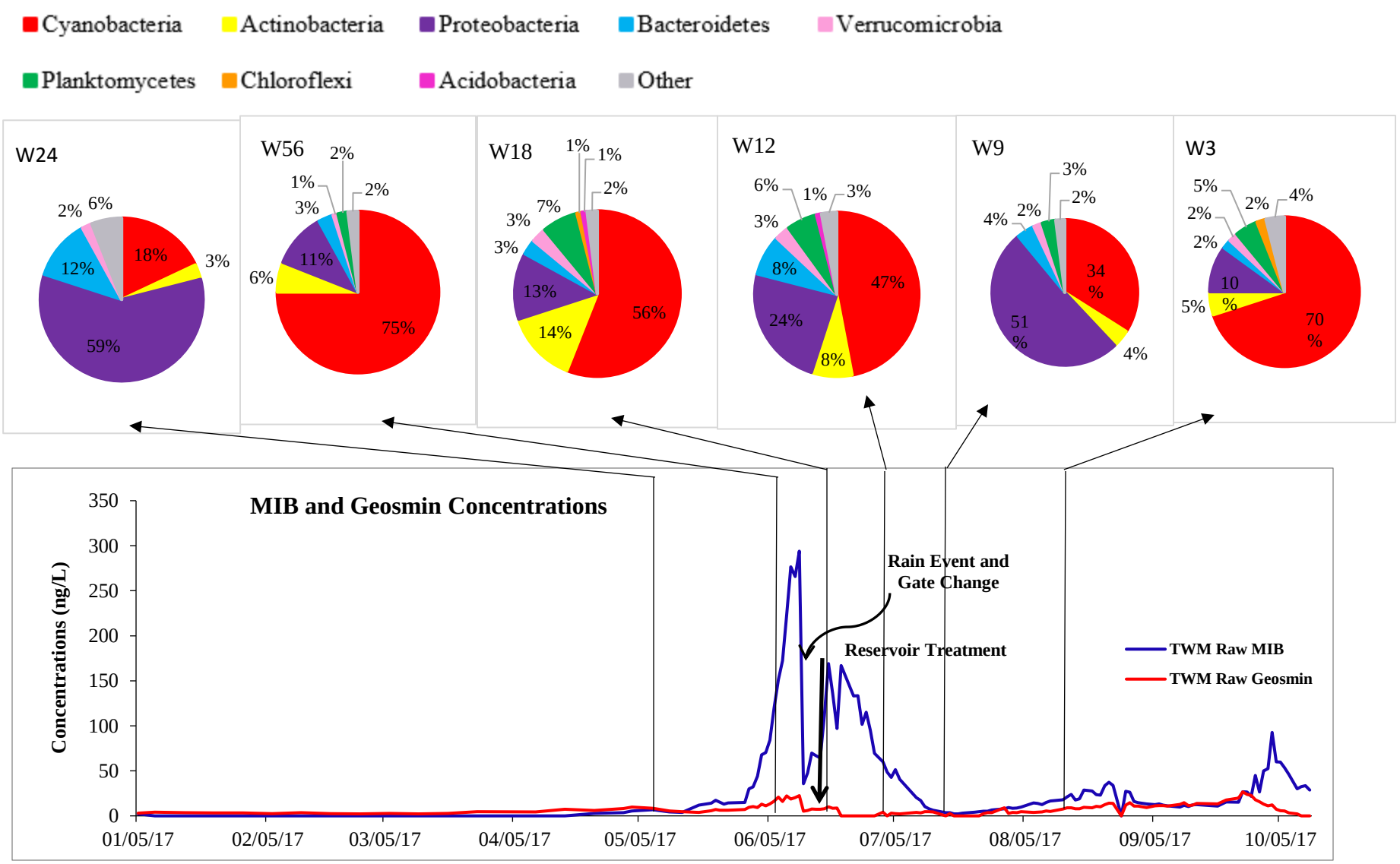


Figure 2.5 Relative abundance of bacterial phyla in comparison to geosmin and 2-MIB concentrations – Dam.

A)

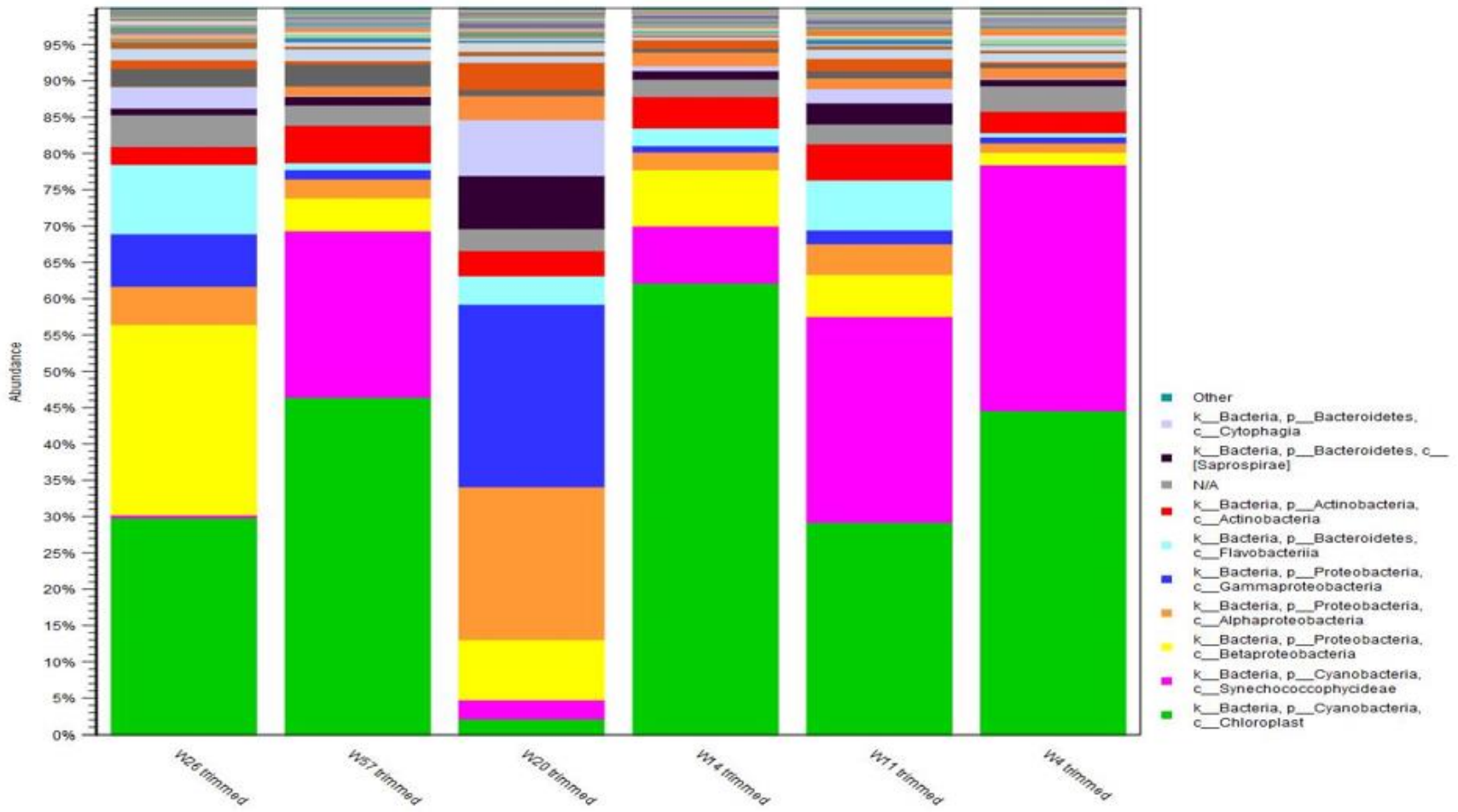
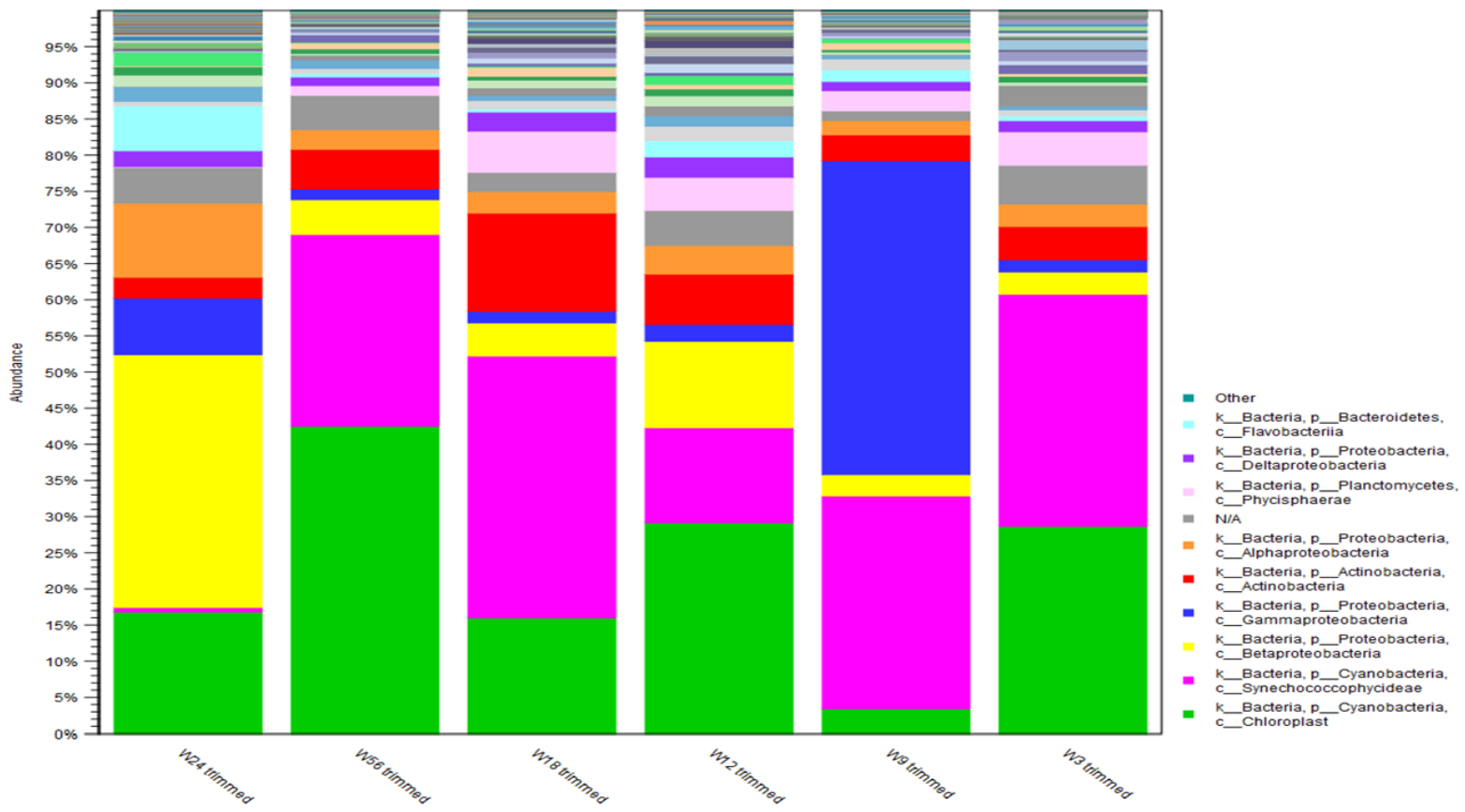


Figure 2.6 Stacked bar charts representing bacterial taxonomy based on OTU clustering analysis. A) North of Intake B) Dam.

Figure 2.6 continued

B)



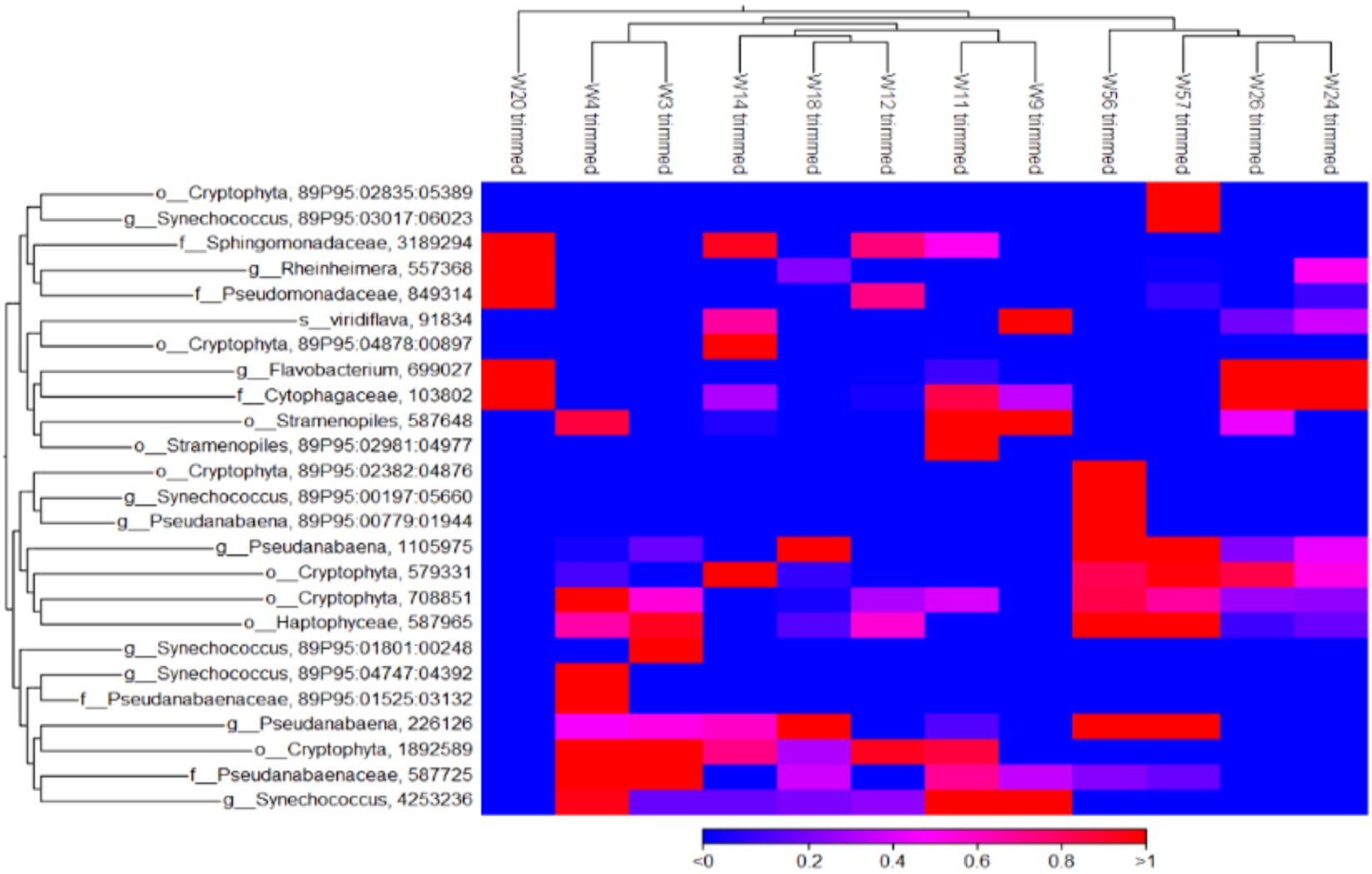


Figure 2.7 Heat Map showing the abundance of each feature in each sample along with the sample clustering.

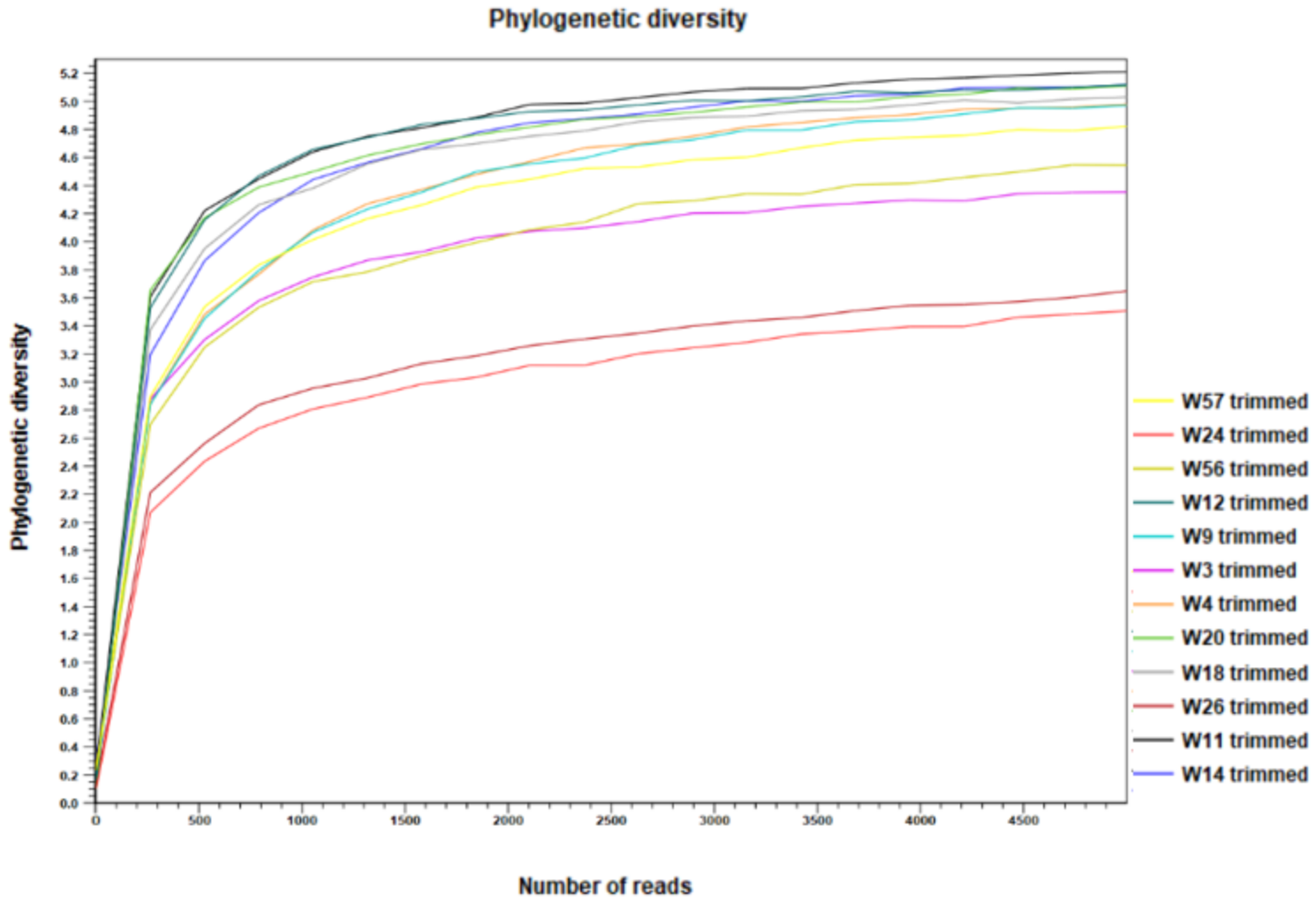


Figure 2.8 Alpha diversity graph based on phylogenetic diversity.

2.4 Discussion

Knowledge of the microorganisms responsible for the synthesis and production of T&O compounds in drinking water environments is critically important for the development of diagnostic tools and early detection and prediction of upcoming odorous episodes. Identification of microorganisms associated with geosmin and 2-MIB production in ECR has been previously achieved using a metagenomic dataset, but recovery of the key enzymes involved in the last step of their synthesis was not successful [*Clercin*, 2018]. In another study, the use of a shotgun metagenomics approach in a drinking water reservoir also failed to identify any 2-MIB synthase genes in the assembled contigs [*Otten et al.*, 2016]. Although the mapping of individual sequencing reads to known 2-MIB synthases obtained from the NCBI database revealed potential 2-MIB producers, only five reads corresponding to an *Oscillatoria* sp. or *Pseudanabaena* sp. 2-MIB synthase were successfully mapped [*Otten et al.*, 2016]. In accordance with those findings, when mapping analysis was performed in the present study to recover the genes of interest, their presence was confirmed but in low number since a few sequences relative to the total number of sample's reads was successfully mapped to the reference sequences. Furthermore, our results revealed the genetic variability of the *geoA* and *MIBS* genes in Actinobacteria, by showing the low percent identity between reference sequences corresponding to the same gene but representing different actinobacterial species. Additional support to this argument is the fact that consensus sequences derived from mapping to actinobacterial reference sequences, when used on a BLASTn search, generated significant alignments with species belonging to the same genus as the reference sequence on the mapping analysis. In contrast, consensus sequences derived from sample mappings to Cyanobacterial reference sequences, when used on the BLASTn search, generated significant alignments with various Cyanobacterial species, indicating that *geoA* and *MIBS* genes are highly conserved among them.

As expected, higher levels of reads mapped to reference sequences in May samples when the reservoir is fully mixed and T&O concentrations are higher. Also, the samples retrieved from lower depths (6-10m) had greater percentages of reads corresponding to the genes of interest. The *geoA* reference sequence of *Streptomyces fradiae* and the *MIBS* sequence of *Streptomyces lividans* had more sample reads mapped to them compared to Cyanobacterial reference sequences, indicating that Actinobacteria might contribute more to the problem. However, we need to take

into consideration the sequence length of the reference sequences, which in the case of *Streptomyces* species was longer, giving an advantage over the number of reads that could successfully align to them.

Other limitations concerning the mapping analysis of the present work, include the inability to assign enzymatic activities to specific organisms as the coverage of microbial population sequencing data is inadequate in available databases and the variability of the genes belonging to the same genera and carrying out the same function is underestimated [*Teske et al.,* 2011]. In addition, metagenomic approaches give information regarding the presence of genes in an environment and not necessarily their activity [*Kennedy et al.,* 2010]. In this study, we did not distinguish between dead or living cells, so the sequences identified might be associated with active as well as dead organisms. Future research should focus on the development of metatranscriptomic approaches to be applied on the study of T&O episodes in ECR so that we can get a better and more accurate understanding of the reservoir's microbial population dynamics.

ECR's bacterial community composition, in relation to geosmin and 2-MIB concentrations was assessed by a 16S sequencing approach. The major bacterial phyla of Cyanobacteria, Proteobacteria, Actinobacteria and Bacteroidetes that were present in ECR are commonly found in freshwater environments [*Eiler* and *Bertilsson,* 2004; *Pope* and *Patel,* 2008; *Poretsky et al.,* 2014]. *Pseudanabaena* and *Synechococcus* genera, present in high abundances in some of the samples are known T&O producers [*Izaguirre* and *Taylor,* 2004], with the latter frequently being abundant in drinking water reservoirs [*Gomez-Alvarez et al.,* 2012; *Poretsky et al.,* 2014]. In general, the relative abundance of Cyanobacteria and Actinobacteria did not correlate well with T&O compound concentrations, indicating that the production of those compounds may be done by a relatively small fraction of the total population. In addition, even though their abundance was high in some samples, the corresponding geosmin and 2-MIB concentrations were low, implying that the organisms might not possess or actively express the genes responsible for the synthesis of those compounds. Sample W18 (Dam) was the only one that had a higher relative abundance of Actinobacteria (14%) compared to the other ones. Since soil is considered to be their major natural reservoir and the dam is a construction where sediment usually accumulates, the increased levels of Actinobacteria at this location can be partially explained [*Asquith et al.,* 2013; *Yang et al.,* 2005]. In terms of Alpha diversity, the samples collected on May 10 had the lowest phylogenetic diversity. Water column mixing, happening in the reservoir at that point of time may act as a disturbance to

the microbial community composition, which is not always resistant to those events [*Poretsky et al.,* 2014; *Shade et al.,* 2011; *Shade et al.,* 2012]. Redistribution of microorganisms across water column might create antagonizing conditions, forcing them to compete for nutrient availability, something that might be against those who cannot quickly adjust to such an environment.

Algaecide treatment application was another factor affecting the microbial community composition in ECR. Even though Cyanobacterial relative abundance decreased to 5% at the north of intake sampling location, the treatment did not seem to be effective towards Actinobacteria. Location also plays an important role in the efficacy of treatment, since the relative abundance of Cyanobacteria in the sample collected near the dam did not decrease as drastically as in the sample from the north of intake location where the algaecide is usually applied. Also, a possible effect in response to the algaecide treatment is the shift of Proteobacterial composition in sample W20, from mainly being comprised of Betaproteobacteria, to increased levels of Gamma and Alpha Proteobacteria, with the latter one previously confirmed as being resistant to other treatment methods [*Chao et al.,* 2013]. Since microbial degradation of geosmin and 2-MIB is possible in ECR, it could be potentially used as an alternative treatment approach in the future for the effective management of T&O problems. Although the 16S rRNA gene sequencing approach gives a good overall view of the microbial community composition in ECR, drawing definite conclusions about their association with T&O production is not feasible since the results only produce relative abundance numbers. Development of techniques for the quantification of the genes responsible for the synthesis of geosmin and 2-MIB, which will be discussed on the next chapter, is something that could provide better insights into the origins of the problem. Furthermore, strain isolation and characterization through culturing experiments would be essential for the correct identification of T&O producers, since not all bacterial strains belonging to the same species can produce those odorous metabolites.

2.5 Conclusions

Understanding the relationship between microbial community composition and function is a major challenge in studying and identifying T&O producers in freshwater environments. The present study was able to recover the synthesis genes of geosmin and 2-MIB through the mapping analysis of a metagenomic dataset. Their presence highlights an active biosynthesis of those

compounds. Existence of bacterial phyla, common in many freshwater environments, was confirmed using a 16S rRNA gene sequencing approach. High abundances of *Pseudanabaena* and *Synechococcus* genera, which are known T&O producers, were also confirmed. Relative abundance of Cyanobacteria and Actinobacteria did not correlate well with T&O compound concentrations. This lack of correlation along with the low number of mapped reads, implies that production may be done by a relatively small fraction of the total bacterial population. However, individual contributions of microorganisms to T&O episodes is difficult to assess. The coper-based algaecide treatment applied to the reservoir, appeared to be effective only for Cyanobacteria and mostly at the north of intake. This emphasizes that the selection of location and type of algaecide is critical for the efficient management of emerging T&O episodes in the reservoir.

2.6 References

Asquith, E. A., Evans, C. A., Geary, P. M., Dunstan, R. H., & Cole, B. (2013). The role of Actinobacteria in taste and odour episodes involving geosmin and 2-methylisoborneol in aquatic environments. *Journal of Water Supply: Research and Technology-Aqua*, *62*(7), 452-467.

Auffret, M., Pilote, A., Proulx, É., Proulx, D., Vandenberg, G., & Villemur, R. (2011). Establishment of a real-time PCR method for quantification of geosmin-producing Streptomyces spp. in recirculating aquaculture systems. *Water research*, *45*(20), 6753-6762.

Chao, Y., Ma, L., Yang, Y., Ju, F., Zhang, X. X., Wu, W. M., & Zhang, T. (2013). Metagenomic analysis reveals significant changes of microbial compositions and protective functions during drinking water treatment. *Scientific reports*, *3*, 3550.

Clercin, N., (2018). Origin and fate of odorous metabolites, 2-methylisoborneol and geosmin in a eutrophic reservoir.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., ... & Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.*, *72*(7), 5069-5072.

Eiler, A., & Bertilsson, S. (2004). Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. *Environmental Microbiology*, *6*(12), 1228-1243.

Giglio, S., Chou, W. K. W., Ikeda, H., Cane, D. E., & Monis, P. T. (2010). Biosynthesis of 2-methylisoborneol in cyanobacteria. *Environmental science & technology*, *45*(3), 992-998.

Gomez-Alvarez, V., Revetta, R. P., & Santo Domingo, J. W. (2012). Metagenomic analyses of drinking water receiving different disinfection treatments. *Appl. Environ. Microbiol.*, *78*(17), 6095-6102.

Harris, T. D., Smith, V. H., Graham, J. L., Van de Waal, D. B., Tedesco, L. P., & Clercin, N. (2016). Combined effects of nitrogen to phosphorus and nitrate to ammonia ratios on cyanobacterial metabolite concentrations in eutrophic Midwestern USA reservoirs. *Inland Waters*, *6*(2), 199-210.

Ho, L., Hoefel, D., Bock, F., Saint, C. P., & Newcombe, G. (2007). Biodegradation rates of 2-methylisoborneol (MIB) and geosmin through sand filters and in bioreactors. *Chemosphere*, *66*(11), 2210-2218.

Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M., & Visser, P. M. (2018). Cyanobacterial blooms. *Nature Reviews Microbiology*, *16*(8), 471.

Izaguirre, G., Wolfe, R. L., & Means, E. G. (1988). Degradation of 2-methylisoborneol by aquatic bacteria. *Appl. Environ. Microbiol.*, *54*(10), 2424-2431.

Izaguirre, G., & Taylor, W. D. (2004). A guide to geosmin-and MIB-producing cyanobacteria in the United States. *Water Science and Technology*, *49*(9), 19-24.

Jüttner, F., & Watson, S. B. (2007). Biochemical and ecological control of geosmin and 2-methylisoborneol in source waters. *Appl. Environ. Microbiol.*, *73*(14), 4395-4406.

Kennedy, J., Flemer, B., Jackson, S. A., Lejon, D. P., Morrissey, J. P., O'gara, F., & Dobson, A. D. (2010). Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism. *Marine drugs*, *8*(3), 608-628.

Lee, J., Rai, P. K., Jeon, Y. J., Kim, K. H., & Kwon, E. E. (2017). The role of algae and cyanobacteria in the production and release of odorants in water. *Environmental pollution*, *227*, 252-262.

Li, L., Pascual, D. L., Tedesco, L. P., Randolph, K. L., Sengpiel, R. E., Hall, B. E., & Wilson, J. S. (2006). Developing a survey tool for the rapid assessment of blue-green algae in central Indiana's reservoirs. *IUPUI. Indianapolis: Central Indiana Resources Partnership/CEES*, *7*(11).

McVeigh, H. P., Munro, J., & Embley, T. M. (1996). Molecular evidence for the presence of novel actinomycete lineages in a temperate forest soil. *Journal of industrial microbiology*, *17*(3-4), 197-204.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., ... & Lopez, R. (2013). Analysis tool web services from the EMBL-EBI. *Nucleic acids research*, *41*(W1), W597-W600.

Otten, T. G., Graham, J. L., Harris, T. D., & Dreher, T. W. (2016). Elucidation of taste-and-odor producing bacteria and toxigenic cyanobacteria by shotgun metagenomics in a Midwestern drinking water supply reservoir. *Applied and Environmental Microbiology*, AEM-01334.

Pope, P. B., & Patel, B. K. (2008). Metagenomic analysis of a freshwater toxic cyanobacteria bloom. *FEMS microbiology ecology*, *64*(1), 9-27.

Poretsky, R., Rodriguez-R, L. M., Luo, C., Tsementzi, D., & Konstantinidis, K. T. (2014). Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PloS one*, *9*(4), e93827.

Seashols-Williams, S., Green, R., Wohlfahrt, D., Brand, A., Tan-Torres, A. L., Nogales, F., ... & Singh, B. (2018). An accurate bacterial DNA quantification assay for HTS library preparation of human biological samples. *Electrophoresis*, *39*(21), 2824-2832.

Shade, A., Read, J. S., Welkie, D. G., Kratz, T. K., Wu, C. H., & McMahon, K. D. (2011). Resistance, resilience and recovery: aquatic bacterial dynamics after water column disturbance. *Environmental microbiology*, *13*(10), 2752-2767.

Shade, A., Read, J. S., Youngblut, N. D., Fierer, N., Knight, R., Kratz, T. K., ... & Whitaker, R. J. (2012). Lake microbial communities are resilient after a whole-ecosystem disturbance. *The ISME journal*, *6*(12), 2153.

Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. *Frontiers in plant science*, *5*, 209.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., ... & Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences*, *103*(32), 12115-12120.

Song, K., Li, L., Li, S., Tedesco, L., Hall, B., & Li, L. (2012). Hyperspectral remote sensing of total phosphorus (TP) in three central Indiana water supply reservoirs. *Water, Air, & Soil Pollution*, *223*(4), 1481-1502.

Suurnäkki, S., Gomez-Saez, G. V., Rantala-Ylinen, A., Jokela, J., Fewer, D. P., & Sivonen, K. (2015). Identification of geosmin and 2-methylisoborneol in cyanobacteria and molecular detection methods for the producers of these compounds. *Water research*, *68*, 56-66.

Teske, A., Durbin, A., Ziervogel, K., Cox, C., & Arnosti, C. (2011). Microbial community composition and function in permanently cold seawater and sediments from an Arctic fjord of Svalbard. *Appl. Environ. Microbiol.*, *77*(6), 2008-2018.

Wang, Z., Song, G., Shao, J., Tan, W., Li, Y., & Li, R. (2016). Establishment and field applications of real-time PCR methods for the quantification of potential MIB-producing cyanobacteria in aquatic systems. *Journal of applied phycology*, *28*(1), 325-333.

Winter, J. G., DeSellas, A. M., Fletcher, R., Heintsch, L., Morley, A., Nakamoto, L., & Utsumi, K. (2011). Algal blooms in Ontario, Canada: increases in reports since 1994. *Lake and Reservoir Management*, *27*(2), 107-114.

Yang, S. L., Zhang, J., Zhu, J., Smith, J. P., Dai, S. B., Gao, A., & Li, P. (2005). Impact of dams on Yangtze River sediment supply to the sea and delta intertidal wetland response. *Journal of Geophysical Research: Earth Surface*, *110*(F3).

Zhang, F., She, Y. H., Chai, L. J., Banat, I. M., Zhang, X. T., Shu, F. C., ... & Hou, D. J. (2012). Microbial diversity in long-term water-flooded oil reservoirs with different in situ temperatures in China. *Scientific Reports*, *2*, 760.

Zwart, G., Crump, B. C., Kamst-van Agterveld, M. P., Hagen, F., & Han, S. K. (2002). Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquatic microbial ecology*, *28*(2), 141-155.

CHAPTER 3. DEVELOPMENT AND APPLICATION OF QUANTITATIVE-PCR ASSAYS TARGETING GEOSMIN AND 2-MIB PROCUCING CYANOBACTERIA IN EAGLE CREEK RESERVOIR

3.1 Introduction

Progressive eutrophication and pollution have caused the accelerated growth of algal blooms in water bodies worldwide which are related to the steady increase of odor incidents in diverse waters such as lakes, aquaculture systems and reservoirs [*Auffret et al.,* 2011; *Huisman et al.,* 2018; *Lee et al.,* 2017; *Wang et al.,* 2016]. Most of the taste and odor episodes in drinking water are caused by the terpenoid bacterial metabolites 2-MIB (2-methylisoborneol) and geosmin (trans-1,10-dimethyl-trans-9-decalol) which produce a mildewed and earthy flavor respectively [*Lee et al.,*2017; *Jüttner and Watson,* 2007; *Watson et al.,* 2008]. Although there is no evidence for adverse health effects related to those compounds, their presence in water cause consumer complaints who consider it unsafe for drinking [*Auffret et al.,* 2011; *Chiu et al.,* 2016].

Each compound exists as (+) and (-) enantiomers (Figure 1.1), but odor outbreaks are caused by the biological production of the (-) stereoisomer which is 10 times more potent than the (+) molecule [*Jüttner and Watson,* 2007]. Their low odor detection by human senses (2-10 ng/L), along with their strong resistance to oxidation, a process commonly applied to water purification, make them the most persistent and disturbing substances in odor incidents [*Suurnäkki et al.,* 2015; *Wang et al.,* 2016]. A wide variety of microorganisms can synthesize and secrete 2-MIB and geosmin, including fungi, proteobacteria and myxobacteria, but cyanobacteria and actinobacteria species are regarded as the main producers in freshwater ecosystems [*Auffret et al*., 2011; *Wang et al*., 2016]. The production mechanism of the terpenoid metabolites in cyanobacteria and actinomycetes is similar with homologous genes involved in the synthesis [*Giglio et al.,* 2010; *Wang et al.,* 2016]. Geosmin is synthesized through the conversion of farnesyl diphosphate to geosmin catalyzed by geosmin synthase enzyme encoded by the *geoA* gene [*Giglio et al.,* 2010; *Suurnäkki et al.,* 2015]. 2-MIB synthesis consists of 2 key reactions: 1) methylation of geranyl diphosphate (GPP) to 2-methyl-GPP, and, 2) cyclization of 2-methyl-GPP to 2-MIB catalyzed by MIB synthase [*Giglio et al.,* 2011]. Most of the production is retained within the cells, with the bulk release occurring during senescence, death and biodegradation [*Lee et al.,* 2017; *Watson et al.,* 2008].

Current approaches followed by water utilities for detection and quantification of odorants in water are mainly based on gas chromatography/mass spectrometry (GC/MS), coupled with various enrichment and extraction techniques, such as SPME [*Liato and Aïder,* 2017; *Tsao et al.,* 2014; *Watson et al.,* 2000]. Although this method is sensitive, it has several drawbacks including the technical expertise required by the operator, the expensive and large instrumentation that cannot be transported and used for on-site monitoring, and most importantly, the fact that it does not identify the biological source of the problem [*John et al.,* 2018; *Su et al.,* 2013]. Microscopy and traditional plate count techniques have been employed and widely used for the identification and enumeration of odor-producing bacteria. However, there are several limitations associated with those methods as most bacteria in nature are difficult to culture and require highly selective media [*Jüttner and Watson,* 2007; *Nielsen et al.,* 2006]. Furthermore, microscopic techniques require taxonomical expertise by the operator, long analytical time and lack the ability of distinguishing between T&O producers and non-producers [*Jüttner and Watson,* 2007; *Su et al.,* 2013].

An alternative approach for investigation of the microbial T&O producers is the use of quantitative real-time polymerase chain reaction (qPCR), which in contrast to the conventional PCR, quantifies the targets rather than identifying its presence or absence. One of the major advantages using this molecular-based assay is its unique identification, as primers and probes are designed to specifically bind to the gene of interest. In addition, it is a sensitive method capable of detecting very low cell numbers, useful for for rapid on-site monitoring of emerging T&O episodes in a relatively inexpensive way [*John et al.,* 2018; *Lyllofff et al.,* 2012]. In the last few years, this method has been widely used for the detection of T&O bacterial producers in various environments such as recirculating aquaculture systems [*Auffret et al.,* 2011], liquors [*Du et al.,* 2013] and reservoirs [*Chiu et al.,* 2016], targeting not only the *geoA* and 2-*MIB* synthase genes but also the 23S rRNA gene in *Streptomycetes* [*Lylloff et al.,* 2012]. Several studies, using primers targeting the *geoA* and *MIBS* genes in Cyanobacteria and Actinobacteria have successfully demonstrated the positive correlation between the genes and the respective geosmin and 2-MIB concentrations [*Giglio et al.,* 2008; *Su et al.,* 2013; *Wang et al.,* 2016]. This demonstrates that qPCR techniques could potentially serve as a valuable tool for the prediction and early detection of emerging T&O events.

Eagle Creek Reservoir, which supplies about 80% of Indianapolis drinking water, experiences frequent and sometimes severe T&O episodes caused by elevated levels of geosmin and 2-MIB. The odorous events are usually observed during the Spring, when the reservoir receives most of its water and is fully mixed and turbid, and at the beginning of Fall. In order to investigate the origin of the problem, qPCR assays targeting the synthesis genes of those compounds in Cyanobacteria were developed and applied to water samples collected during the 2018 sampling season (April-October). The objectives of the present work are to explore the correlation between the copy number of *geoA* and *MIBS* genes and the geosmin and 2-MIB concentrations, investigate the spatiotemporal variation of gene abundance in the reservoir and finally, evaluate the effectiveness of 16S rRNA qPCR and cell counting techniques as predictors of T&O episodes.

3.2 Materials and Methods

3.2.1 Study site

Eagle Creek Reservoir (ECR; 86°18′13.07″W, 39°51′09.84″N; A = 5.0 km^2; Z = 4.2 m), was constructed in 1967 in the northwest corner of Marion County to initially provide flood control and then drinking water for the city of Indianapolis when T.W. Moses Drinking Water Plant became available in 1976 [*Harris et al.,* 2016; *Li et al.,* 2006]. The depth ranges from 4 to 13 meters, with the deepest areas located in the southern basin, near the dam. The reservoir is classified as mesotrophic to eutrophic [*Song et al.*, 2012], and its water column is mixed and turbid in April/May when it receives most of its water from snow melt and intense rainfall and then in October after the seasonal thermal stratification is over.

3.2.2 Sample collection and processing

Sampling season started in late April and lasted through October, resulting in 24 collection dates and 328 samples in total in 2018. Water samples were collected using a vertical Van Dorn sampler at 14 locations over the reservoir (Figure 3.1) and 50ml of water were immediately filtered through a 0.22 μm Sterivex filter unit. For preservation purposes, 3ml of RNA*later* stabilization solution was added to the filter. Filters were kept on ice and stored in a -80 °C freezer until further processing.

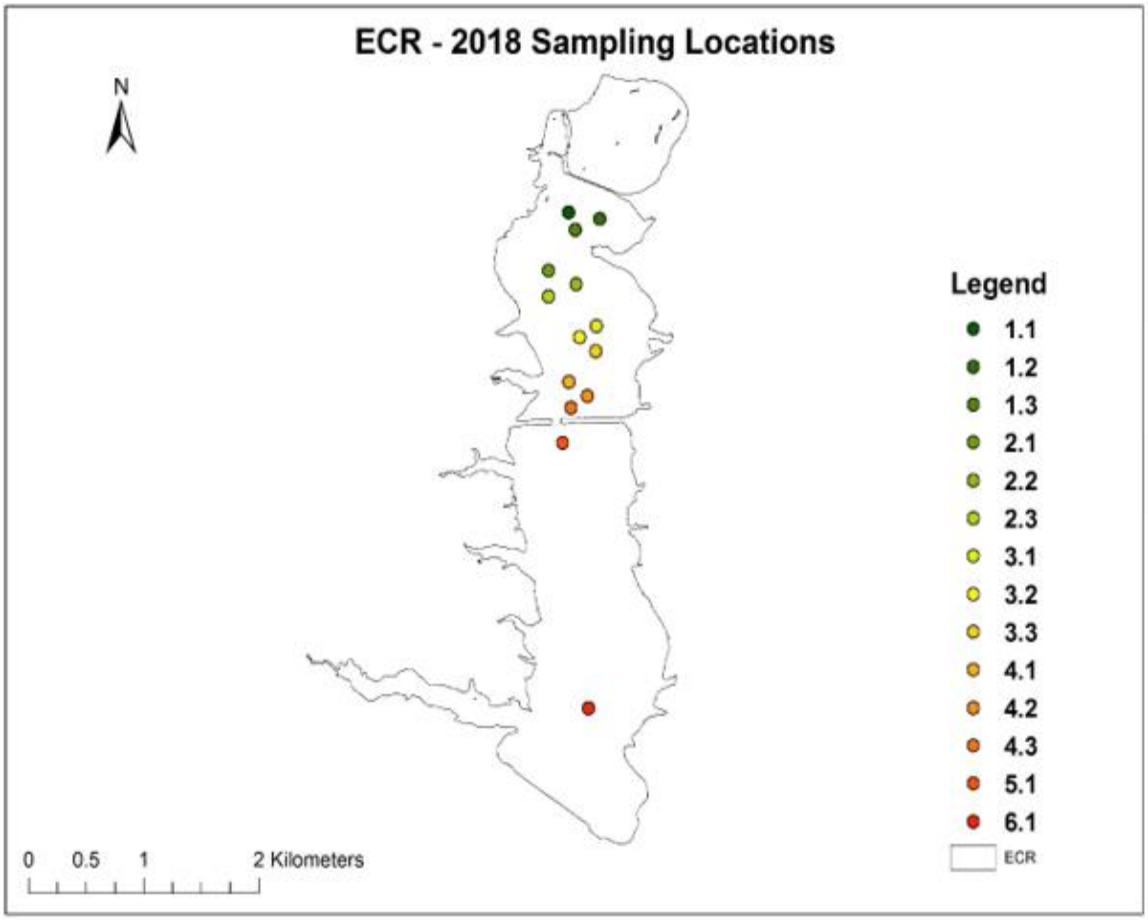


Figure 3.1 ECR-Sampling locations 2018.

3.2.3 DNA extractions

Genomic DNA of the water samples was extracted using the DNeasy PowerWater Sterivex Kit (Qiagen). Lysis buffer was added to the Sterivex filter units, which were then mixed. Subsequently, the lysate was removed for additional lysis using a 5 ml bead beating tube. After the removal of proteins and inhibitors, total genomic DNA was captured on an MB Spin Column. The column was washed, and purified DNA was eluted in 100 µL of EB solution (sterile elution buffer) and stored at −20 °C. The concentration and purity of the DNA was measured with a NanoDrop 2000/2000c Spectrophotometer (Thermo Scientific Inc.). Bacterial DNA, used as a standard for the qPCR assay, was extracted using the Quick-DNA Fungal/Bacterial kit (Zymo Research).

3.2.4 Primer design

Sequences of the *geoA* and *MIBS* were obtained from the NCBI database and were aligned using Clustal Omega (multiple sequence alignment program) and BioEdit. Primers (Table 3.1) and probes (Table 3.2) were designed in conserved locations along the length of the genes using the PrimerQuest Tool. For MIB synthase detection in Cyanobacteria, primers were designed based on the following sequences: *Pseudanabaena* sp. NIVA-CYA 111 (HQ630887.1), *Oscillatoria* sp. 327/2 (KJ658377.1) and *Planktothrix* sp. 328 (KJ658378.1). For geosmin synthase detection, primers were designed based on the following cyanobacterial sequences: *Nostoc* sp. UK1 (KJ658372.1), *Aphanizomenon* sp. PMC9501 (KJ658367.1), *Planktothrix* sp. 18 (KJ658376.1), and *Oscillatoria* sp. PCC 9240 (KJ658375.1). Estimation of total bacterial load was based on the universal 16S primers and probe retrieved from *Nadkarni et al.* [2002].

Table 3.1 Primers used in this study.

Primers	Sequence (5' to 3')	Expected Length (bp)	Target	Reference
CYN_GEO_S5 F	ATACTYGGMTCARAAGARGAAGC	103	***geoA***	This study
CYN_GEO_S5 R	GYATCTGGATGRGTATAKGARCA			
Cy_MIB_S3.F	CAGCTTCTACACCTCCATGAC	135	***MIBS***	This study
Cy_MIB_S3.R	AGGAGATCGTTGACCAKRACG			
16S_uni.F	TCCTACGGGAGGCAGCAGT	466	***16S* rRNA**	*Nadkarni et al.* (2002)
16S_uni.R	GGACTACCAGGGTATCTAATCCTGTT			

Table 3.2 Probes used in this study.

Probes	Sequence (5' to 3')	Reference
CYN_GEO	**NED**-GATKCYCAYGACTACGCCYTRCT-**MGBNFQ**	This study
CYN_MIB	**VIC**-TCCGTCACGCAGCGTTCYTRGC-**MGBNFQ**	This study
16SU	**FAM**-CGTATTACCGCGGCTGCTGGCAC-**TAMRA**	*Nadkarni et al.* (2002)

3.2.5 PCR amplification

Initial test for amplification of the desired target genes, was performed in a total volume of 10.5 µL, containing 3µL of extracted DNA, 1 µL (10µM) of each of the forward and reverse primers for either *geoA* or *MIBS*, 5 µL of 2X PCR Master mix and 0.5 µL of 2mg/ml bovine serum

albumin. For the primers targeting the *geoA* gene, amplification was performed at 94°C for 8 min, then 40 cycles at 94 °C for 15 sec, followed by 1 min at 52 °C, 72°C for 30 sec, and a final extension period of 1 min at 72°C. Amplification conditions were exactly the same for 16S and *MIBS* detection except from the annealing temperatures that were 55°C and 53°C respectively. *Nocardia* sp. (ATCC 202099) which belongs to Actinobacteria was used to evaluate the specificity of the primers towards Cyanobacteria and nuclease-free water was used as a negative control. Positive controls included genomic DNA from *E.coli* (ATCC 700928DQ), *Nostoc sp.* (ATCC 53789) and *Pseudanabaena sp.* (ATCC 29210) for the primers targeting the 16S rRNA, *geoA* and *MIBS* genes respectively. Field samples were used as well to test for the primer's effective amplification. PCR product quality was evaluated after electrophoresis on agarose gels.

3.2.6 Standards for qPCR analysis

For the 16S assay, genomic DNA from *E.coli* (ATCC 700928DQ) was used as a standard, and 6 serial dilutions were performed on an initial concentration of 35000 copies/μL, with a dilution factor of 1/5. For the *geoA* assay, *Nostoc sp.* (ATCC 53789) was used as a standard and 6 serial dilutions were performed on an initial concentration of $4.12*10^{10}$ copies/μL with a dilution factor of 1/5. Based on the MIB synthase sequence of Pseudanabaena sp. NIVA-CYA 111, a DNA fragment (Table 3.3) was synthesized (Integrated DNA Technologies) and used as a standard for the *MIBS* assay. The standard was serially diluted 7 times, starting at a concentration of $60.22*10^{7}$ copies/μL with a dilution factor of 1/10. The number of gene copies was calculated as follows:
N (copies/ml) = C g/ml * $[Y\ bp\ *\ 660\ g\ mole^{-1}bp^{-1}]^{-1}N_A$ [*Auffret et al.*, 2011; *Su et al.*, 2013], where C is the DNA concentration, Y is the PCR fragment length and N_A represents Avogadro's constant.

Table 3.3 Sequence of the synthesized *MIBS* DNA fragment.

***MIBS* Fragment (5'-3')**
CTGGCTGCGCGACAGCACGACAGCTTCTACACCTCCATGACGCTAATCGACCCCA TCGGAGGCTACGTCCTCCCACCAGATATTTTCTTCGATTCGCGCGTCCGTCACGCA GCGTTCCTAGCTGGGACGGCGGTCGTTCTGGTCAACGATCTCCTTTCGGTCGCCAA AGATCTGG

3.2.7 qPCR protocol

DNA from the water samples was aliquoted into 96-well plates and reactions were performed in triplicate for each assay. Each qPCR reaction was performed in a total volume of 12.5 µL. For the 16S assay, the reaction consisted of 6.25 µL TaqMan Universal PCR Master Mix, 0.25 µL of 10µM of each of the primers and probe, 2.5 µL DNA, 1µL bovine serum albumin and 2µL of nuclease free water. The amplification conditions were as follows: an initial denaturation step of 94°C for 8 min, followed by 40 cycles of 94 °C for 15s, 55 °C for 60s and 72°C for 30s. For the *geoA* assay, the reaction consisted of 6.25 µL TaqMan Universal PCR Master Mix, 0.4 µL of 10µM of each of the primers and probe, 2.5 µL DNA, 1µL bovine serum albumin and 1.55µL of nuclease free water. The amplification conditions were as follows: an initial denaturation step of 94°C for 8 min, followed by 40 cycles of 94 °C for 15s, 52 °C for 60s and 72°C for 30s. Finally, the *MIBS* assay consisted of 6.25 µL TaqMan Universal PCR Master Mix, 0.5 µL of 10µM of each of the primers and probe, 2.5 µL DNA, 1µL bovine serum albumin and 1.25µL of nuclease free water. The amplification conditions were as follows: an initial denaturation step of 94°C for 8 min, followed by 40 cycles of 94 °C for 15s, 53 °C for 60s and 72°C for 30s on a 7500 real time PCR system (ThermoFisher).

3.2.8 T&O analysis

2-Methylisoborneol and geosmin concentrations were quantified by a Head-Space Solid-Phase Micro-Extraction (HS-SPME) combined with a Gas Chromatography-Mass Spectrometry (GC-MS) at the Citizens Energy lab.

3.2.9 Cell counts

Algae cell count analysis was performed by Chase Howard and conducted by hand using microscopy with a Nagoette Bright Line Hemacytometer stage (Hausser Scientific, 0.500 mm stage depth) and an Olympus BX53 microscope (100x, 200x, and 400x magnifications) on unfiltered water samples treated with 1-2% Lugol's solution.

3.2.10 Statistical analysis

For statistical analyses, R v64 3.5.3 software package was used [R Development Core Team, 2011]. Spearman's rho correlations were performed to assess potential links between gene

densities and geosmin and 2-MIB concentrations. The Kruskal-Wallis test was used to determine if there are statistically significant differences between sampling locations for the measured gene abundances by qPCR.

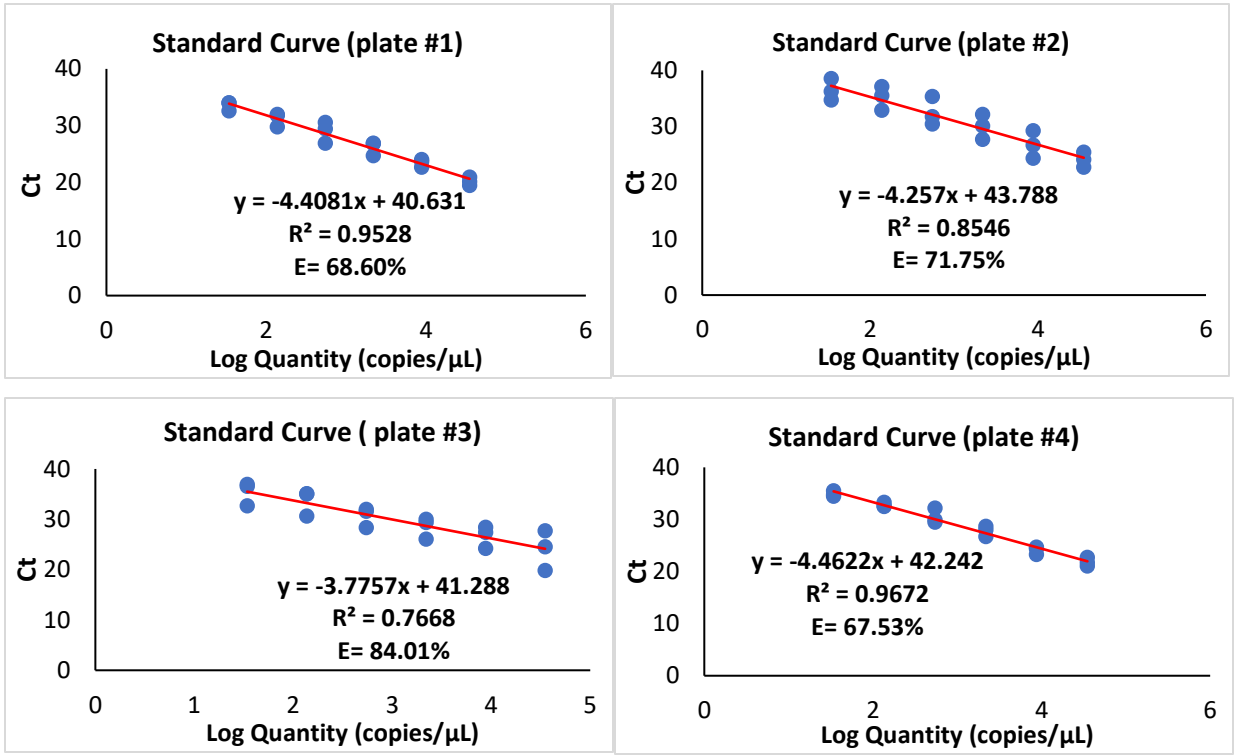
3.2.11 Spatial maps

Spatial maps were created with ArcMap, a feature within the ArcGIS [ESRI Inc., 2019].

3.3 Results

The standard curves for each of the 16S, *geoA* and *MIBS* TaqMan assays were generated by using serial dilutions of genomic DNA from *E.coli* (ATCC 700928DQ), *Nostoc sp.* (ATCC 53789) and the synthesized *MIBS* DNA fragment (Table 3.3), respectively. Due to the large volume of samples, a separate standard curve was generated for each plate of each assay. The standard curves shown in Figure 3.2 were linear with high correlation coefficients. For the 16S, *geoA* and *MIBS* assays, R^2 had a range of 0.7668 – 0.9672, 0.971 – 0.9922 and 0.9576 – 0.9879, respectively. qPCR efficiency (E) was calculated based on the equation, E = ($10^{-1/s}$- 1) where s is the slope of the standard curve [*Tsao et al.,* 2014]. The efficiencies ranged between 67.53-84.01% for the 16S assay, 71.22-81.86 % for the *geoA* assay and 64.5-86.53% for the *MIBS* assay. Those values are outside the desired range of 90-105%, which is somewhat expected considering the degenerate nature of our primers and probes [*Seashols-Williams et al.,* 2018]. If we assume that the methodological threshold was C_t=40 cycles [*Lylloff et al.,* 2012], then the detection limit for the 16S assay for plates 1 through 4 was 1.39, 7.76, 2.19 and 3.18 copies/µL. In the same way, regarding the *geoA* assay the detection limit for plates 1 through 4, was 28,438.1 (copies/µL), 37,792 (copies/µL), 42,169.7 (copies/µL) and 92,257.1 (copies/µL). For the *MIBS* assay detection limits for plates 1 through 4 were 3.19 (copies/µL), 2.57 (copies/µL), 4.16 (copies/µL), and 47.93 (copies/µL).

A)



B)

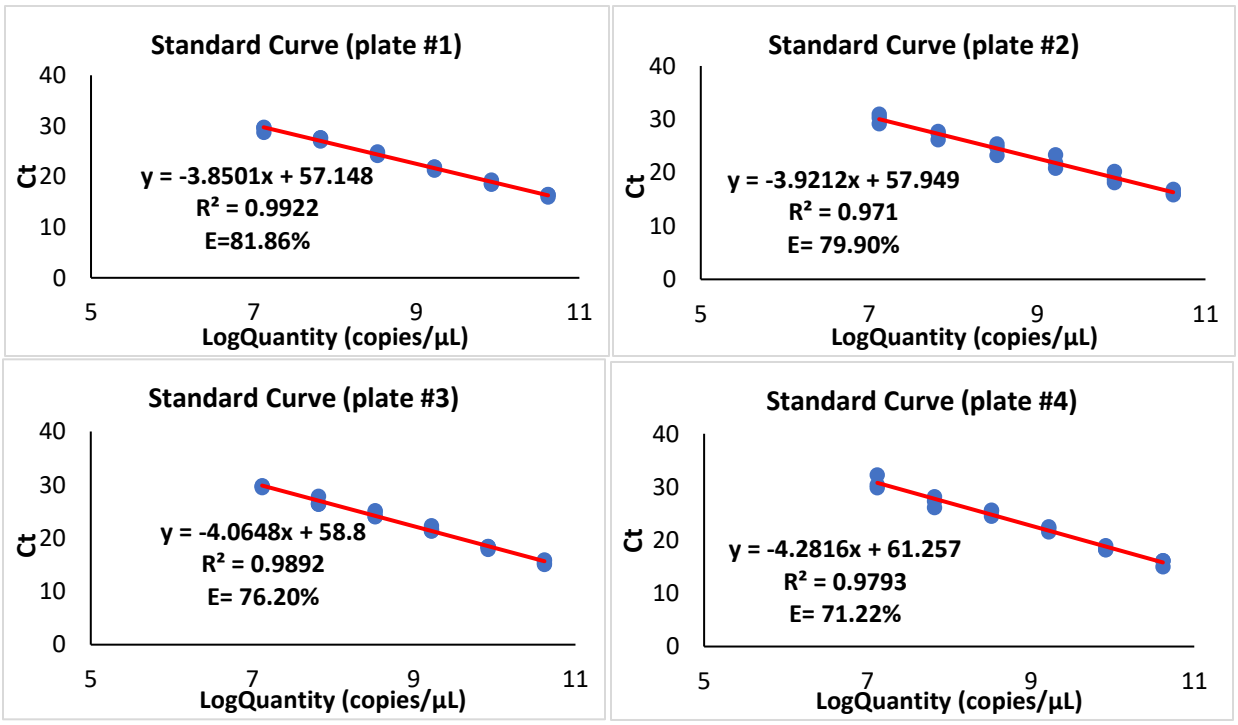
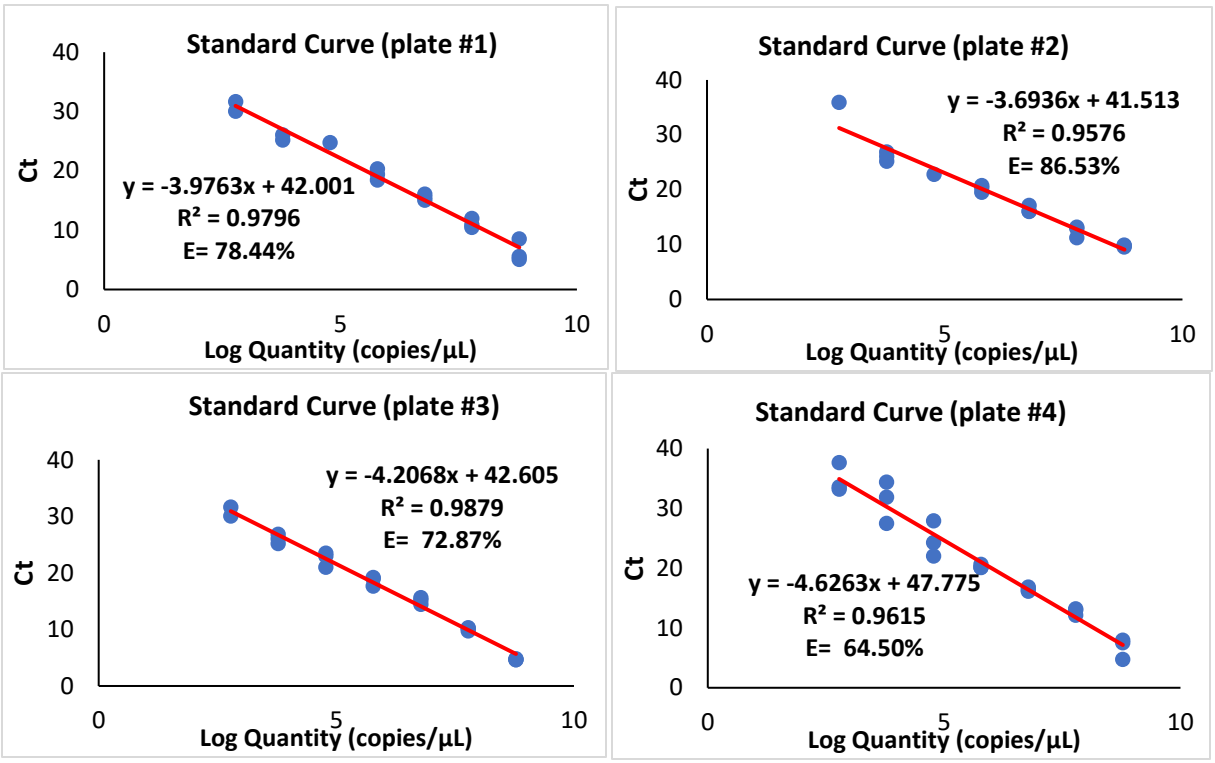


Figure 3.2 Standard curves A) 16S assay B) *geoA* assay C) *MIBS* assay.

Figure 3.2 continued

C)



Raw water samples were analyzed for odorous compounds by the Citizens Energy lab. Looking at geosmin concentrations at the first sampling location of each grid (Figure 3.3), it is noticeable that major peaks occurred on early May and then during late August and at the beginning of Fall. Geosmin detections show minima during the Summer months. Maximal value of geosmin was observed in May 8 for location 4.3 (Figure 3.4), which was 24.38 ng/L and exceeded the odor threshold of 4 ng/L [*Clercin,* 2018]. Time series plots (Figure 3.4) for the *geoA* gene detected by the qPCR assay showed that the major peaks for all locations occurred in early May, and some minor ones during late August and at the beginning of Fall. Algaecide treatments to terminate T&O production were applied on May 9th, June 26th and September 18th. Geosmin showed an immediate response to the May 9th treatment as concentration dropped to below detection limits on the following collection day (May 10th) for all sampling locations. Response to the June 26th treatment was not immediate for the sampling points located north of intake, as geosmin concentrations peaked a few days later on June 28 and sharply decreased after that. The June 26th treatment did not seem to affect geosmin concentrations at locations 5.1 and 6.1 where

the levels were already minimal and remained this way after the treatment as well. The treatment applied on September 18th successfully decreased geosmin concentration levels on locations within grids 1,2,3 and 4 as observed on the subsequent sampling date which was on September 20. This treatment, however, did not disrupt the geosmin concentration on locations 5.1 and 6.1. Overall the effects of the last algaecide treatment did not last long since concentrations peaked again, beginning September 27. Regarding the *geoA* gene levels, response to the May 9th treatment was either immediate or delayed as observed in the case of locations 1.1, 1.2 and locations in grid 2 & 3, where an increase occurred on May 10 and the levels dropped after that. On June 28, which was the immediate sampling date after the June 26th treatment, all sampling points except from location 4.3 showed minimal or below detection levels of *geoA* gene concentration. Finally, on September 20, the immediate sampling collection date after the September 18th algaecide treatment, *geoA* concentrations were below detection limits except from location 1.1. Although the fluctuation patterns between geosmin concentration and *geoA* gene levels in our samples follow one another closely in some parts, later in the sampling season the major peaks in geosmin concentration do not correspond well with gene abundances. In order to evaluate the association between those two variables, Spearman's rho correlation analysis was performed. Results (Table 3.4) indicated that a statistically significant positive association between geosmin concentration and *geoA* gene existed only for location 3.3 with rho=0.70 and for location 4.3 with rho=0.51.

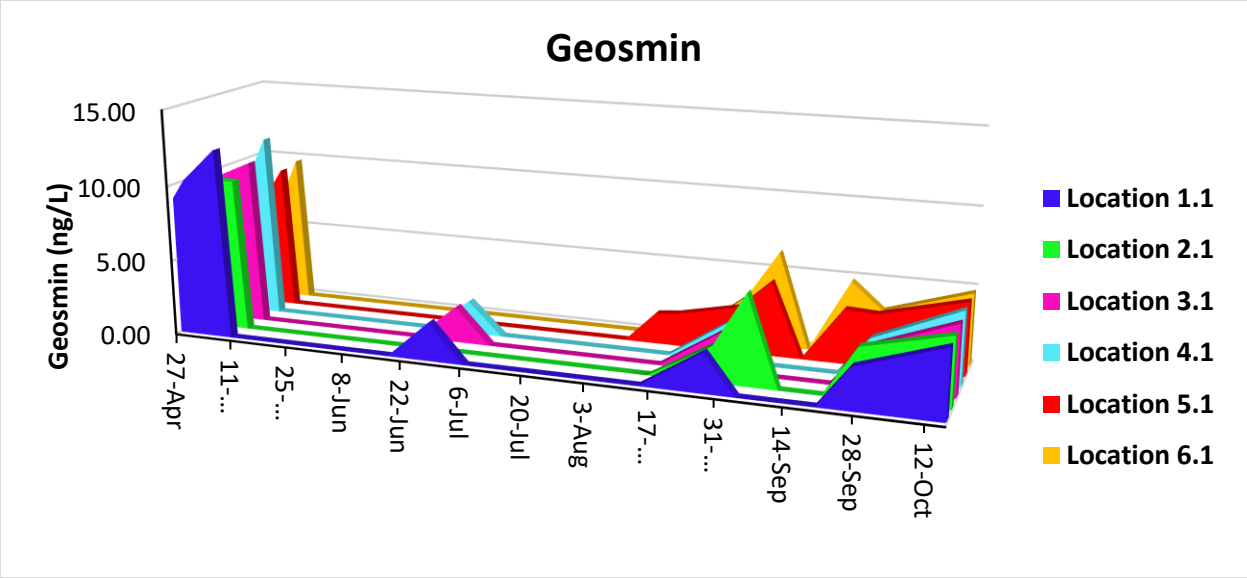


Figure 3.3 Geosmin concentration over time.

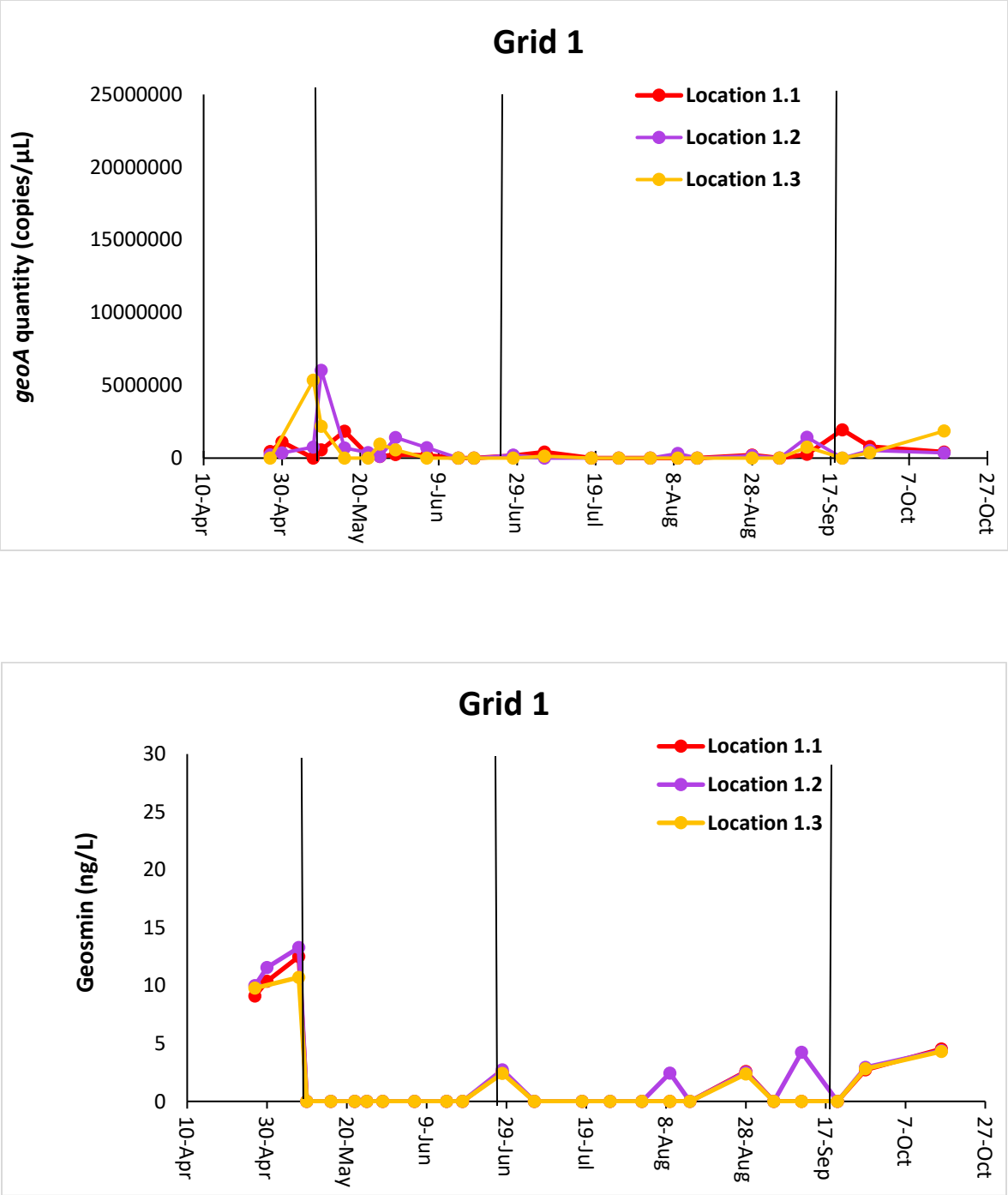
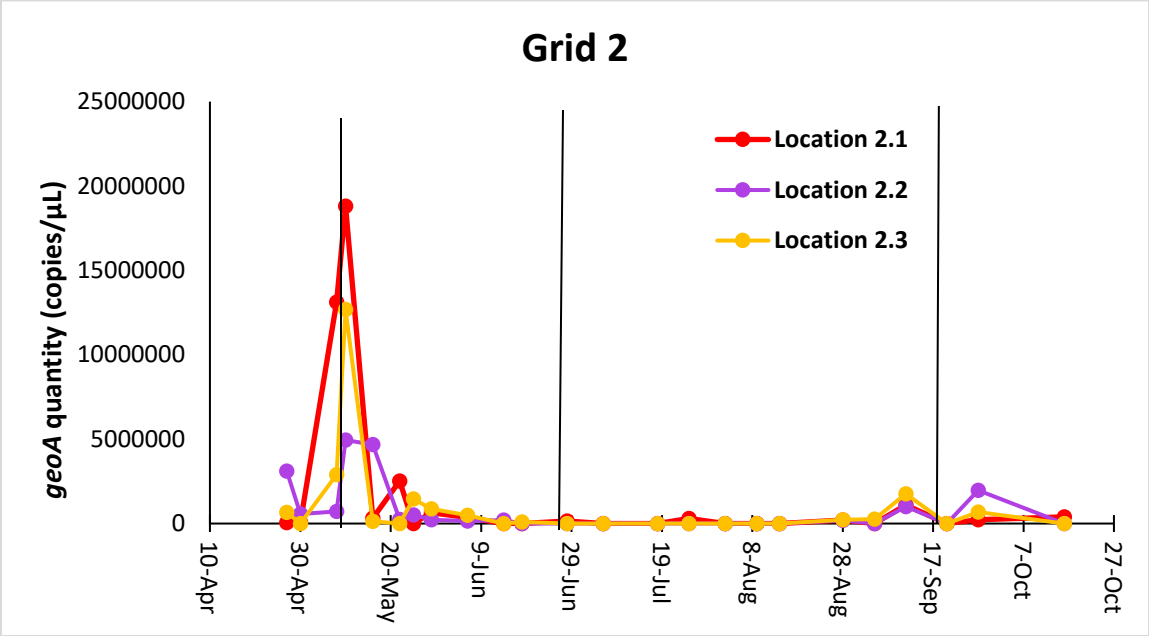


Figure 3.4 Time series graphs of geosmin concentration and *geoA* gene quantity for the sampling locations within each grid. Vertical black lines indicate algaecide treatment dates.

Figure 3.4 continued



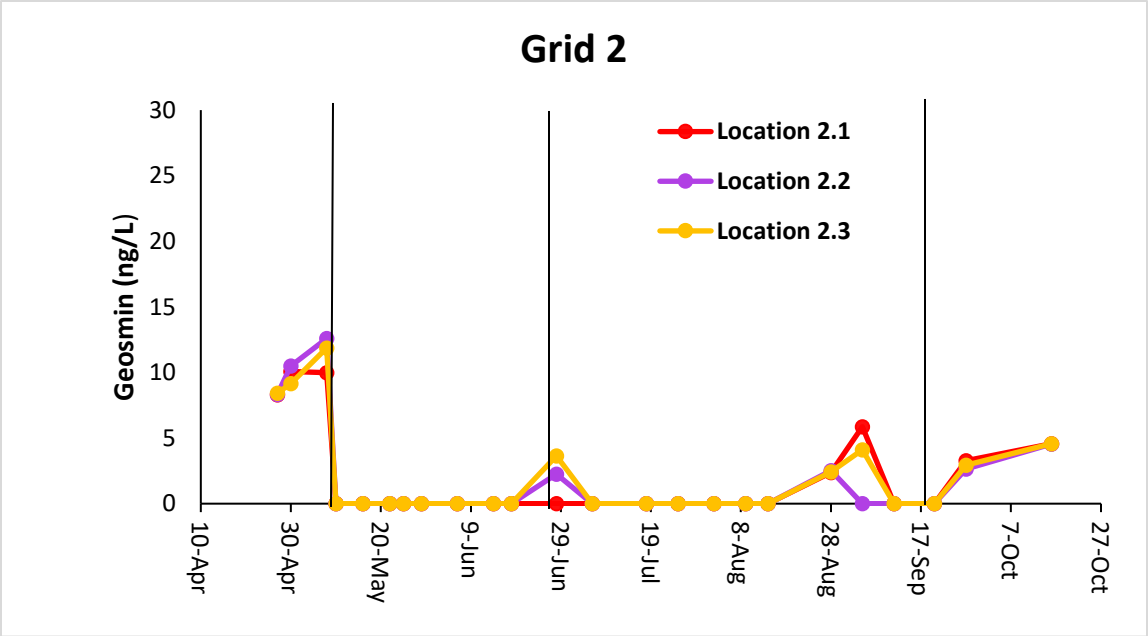


Figure 3.4 continued

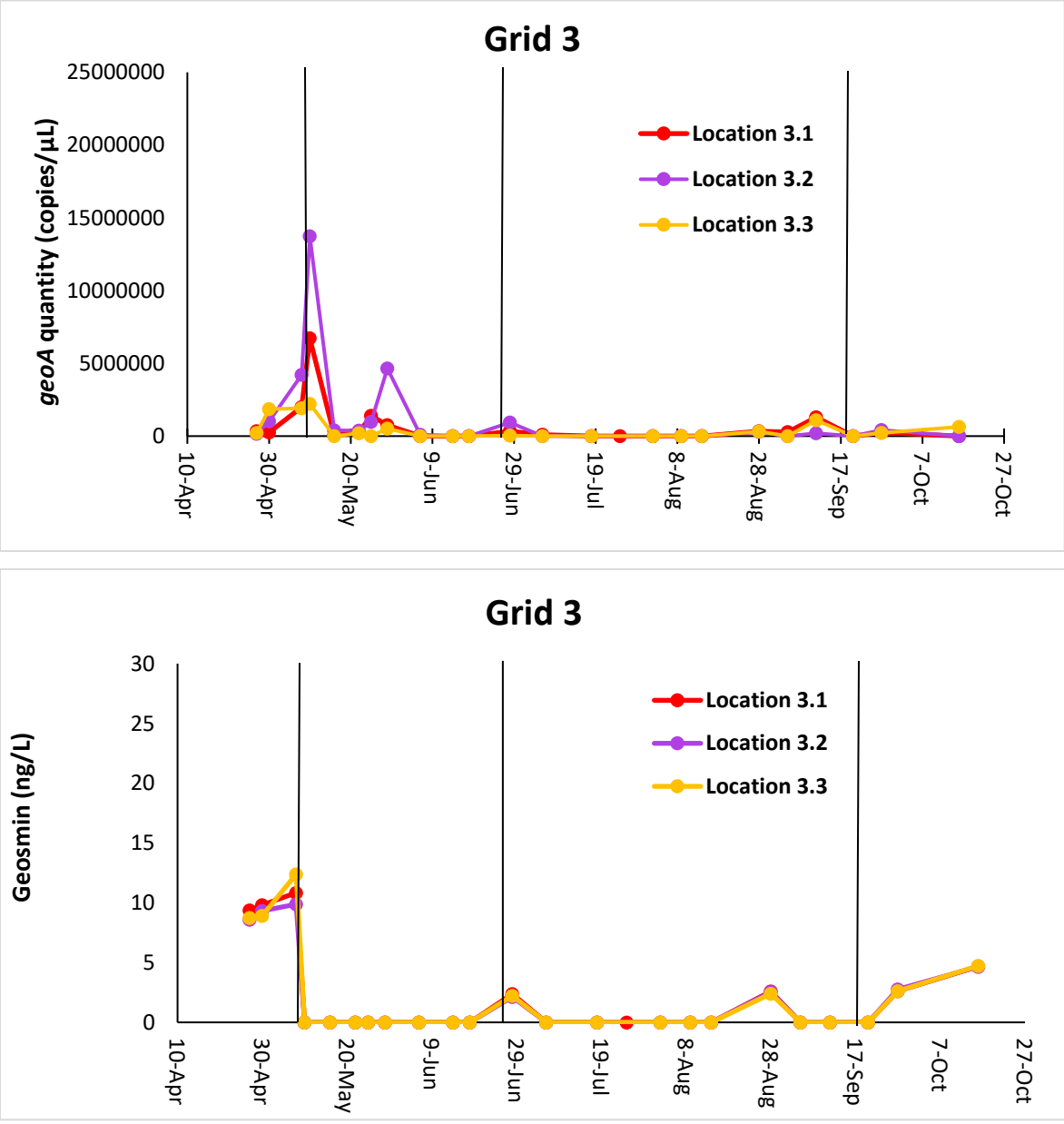


Figure 3.4 continued

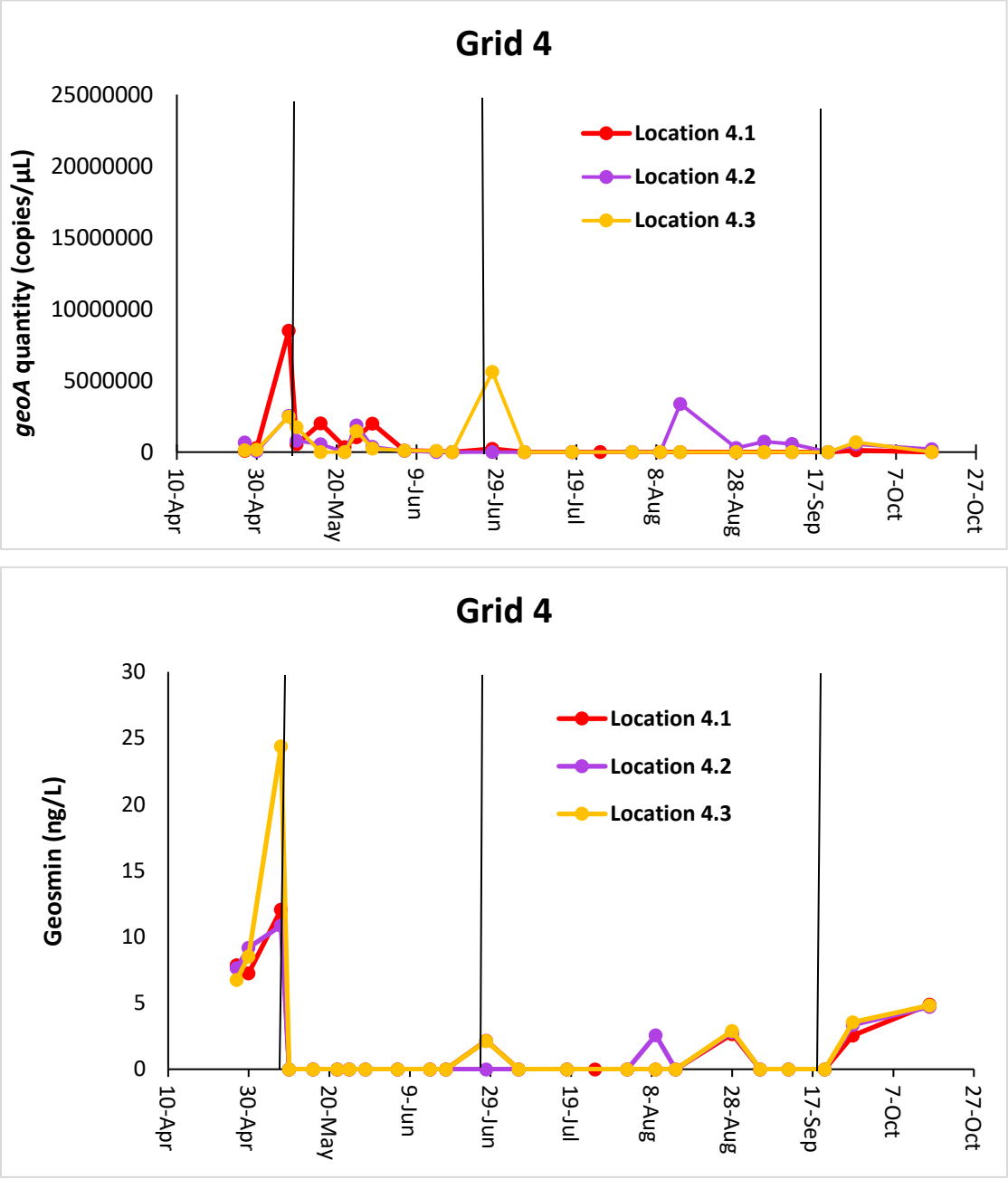
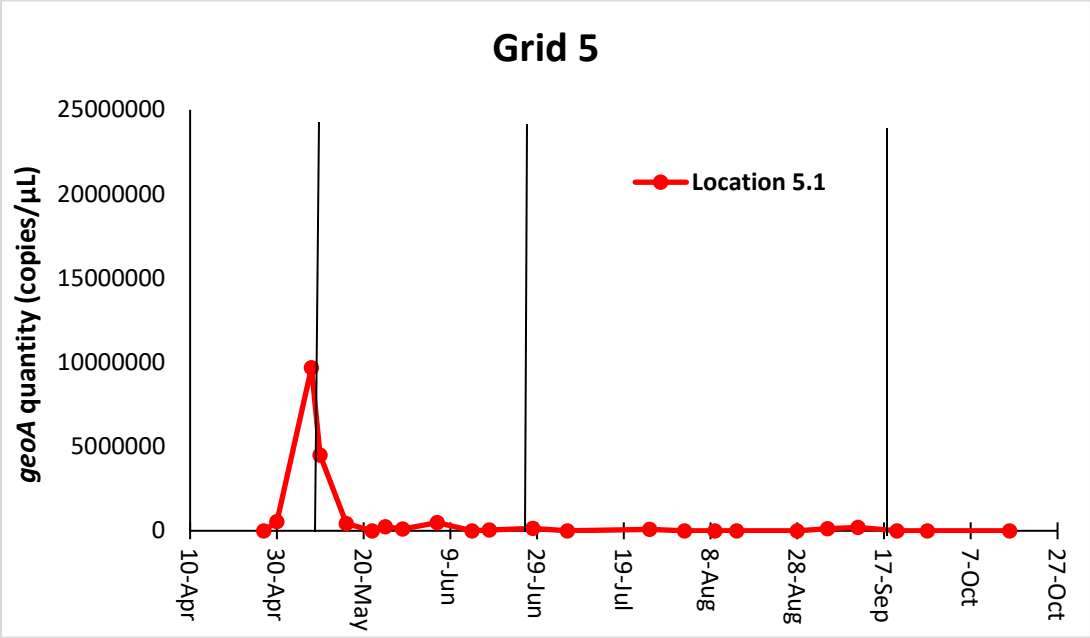


Figure 3.4 continued



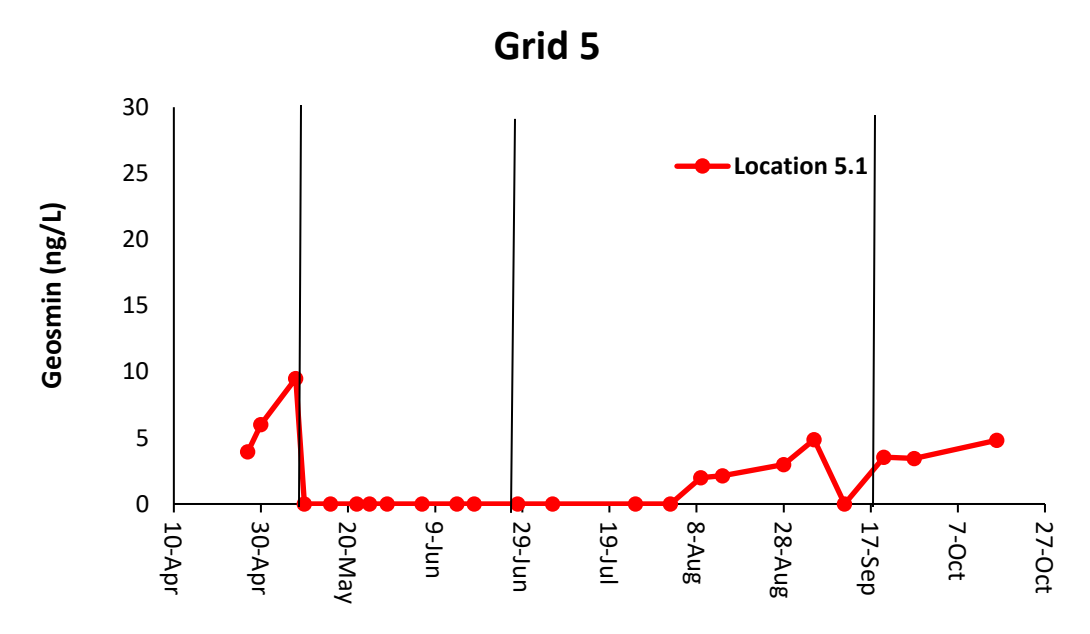
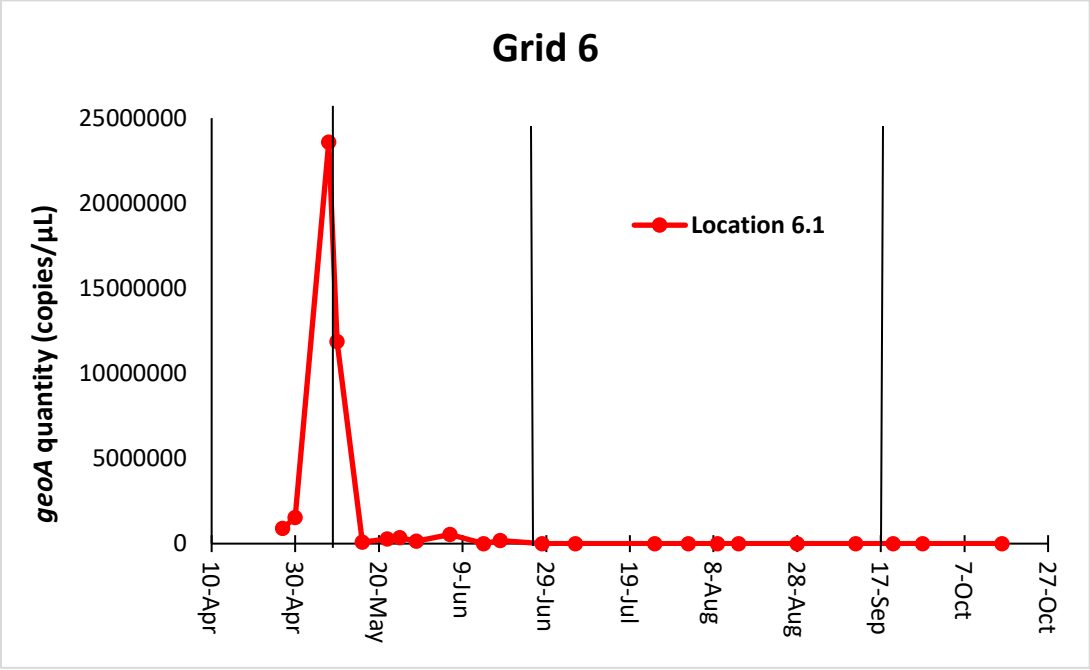


Figure 3.4 continued



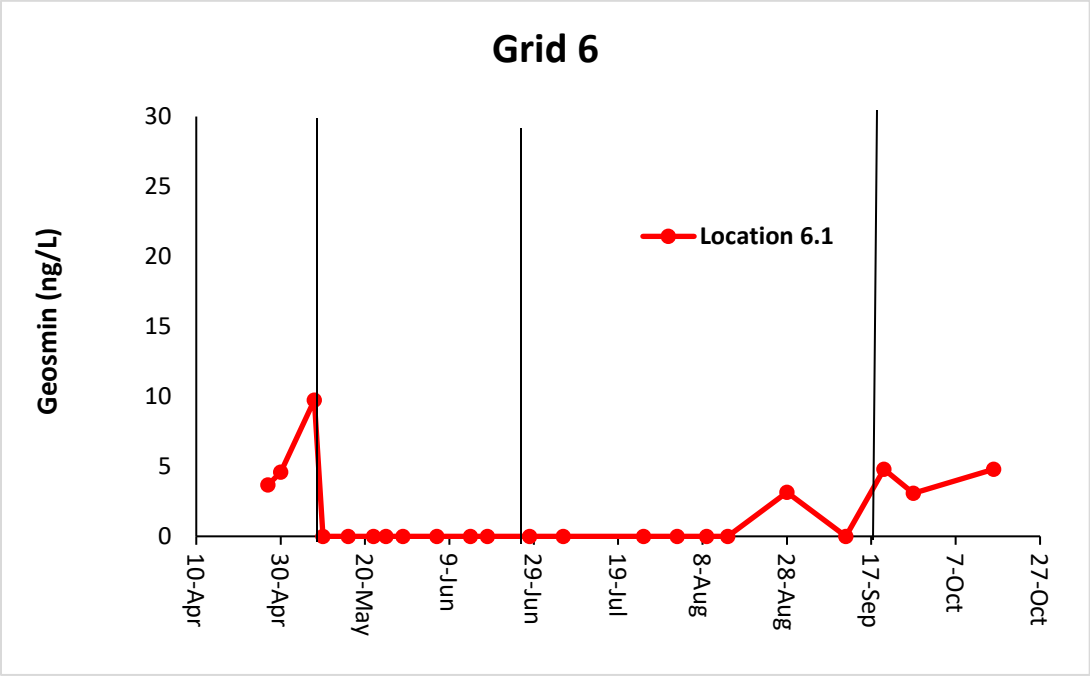


Table 3.4 Correlation analysis results between geosmin and *geoA* gene quantity.

Location	Spearman rho	*p*-value
1.1	0.248648	0.2526
1.2	0.3922697	0.06412
1.3	0.3314672	0.1318
2.1	0.0528011	0.8202
2.2	0.3089066	0.1515
2.3	0.1497546	0.4849
3.1	0.322475	0.1243
3.2	0.3621898	0.08944
3.3	**0.701068**	**0.0003995**
4.1	0.226925	0.3098
4.2	0.3071725	0.1877
4.3	**0.5098368**	**0.02165**
5.1	-0.04282962	0.8499
6.1	0.1428875	0.5367

Looking at MIB concentrations for the first sampling location of each grid (Figure 3.5), it is noticeable that major peaks occurred during May and September. MIB detections show minima during the Summer months when the reservoir stratifies. Maximal value of MIB was observed in May 16 for location 6.1 (Figure 3.6), which was 631.85 ng/L and exceeded the odor threshold of 10 ng/L [*Clercin,* 2018]. Time series plots (Figure 3.6) for the *MIBS* gene, detected by the qPCR assay, show that the major peak occurred during May, with location 6.1 having the highest gene concentration of 157,427.5 (± 54,553.9 copies/μL) among all sampling points on May 8th. As mentioned above, algaecide treatments to terminate T&O production were applied on May 9th, June 26th and September 18th. On May 10th which was the following sample collection date after the first treatment, MIB concentration levels started decreasing for locations in grid 1, 2, 3 and 4 but without immediately reaching levels below the odor threshold. In contrast, the MIB concentration for locations 5.1 and 6.1 kept increasing even after the treatment, reaching a maximum on May 16th and started drastically decreasing after that. When the second algaecide treatment was applied, MIB concentrations were already at low levels across all locations, implying that geosmin might have been the main target. After the third treatment, MIB concentration did not show an immediate response, since on September 20th it peaked across all locations. Looking at the *MIBS* gene levels and their response to algaecide treatments, a gradual

decrease started following the application of the first one for most locations. The subsequent treatments did not seem to influence *MIBS* levels detected by our qPCR assay, since they were already low or below detection levels. It is worth noting that although *MIBS* gene concentrations were close to zero during September, MIB concentrations showed some minor peaks during this time. This highlights the possible existence of other bacterial MIB producers that contribute to the problem and are not detectable by this qPCR assay. In order to evaluate the association between the *MIBS* gene abundance and the corresponding MIB concentration, Spearman's rho correlation analysis was performed. Results (Table 3.5) indicated a statistically significant positive correlation between the two variables for all of our sampling locations.

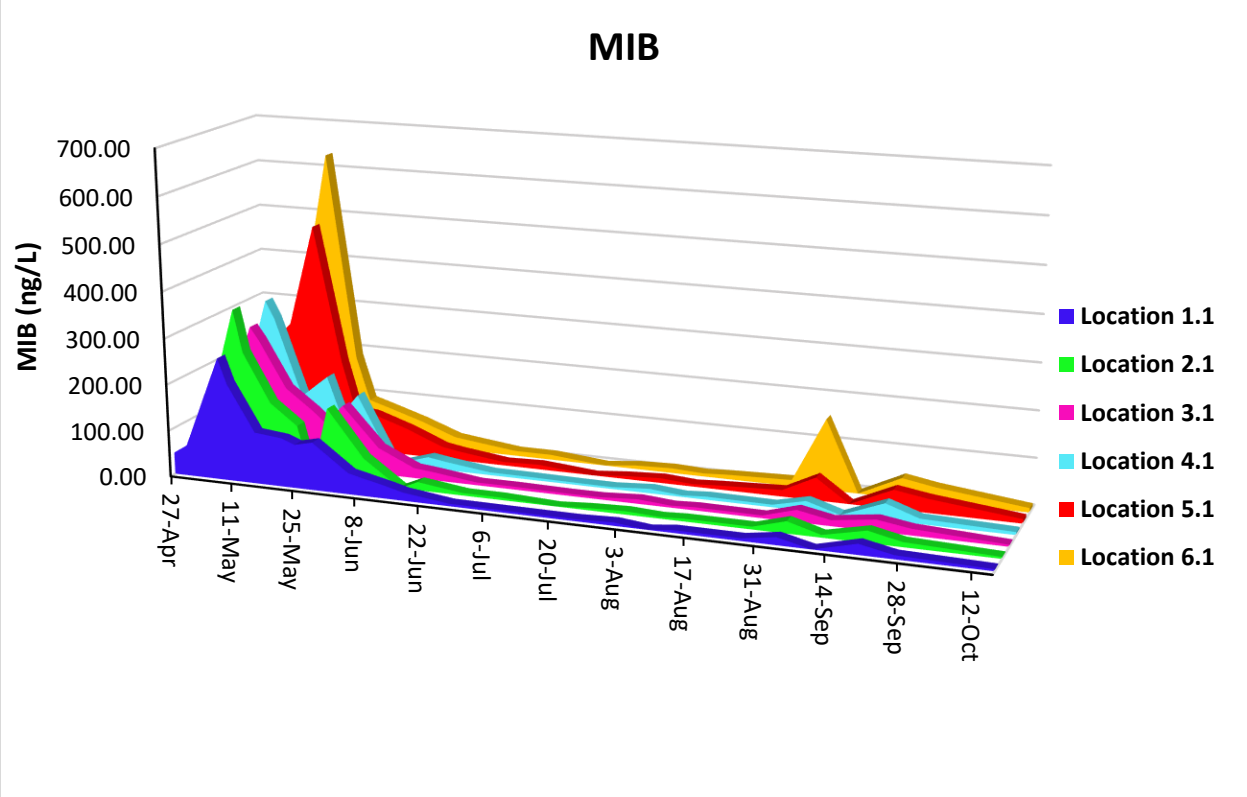


Figure 3.5 MIB concentration over time.

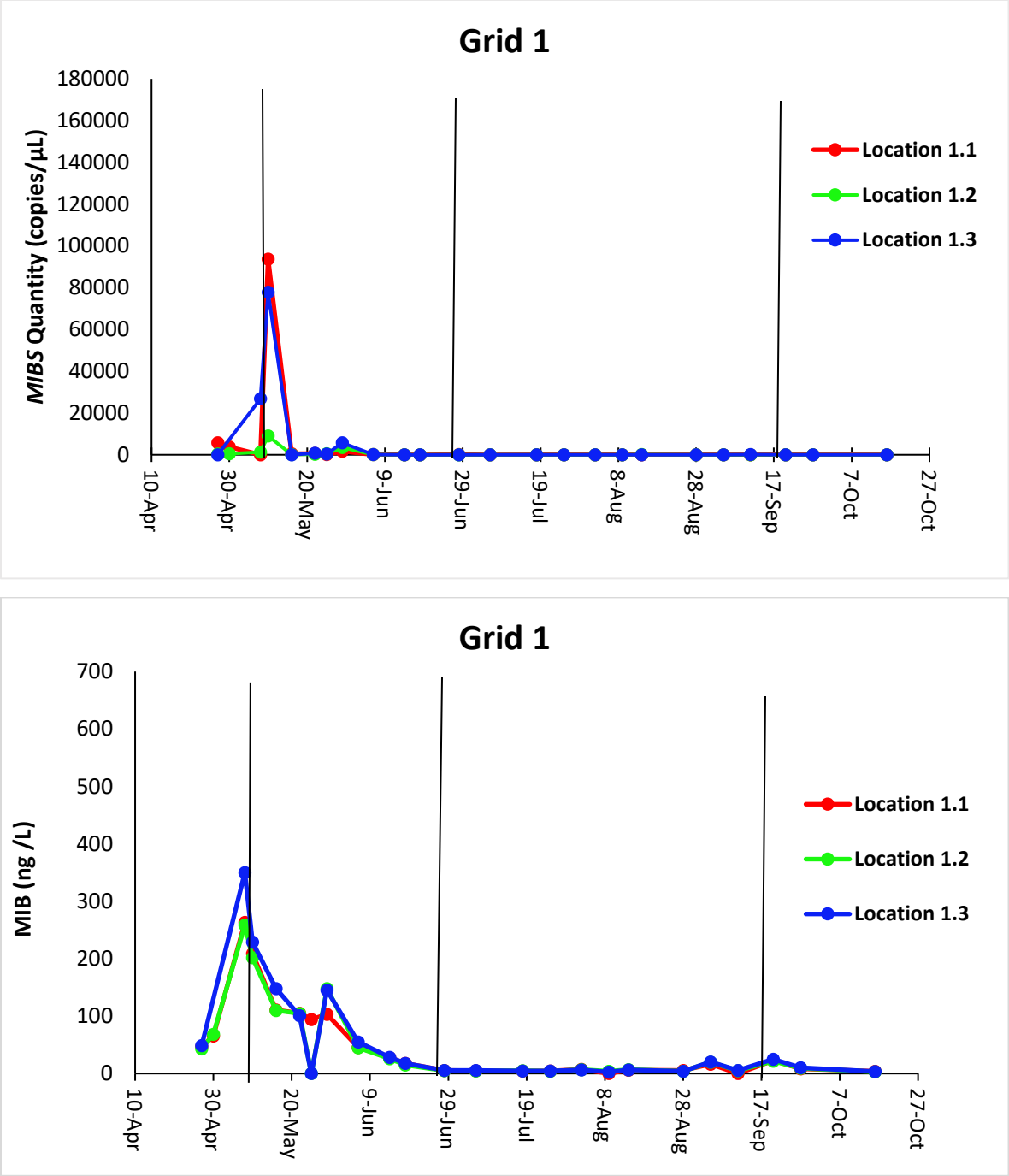
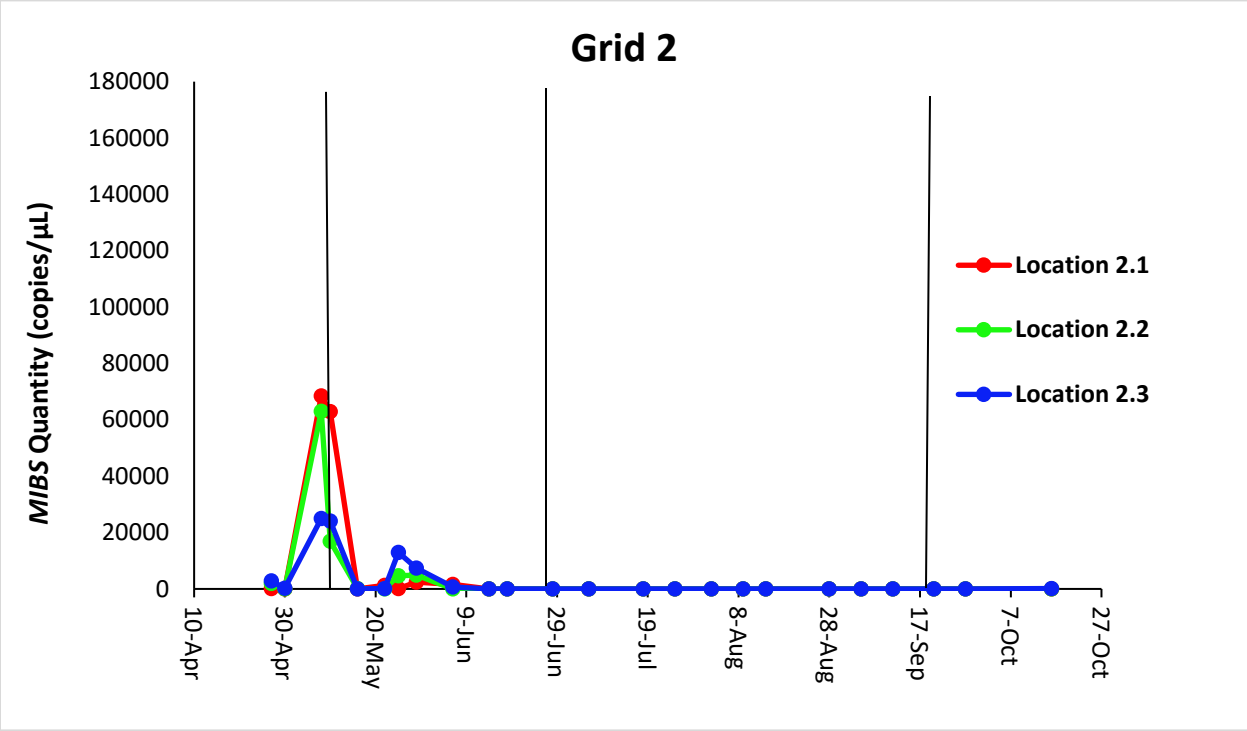


Figure 3.6 Time series graphs of MIB concentration and *MIBS* gene quantity for the sampling locations within each grid. Vertical black lines indicate algaecide treatment dates.

Figure 3.6 continued



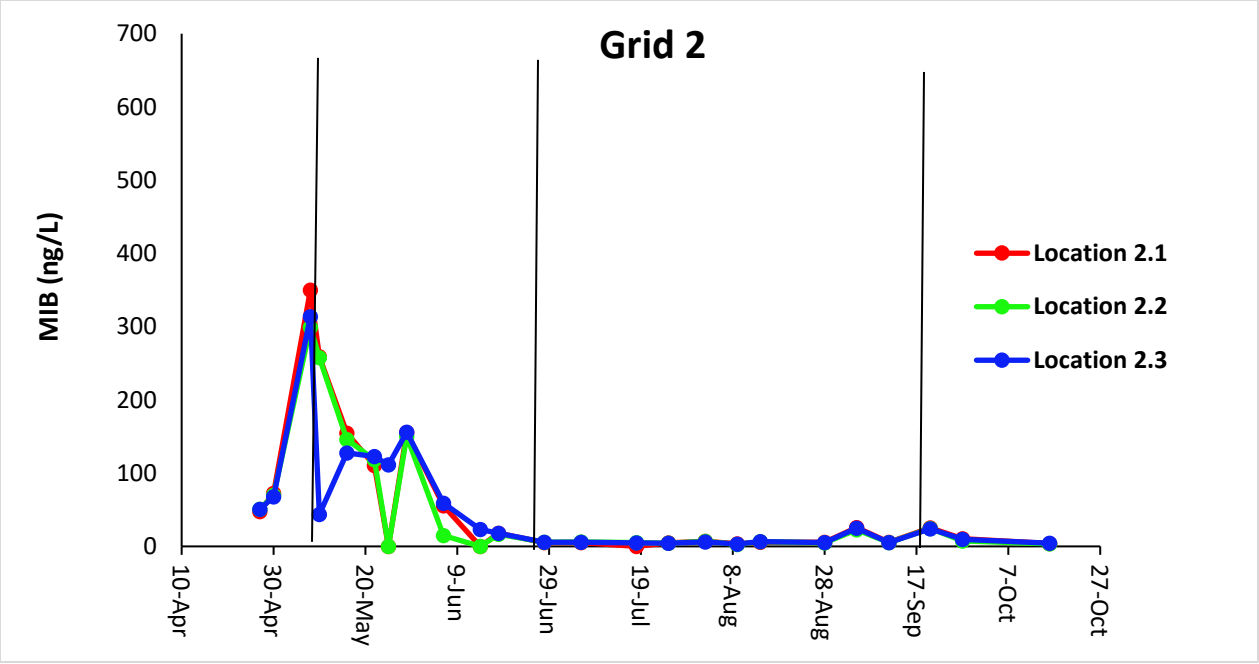
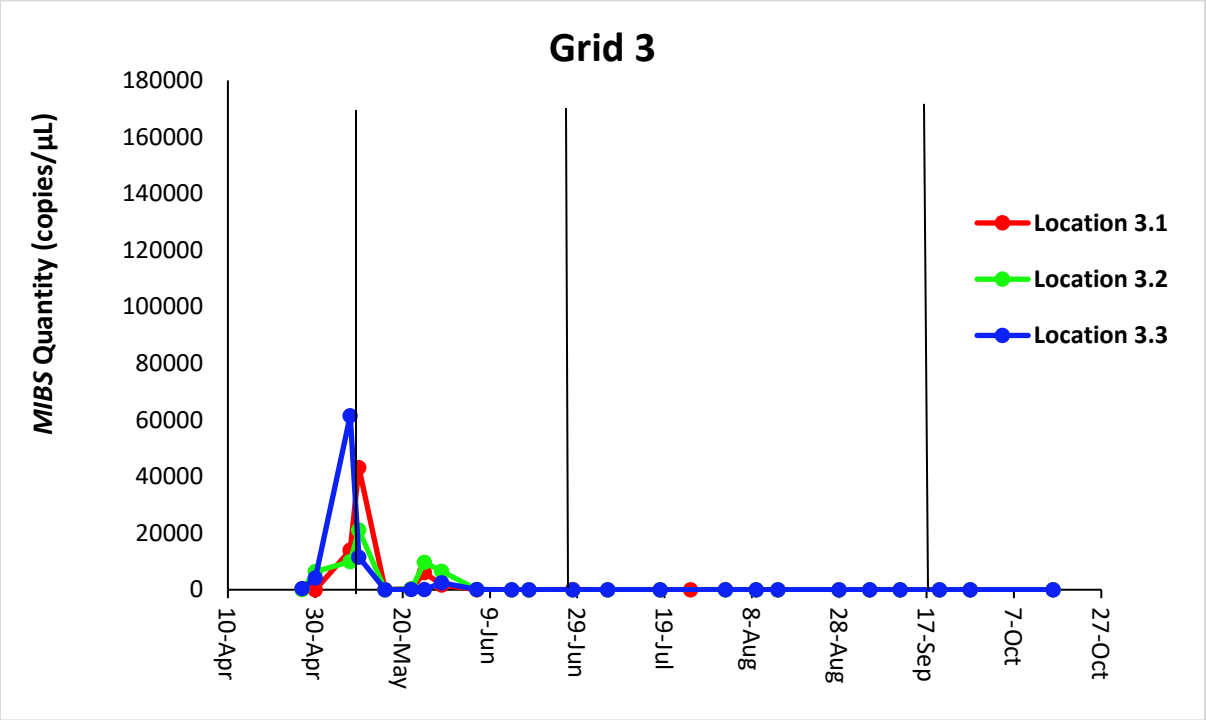


Figure 3.6 continued



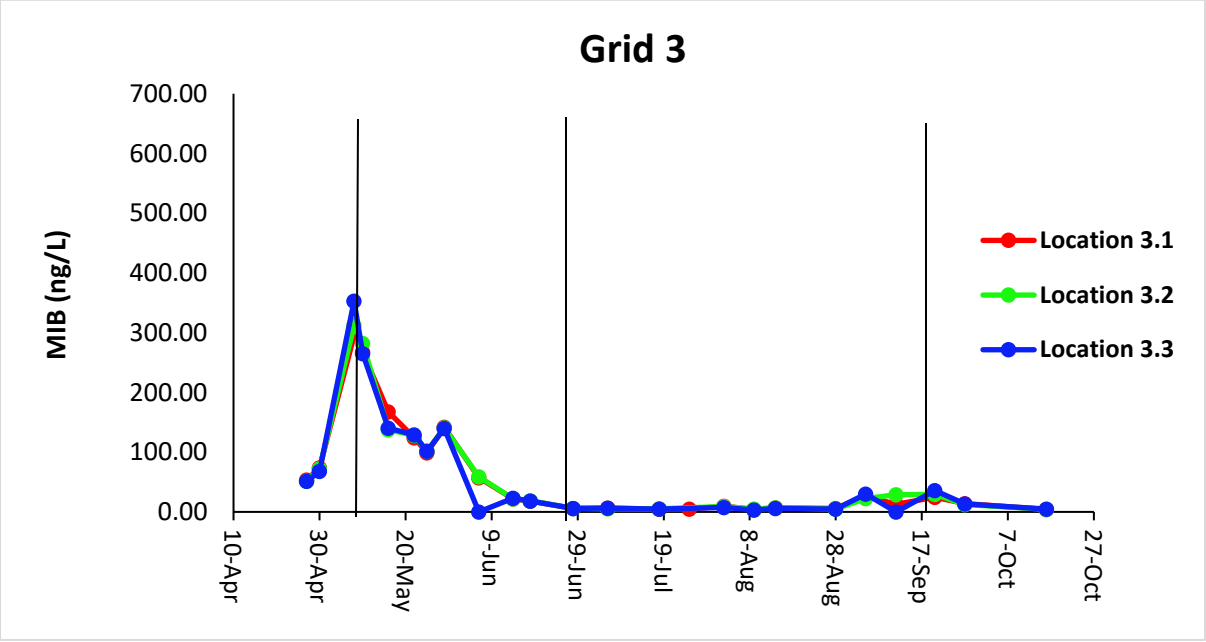
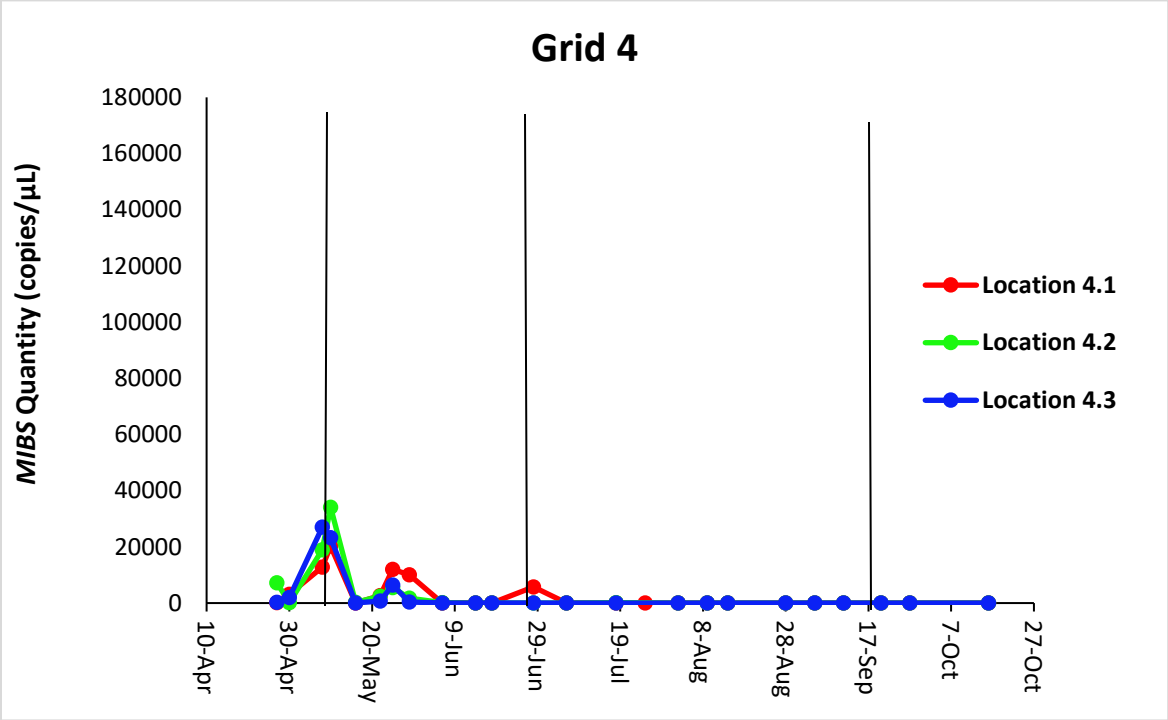


Figure 3.6 continued



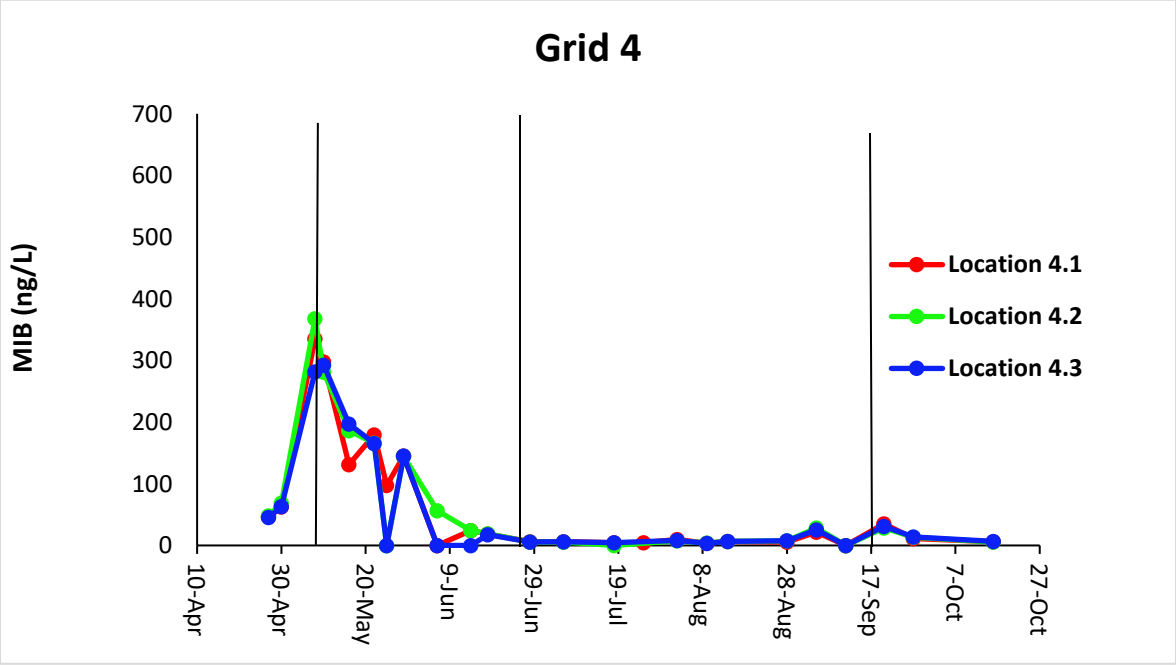
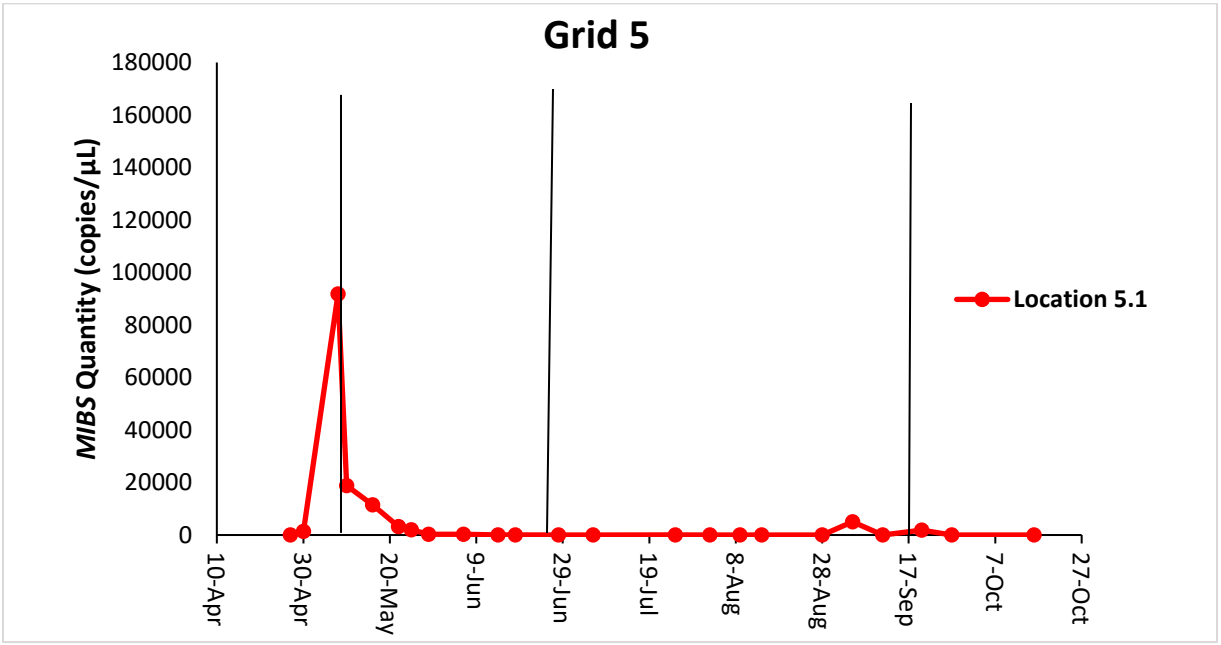


Figure 3.6 continued



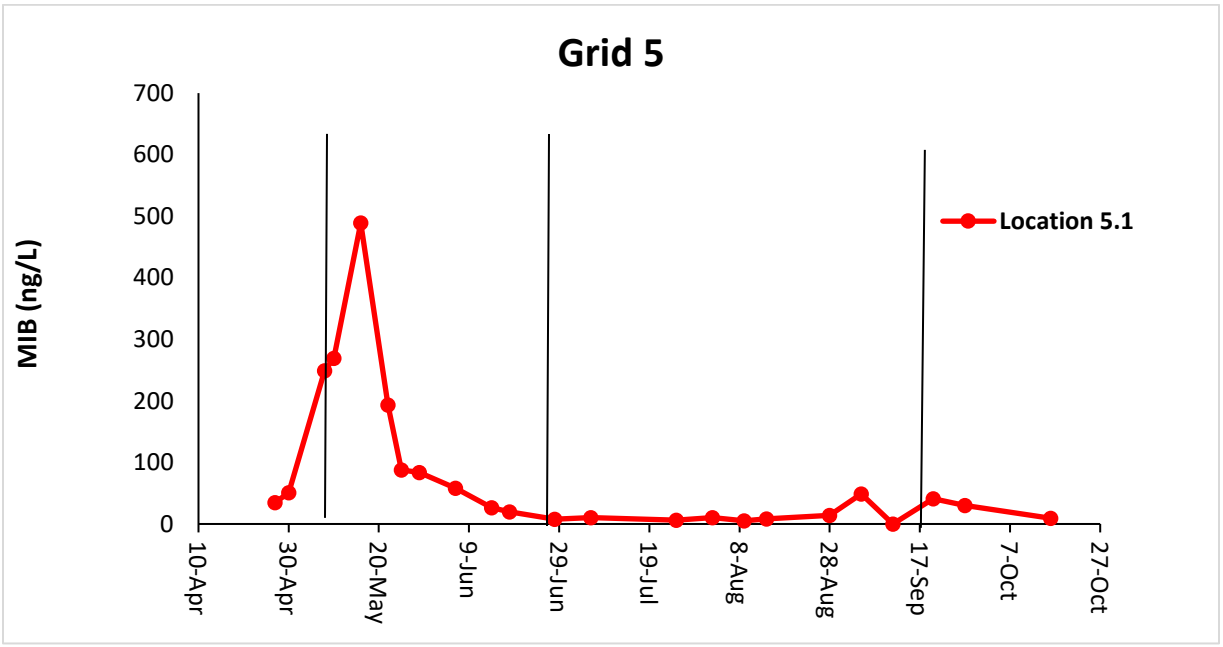
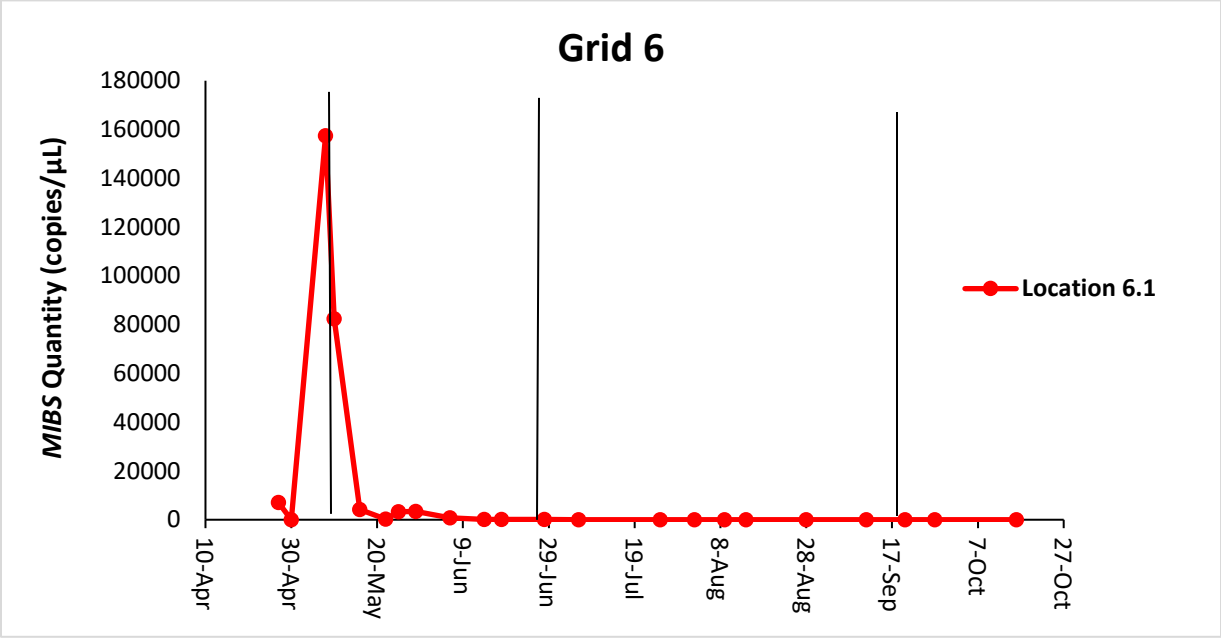


Figure 3.6 continued



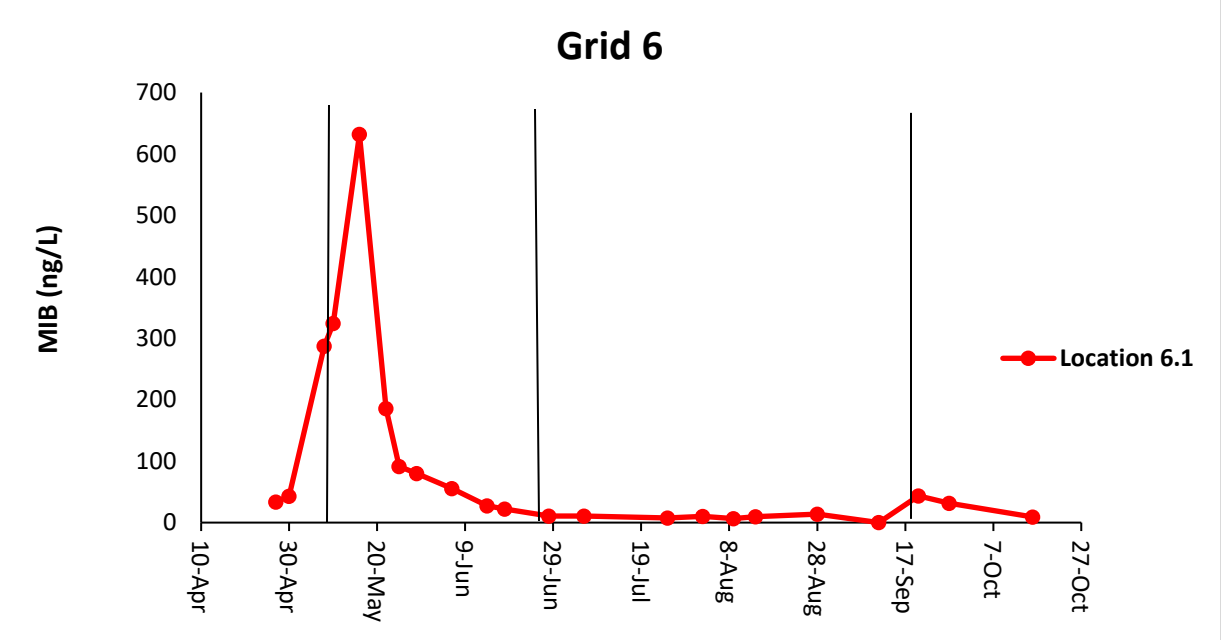
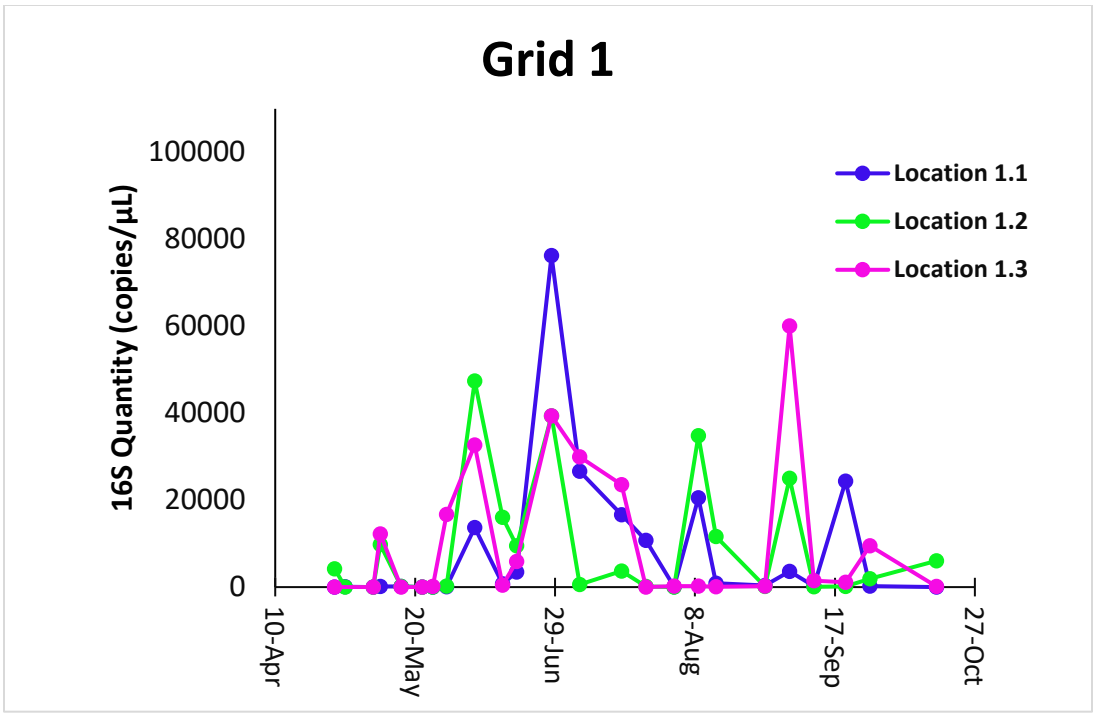


Table 3.5 Correlation analysis results between MIB and *MIBS* gene quantity.

Location	Spearman rho	p-value
1.1	0.6222153	0.001523
1.2	0.8131144	2.41E-06
1.3	0.6261638	0.001824
2.1	0.7991441	1.39E-05
2.2	0.8102751	4.86E-06
2.3	0.7117924	9.59E-05
3.1	0.6818296	0.0002433
3.2	0.727774	8.29E-05
3.3	0.7488591	9.39E-05
4.1	0.7330964	0.000104
4.2	0.8526532	1.80E-06
4.3	0.7901149	5.70E-05
5.1	0.8539246	4.31E-07
6.1	0.8033632	1.16E-05

In order to measure total bacterial density and evaluate its possible relationship with fluctuations in MIB and geosmin concentrations, a qPCR assay was used, targeting the 16S rRNA gene. Compared to the *geoA* and *MIBS* results, where major peaks occurred during May and September, the 16S rRNA gene quantities (Figure 3.7), had peaks occurring throughout the sampling season. Spearman's correlation coefficients were used to evaluate the association between the 16S rRNA gene quantity and T&O concentrations. Results (Table 3.6) indicated that a statistically significant negative correlation between 16SrRNA quantity and MIB concentration occurred only for locations 1.1, 2.2, 3.1, and 4.3. A statistically significant negative correlation between the 16SrRNA gene quantities and geosmin concentrations occurred only for locations 4.3 and 5.1 (Table 3.7).

Since cell counting techniques are still widely used by water utilities to monitor T&O episodes, we wanted to assess the relationship between total cell counts and geosmin and MIB concentrations [*Jüttner and Watson,* 2007]. As expected by the fact that cell counting techniques do not take into consideration the synthetic capabilities of microorganisms, there were no statistically significant correlation between T&O concentrations and total cell counts (Table 3.8 & Table 3.9). This highlights that techniques that are based on the estimation of total bacterial densities are not the most appropriate tools in monitoring and predicting upcoming T&O problems.



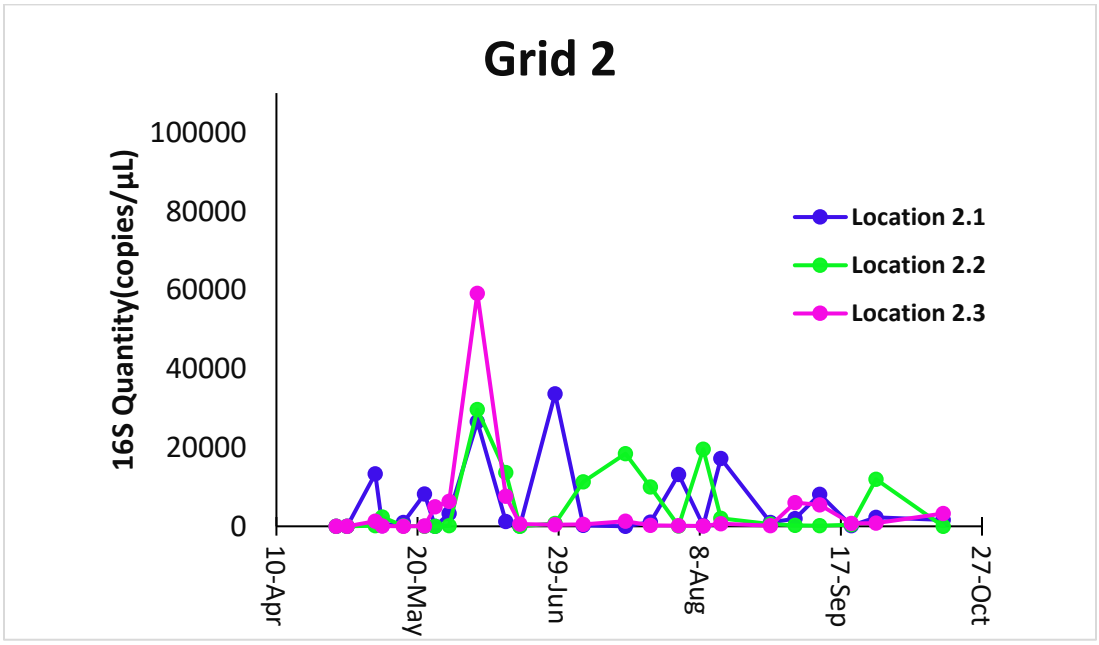


Figure 3.7 Time series plots of the 16S rRNA gene quantities measured by qPCR across all locations within each grid.

Figure 3.7 continued

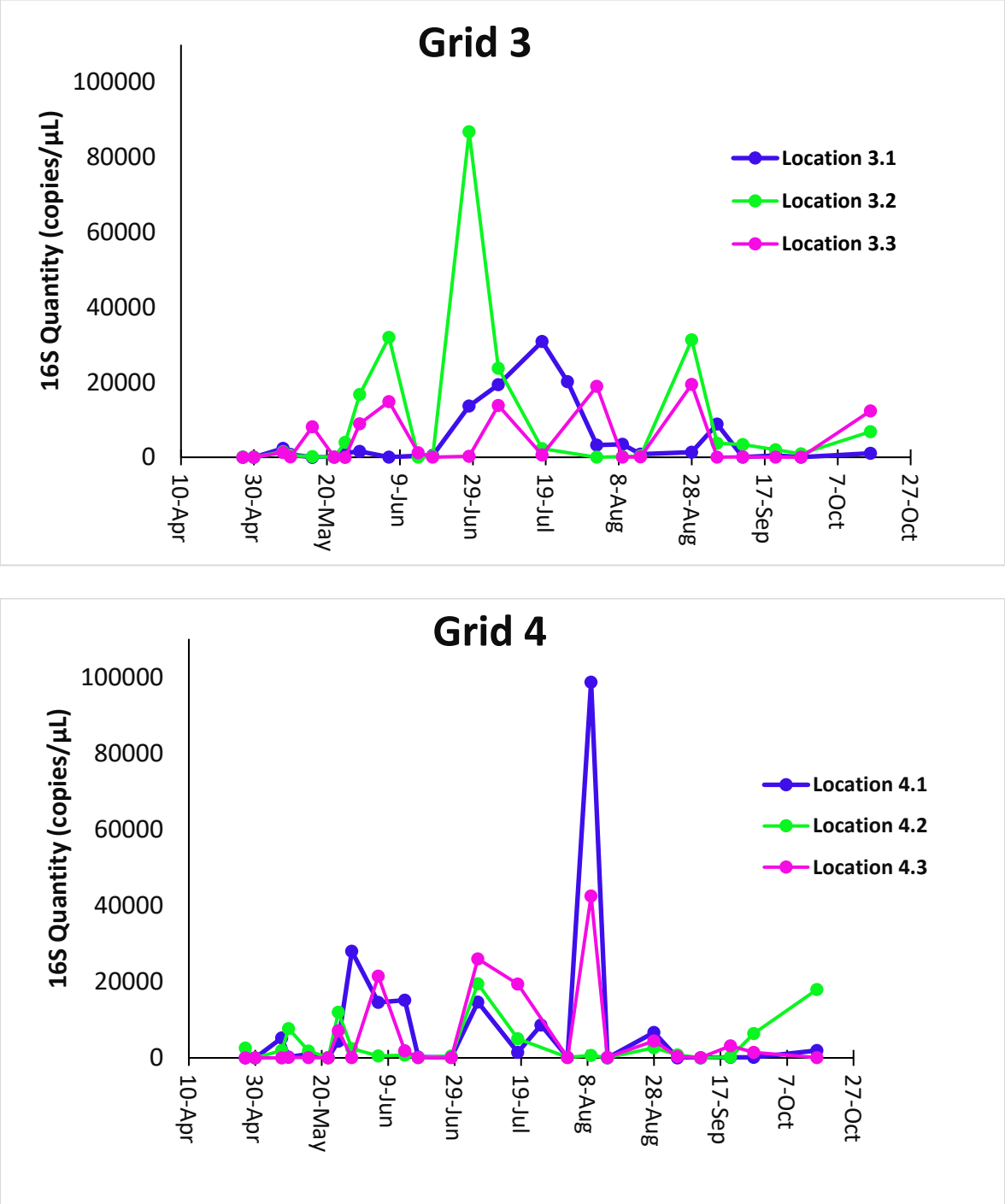


Figure 3.7 continued

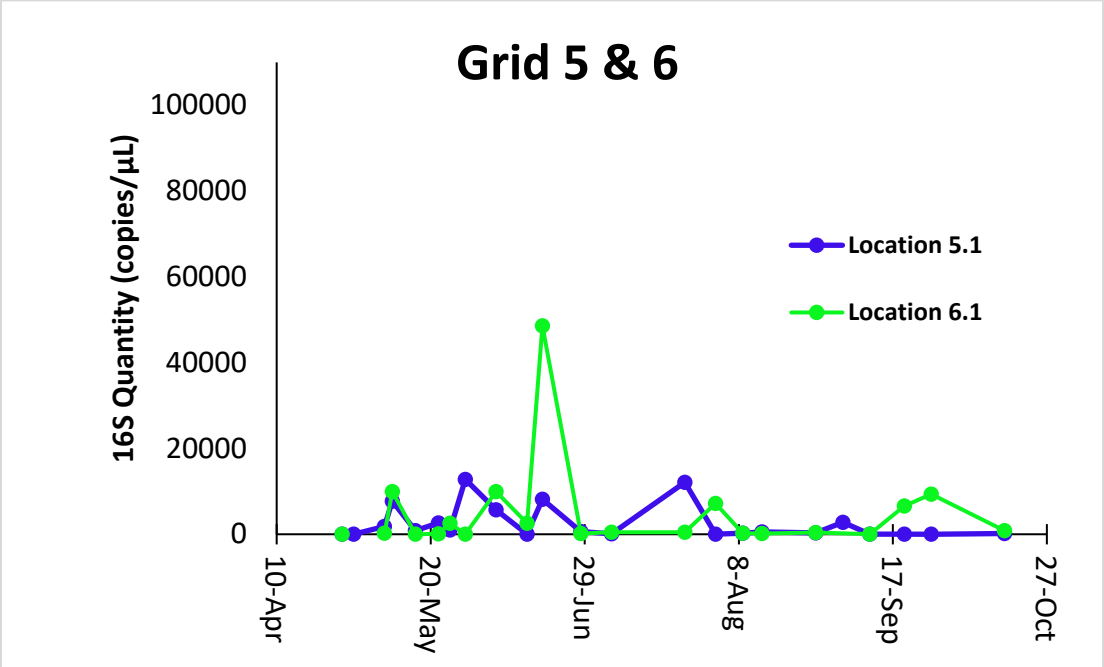


Table 3.6 Correlation analysis results between 16S rRNA gene and MIB concentration.

Location	Spearman rho	p-value
1.1	**-0.4990119**	**0.01648**
1.2	-0.2262846	0.2977
1.3	-0.09768492	0.6645
2.1	-0.02077922	0.9303
2.2	**-0.445072**	**0.03793**
2.3	0.0152207	0.9437
3.1	**-0.5148946**	**0.01004**
3.2	-0.2005929	0.3571
3.3	-0.2636364	0.2471
4.1	-0.2546584	0.2516
4.2	-0.0481203	0.8412
4.3	**-0.6157895**	**0.005975**
5.1	0.2626377	0.2377
6.1	-0.1025974	0.6573

Table 3.7 Correlation analysis results between 16S rRNA gene and geosmin concentration.

Location	Spearman rho	p-value
1.1	-0.4086533	0.05286
1.2	-0.202016	0.3553
1.3	-0.2388452	0.2844
2.1	-0.2111968	0.3581
2.2	-0.3550825	0.09638
2.3	-0.09560253	0.6568
3.1	-0.2003158	0.348
3.2	0.005456795	0.9803
3.3	-0.009283378	0.9681
4.1	-0.1809525	0.4203
4.2	0.2558489	0.2763
4.3	**-0.4869952**	**0.02943**
5.1	**-0.4367002**	**0.04215**
6.1	-0.1276464	0.5814

Table 3.8 Correlation analysis results between total cell counts and geosmin concentration.

Location	Spearman rho	p-value
1.1	0.03557389	0.8816
1.2	0.1155376	0.6376
1.3	0.1778694	0.4531
2.1	0.04571182	0.8571
2.2	0.102769	0.6664
2.3	0.2790208	0.2335
3.1	0.1228351	0.6164
3.2	0.2937021	0.1963
3.3	-0.1345233	0.5946
4.1	0.2826148	0.2273
4.2	0.4219671	0.1035
4.3	0.2976703	0.2812
5.1	-0.1231587	0.6495
6.1	0.08819521	0.7364

Table 3.9 Correlation analysis results between total cell counts and MIB concentration.

Location	Spearman rho	p-value
1.1	0.4015038	0.08042
1.2	0.09122807	0.7101
1.3	0.2421053	0.3024
2.1	0.2714138	0.2749
2.2	0.1849624	0.4333
2.3	0.01203008	0.962
3.1	0.3194384	0.1825
3.2	-0.05974026	0.7973
3.3	0.2301342	0.3567
4.1	0.02255639	0.9265
4.2	0.1647059	0.5412
4.3	0.2571429	0.3538
5.1	0.2382353	0.3729
6.1	-0.04411765	0.8686

Spatial maps were constructed using the ArcMap software in order to better visualize the spatial variations of the 16S rRNA genes (Figure 3.8) and the T&O synthesis genes (Figure 3.9 & Figure 3.10). Based on the Kruskal-Wallis test that was performed (Table 3.10) to investigate if there was a location that had higher gene quantities compared to others during the time span of our sampling, there was no statistically significant difference between locations for *geoA, MIBS* and 16S rRNA gene quantities.

Table 3.10 Kruskal-Wallis test results.

	Kruskal-Wallis test	Conclusion
MIBS	chi-squared = 7.0291, df = 13, p-value = 0.9006	No statistically significant difference between locations
geoA	chi-squared = 5.9861, df = 13, p-value=0.9467	No statistically significant difference between locations
16S rRNA	chi-squared = 5.6415, df = 13, p-value=0.9582	No statistically significant difference between locations

In order to evaluate the change of gene quantities before and after the application of the algaecide treatment the Wilcoxon Signed Rank Test was performed, which is the non-parametric analogue to the paired t-test. Based on those results (Table 3.11), the null hypothesis which

assumes that the median difference between pairs of measured gene quantities before and after treatment is zero, is accepted for most of the target genes and treatment dates. The only exception is for the *geoA* gene quantity in respect to the second and third treatment, in which the null hypothesis was rejected based on the p-value. Regarding the second treatment, there was a statistically significant increase in the *geoA* quantity a few days after its application. In contrast, after the third treatment, there was a statistically significant decrease in the *geoA* quantity. Since there is not a consistent trend, specific conclusions cannot be drawn about the response of gene quantities to the algaecide treatment. The fact that sample collection was performed the exact day before and right after just for the first treatment could pose a limitation to the interpretation of the results. During the days between the treatment and sample collection, both biotic and abiotic factors could have interfered, impacting the gene levels and preventing an accurate evaluation of the pre and post algaecide behaviour.

Table 3.11 Wilcoxon Signed Rank test results for the gene quantities measured for the collection dates before and after the algaecide treatments.

	Wilcoxon Signed Rank Test		
Quantity (copies/µL)	**Pre/Post Treatment (May-9)**	**Pre/Post Treatment (June-26)**	**Pre/Post Treatment (September 18)**
16S	Z = -0.78471, p-value = 0.4326	Z = -1.7264, p-value = 0.08428	Z = -0.59638, p-value = 0.5509
MIBS	Z = -0.15694, p-value = 0.8753	Z = -0.44455, p-value = 0.6566	Z = -0.44721, p-value = 0.6547
geoA	Z = -0.47082, p-value = 0.6378	Z = -2.1339, **p-value = 0.03285**	Z = 2.0449, **p-value = 0.04086**

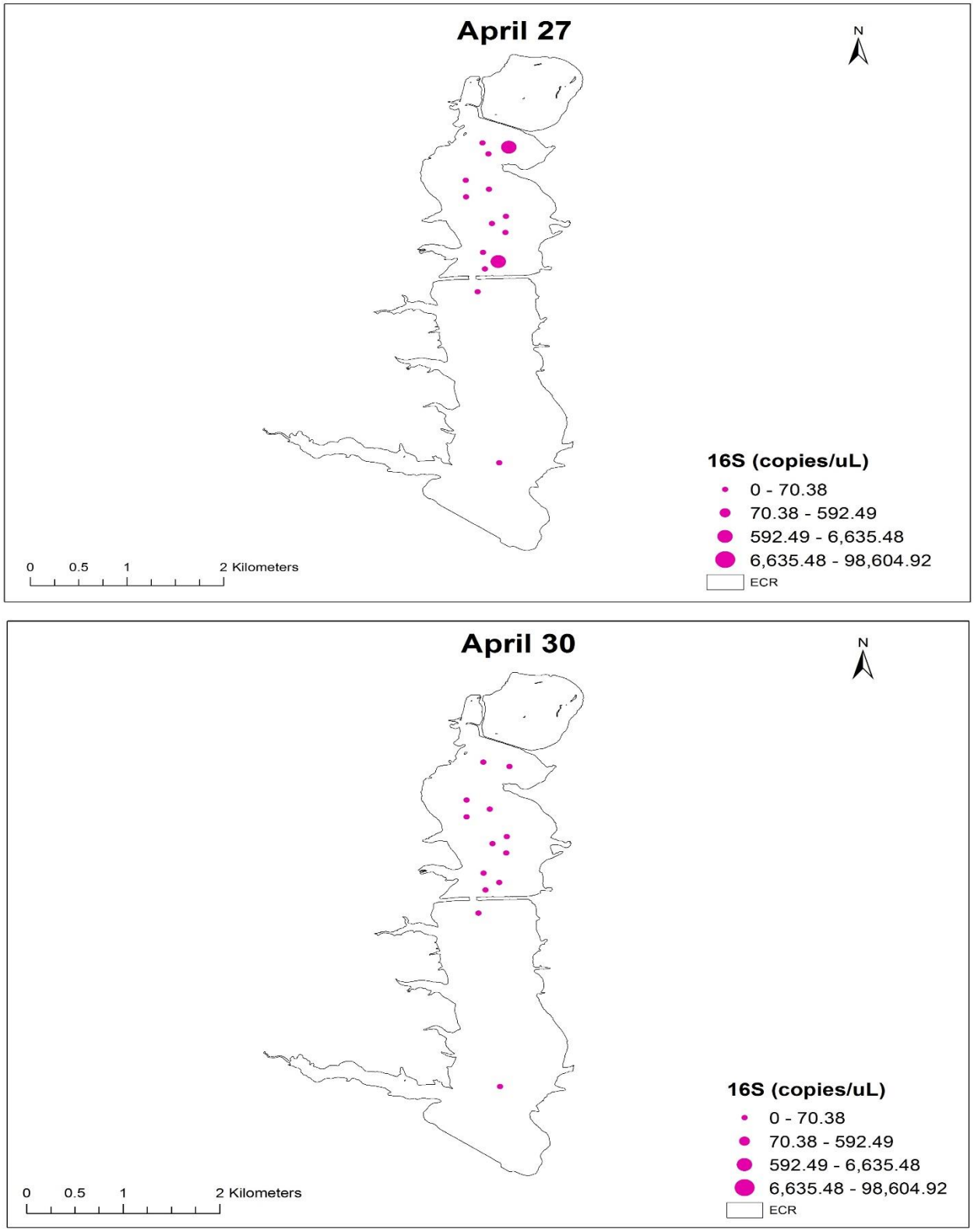
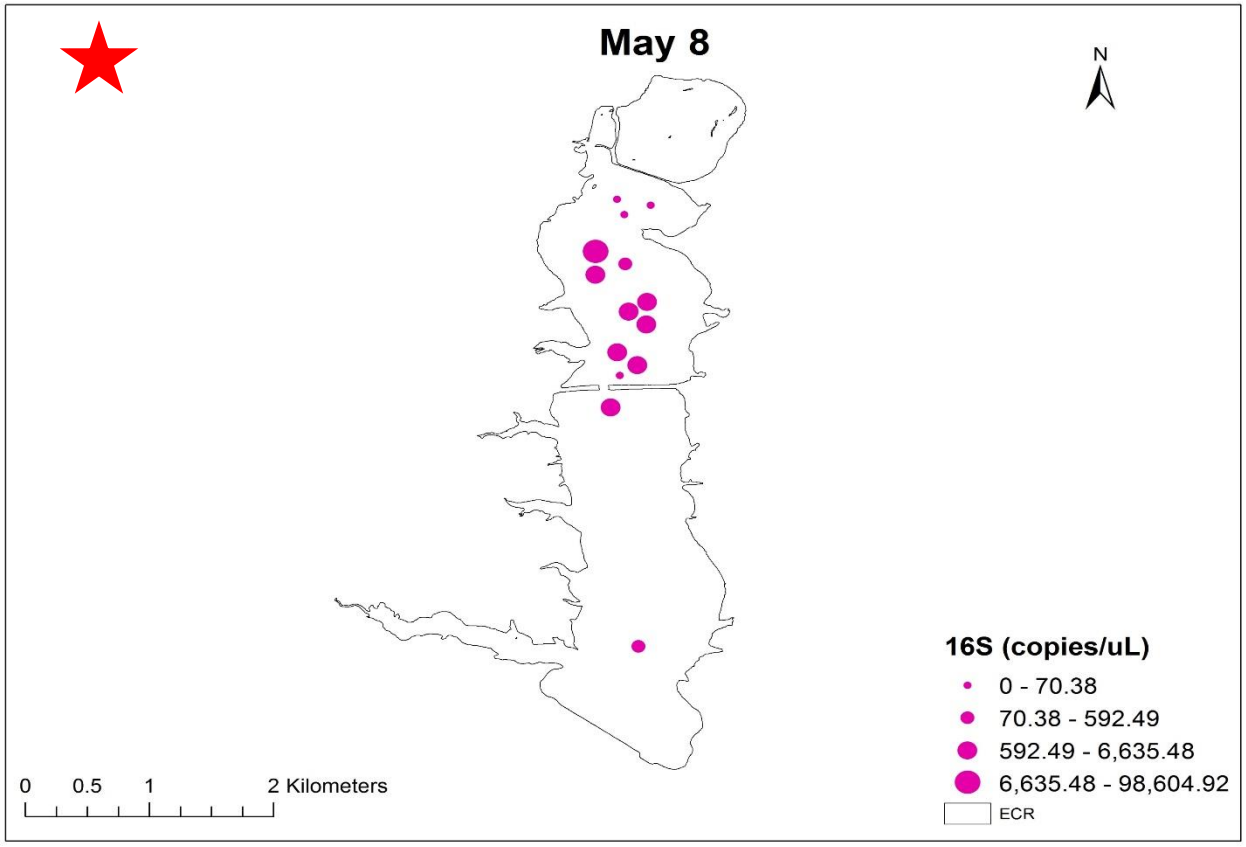


Figure 3.8 Spatial maps of 16S rRNA gene quantity for each sampling date. Data are sorted on a quantile basis. Red stars indicate sampling date before the algaecide treatment and blue stars indicate the sampling date after treatment.

Figure 3.8 continued



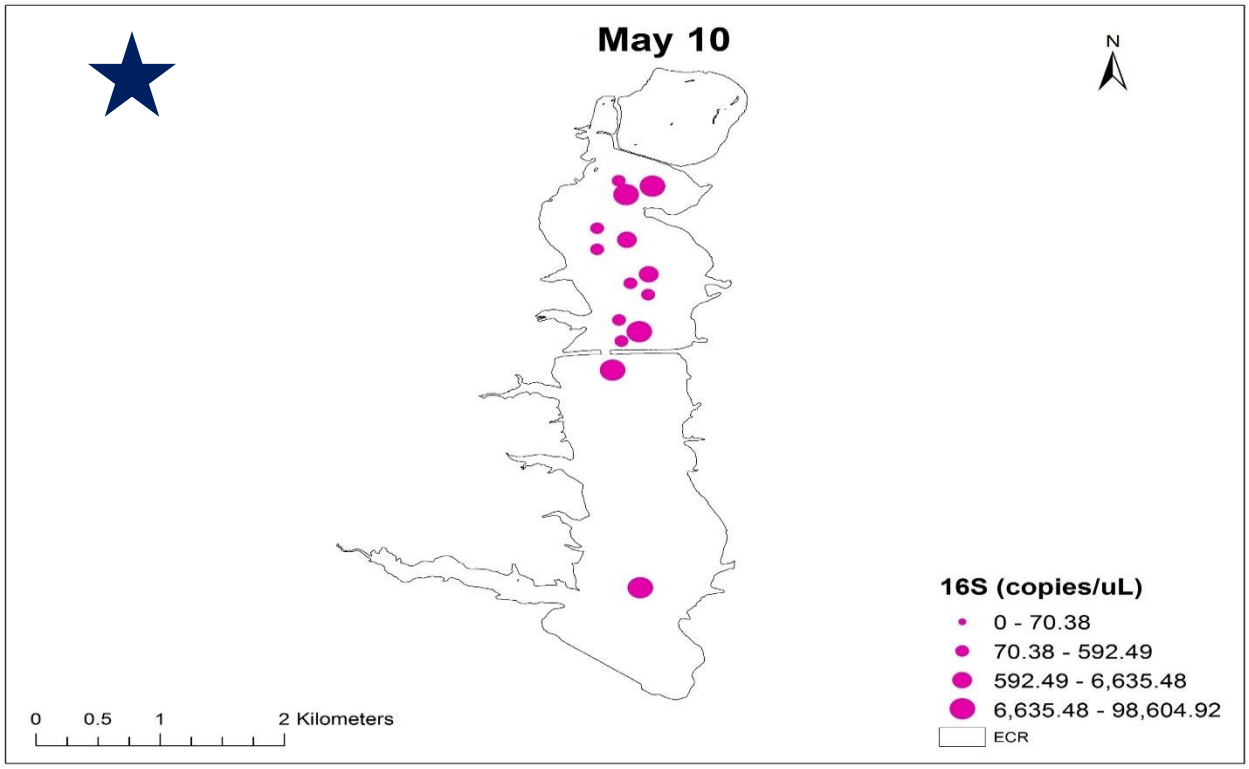
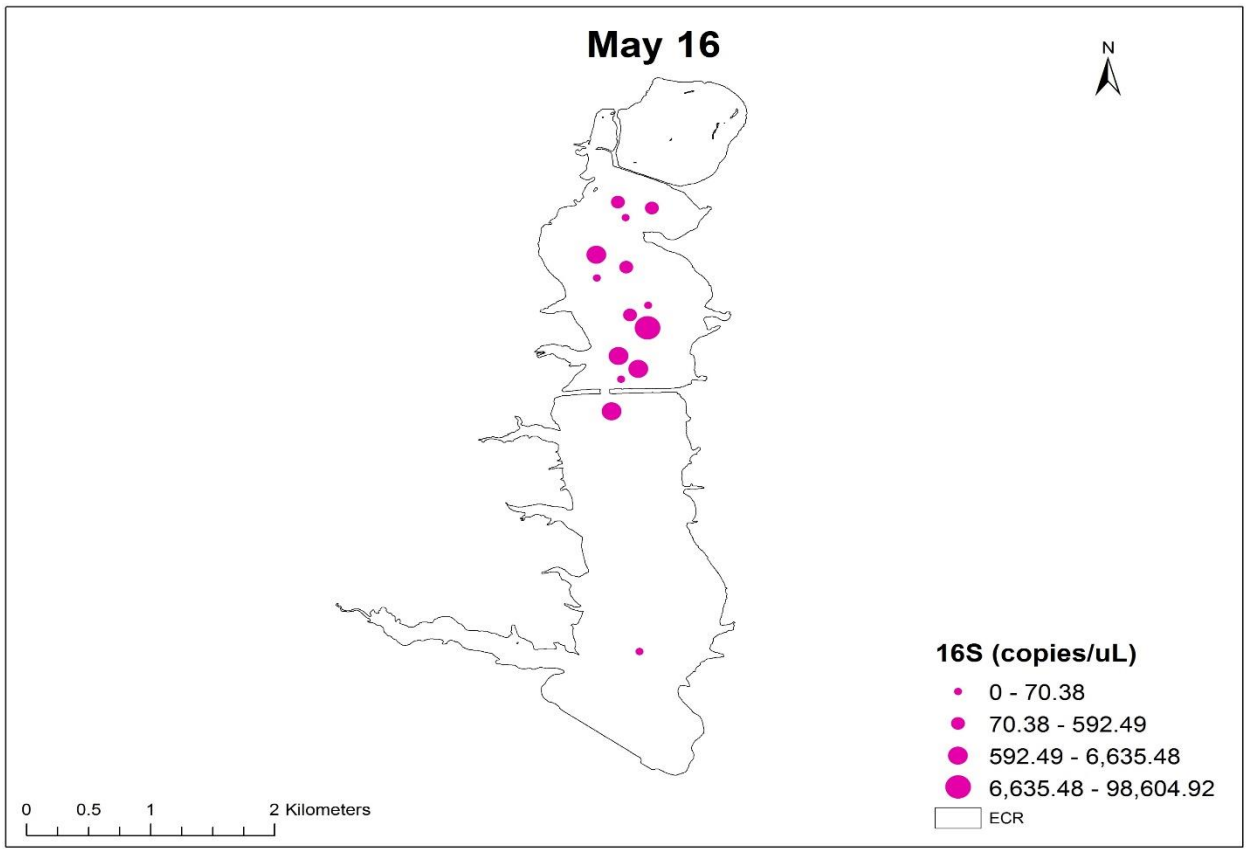


Figure 3.8 continued



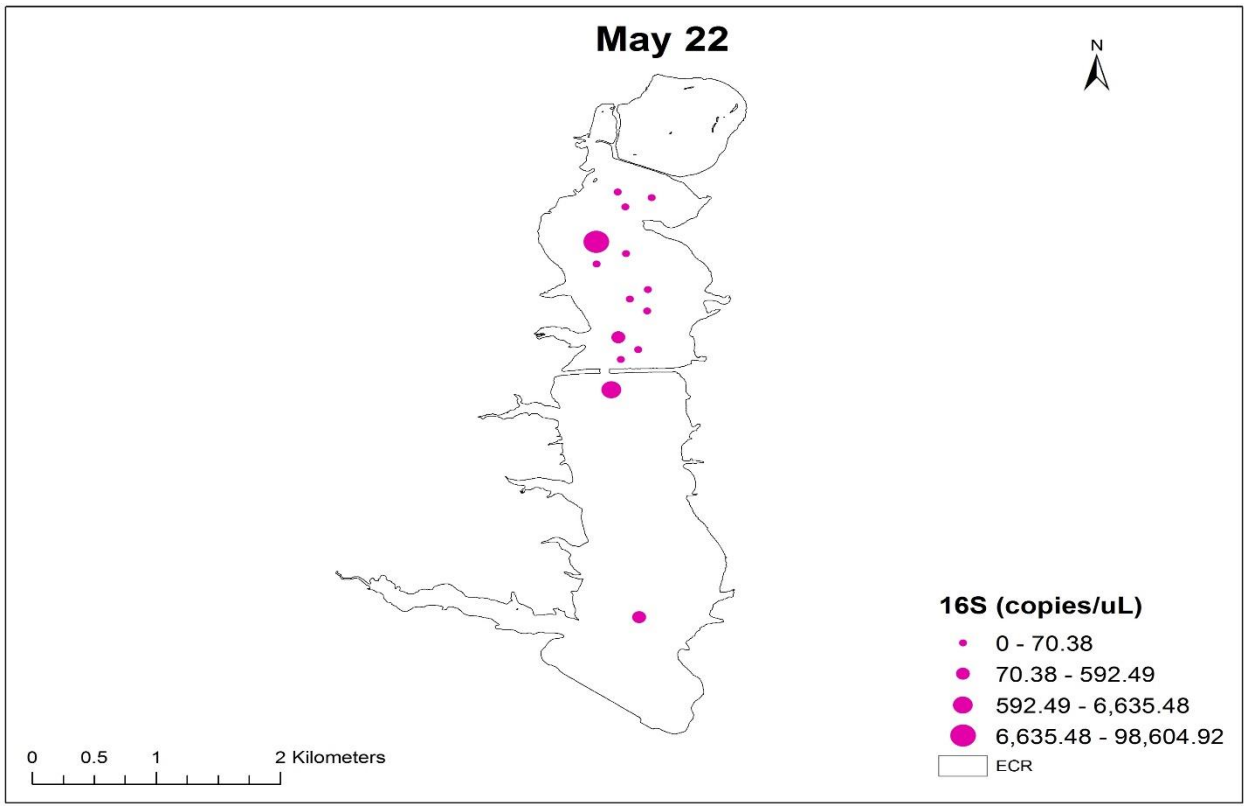
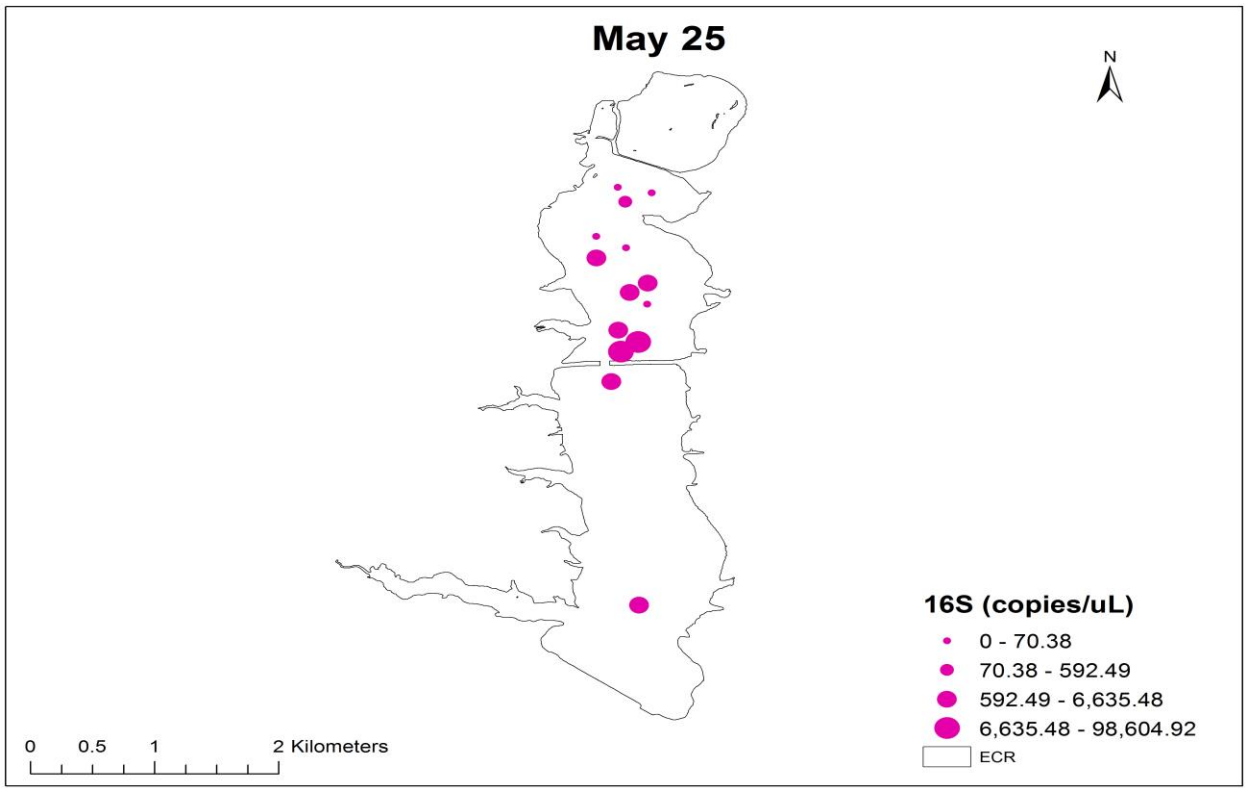


Figure 3.8 continued



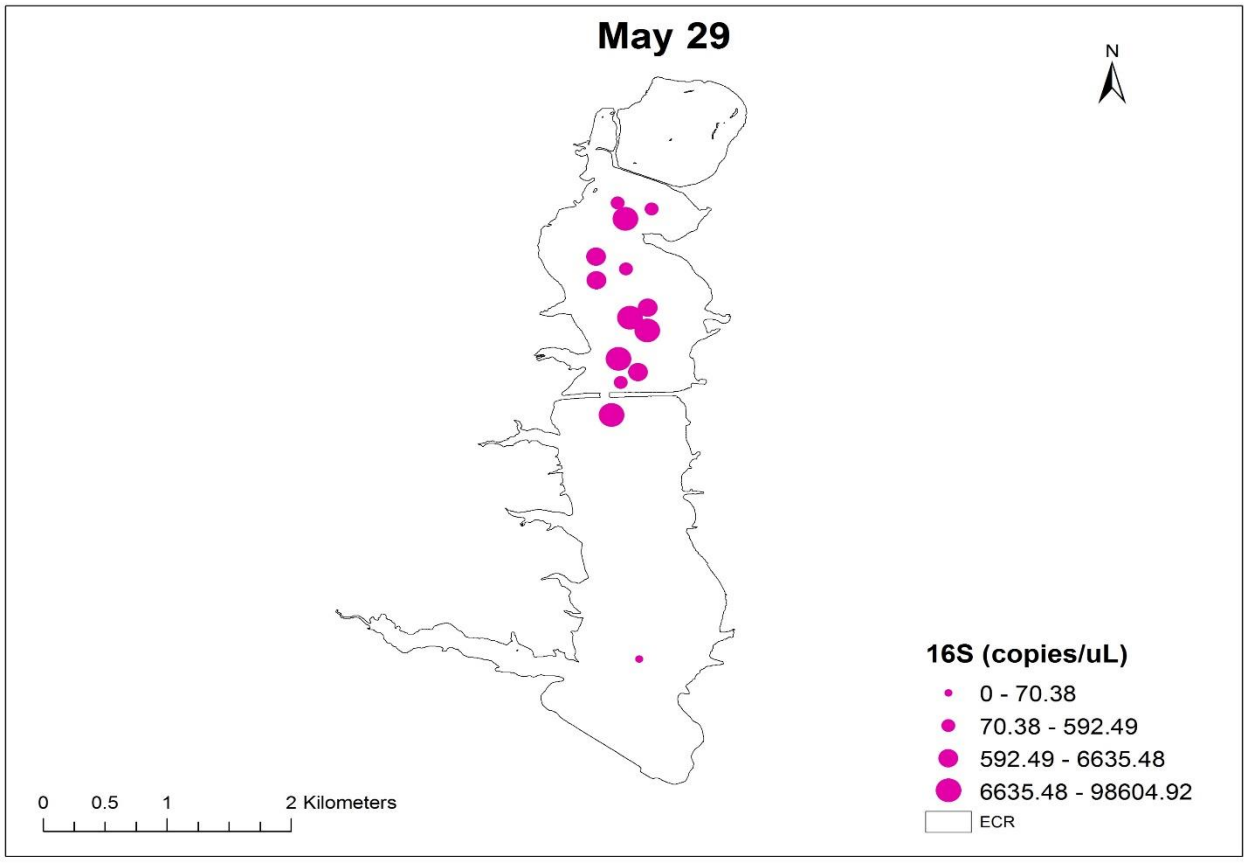


Figure 3.8 continued

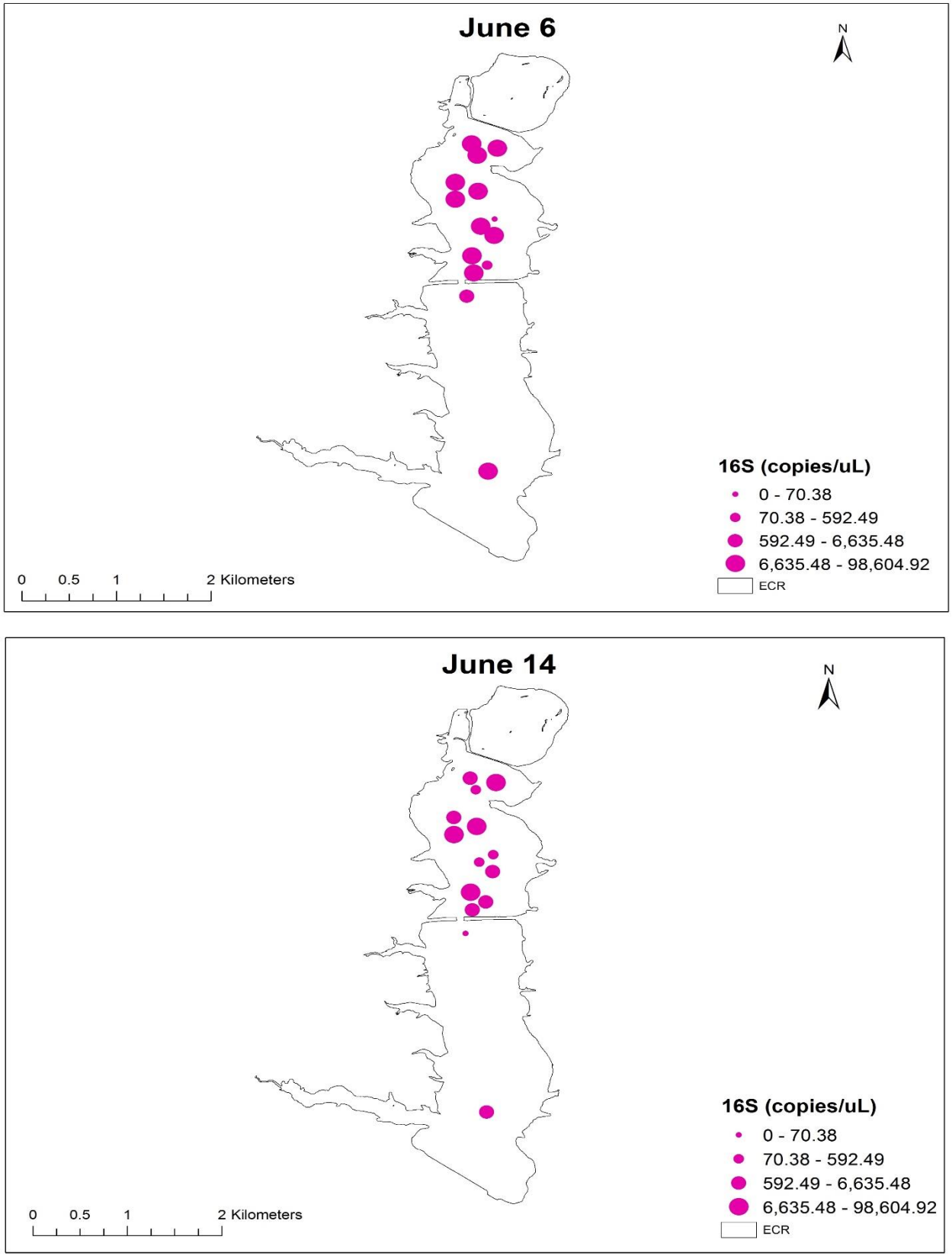
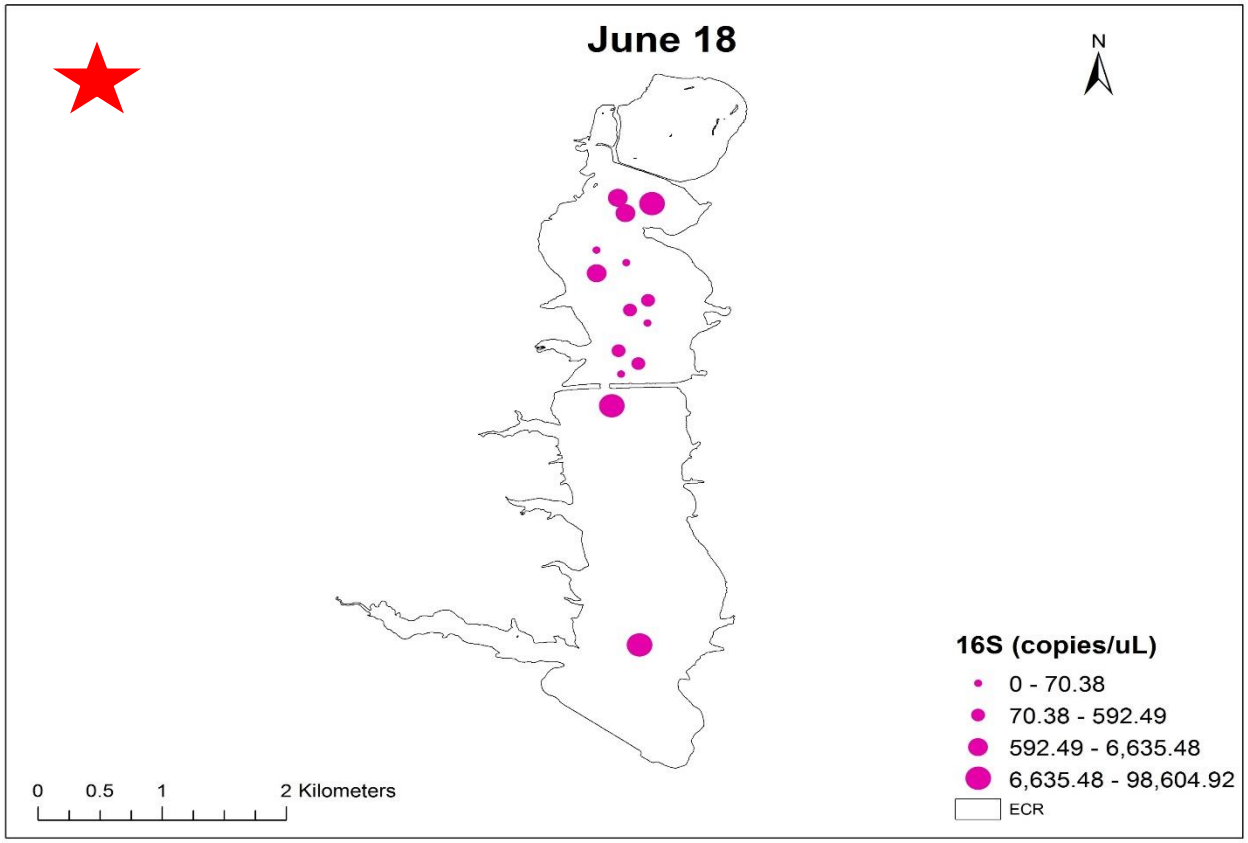


Figure 3.8 continued



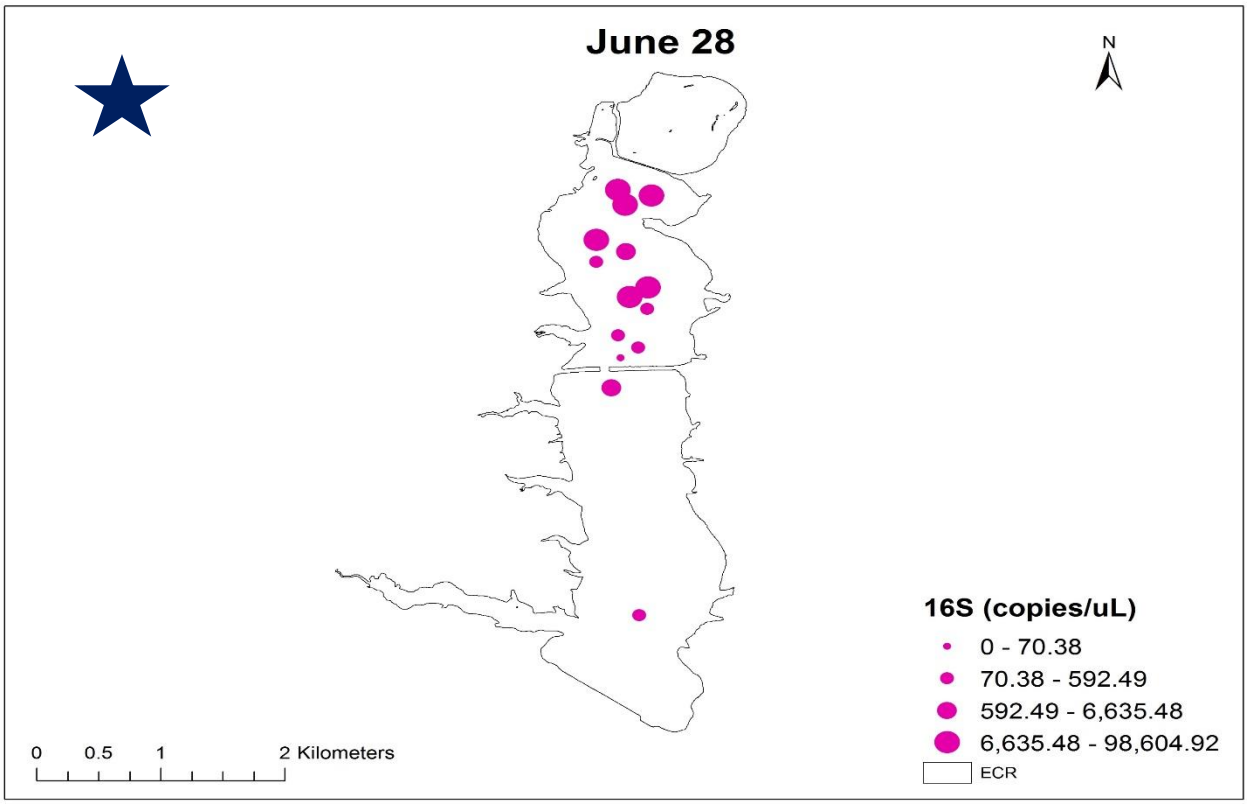
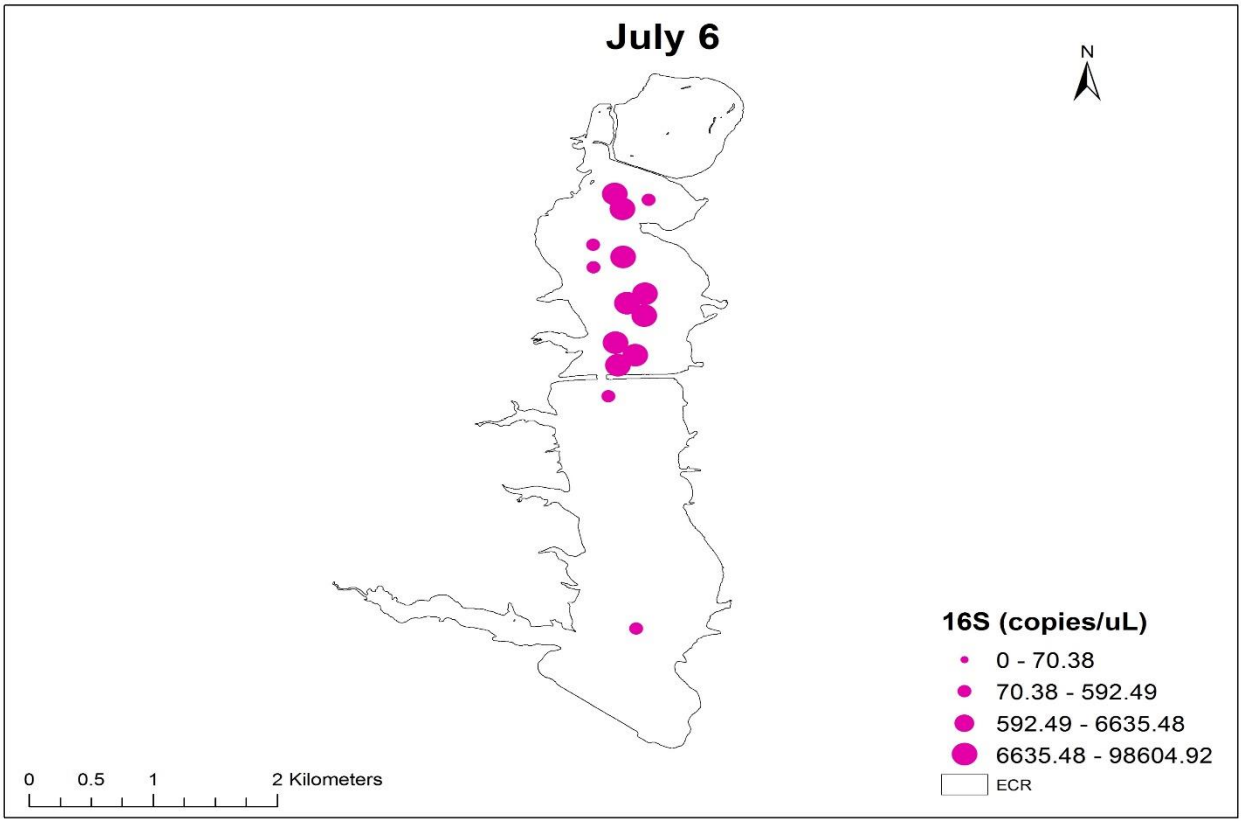


Figure 3.8 continued



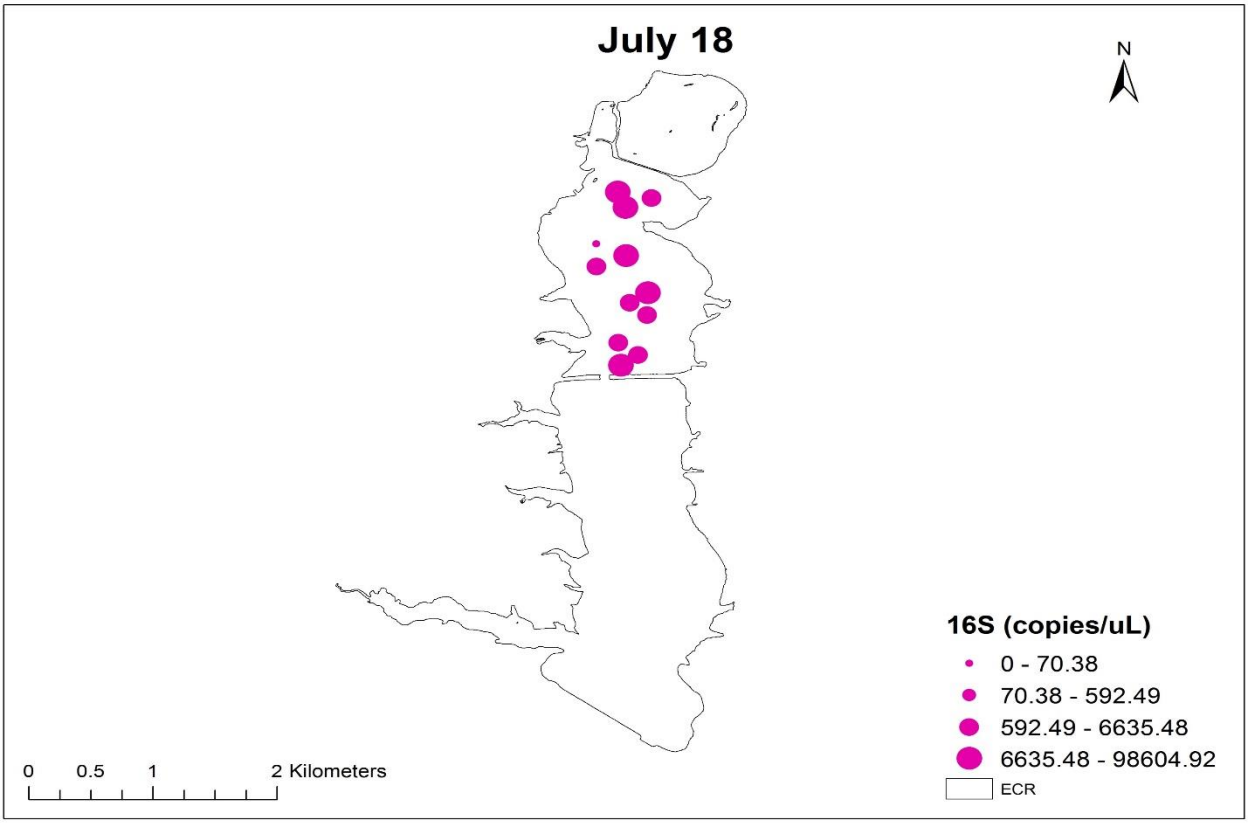
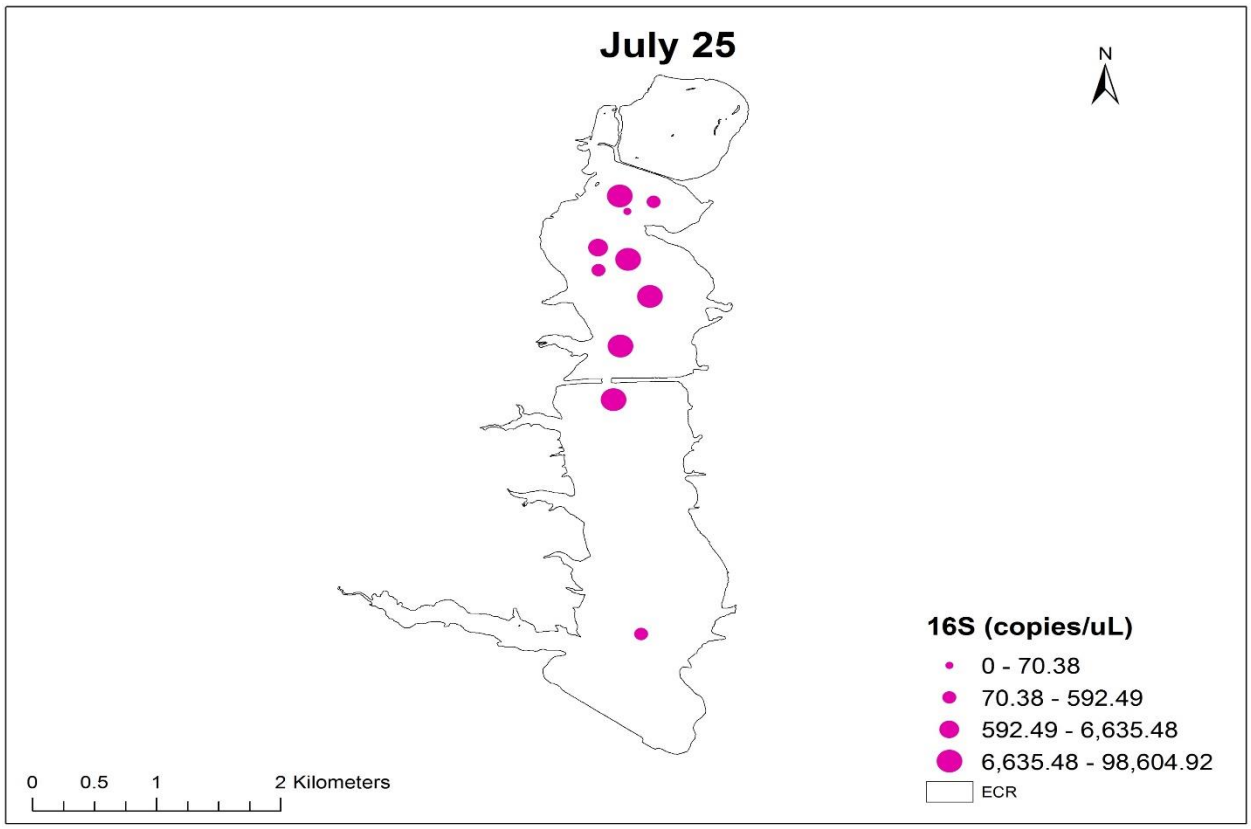


Figure 3.8 continued



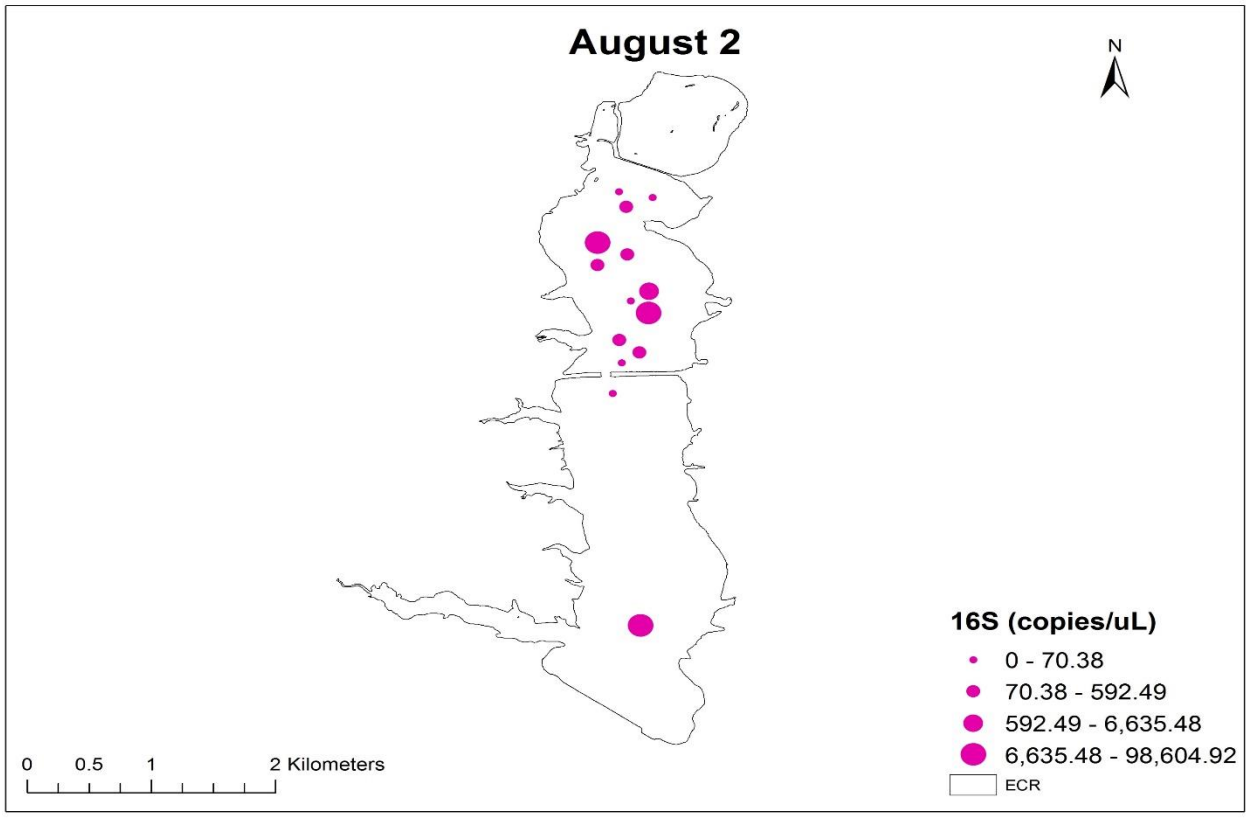
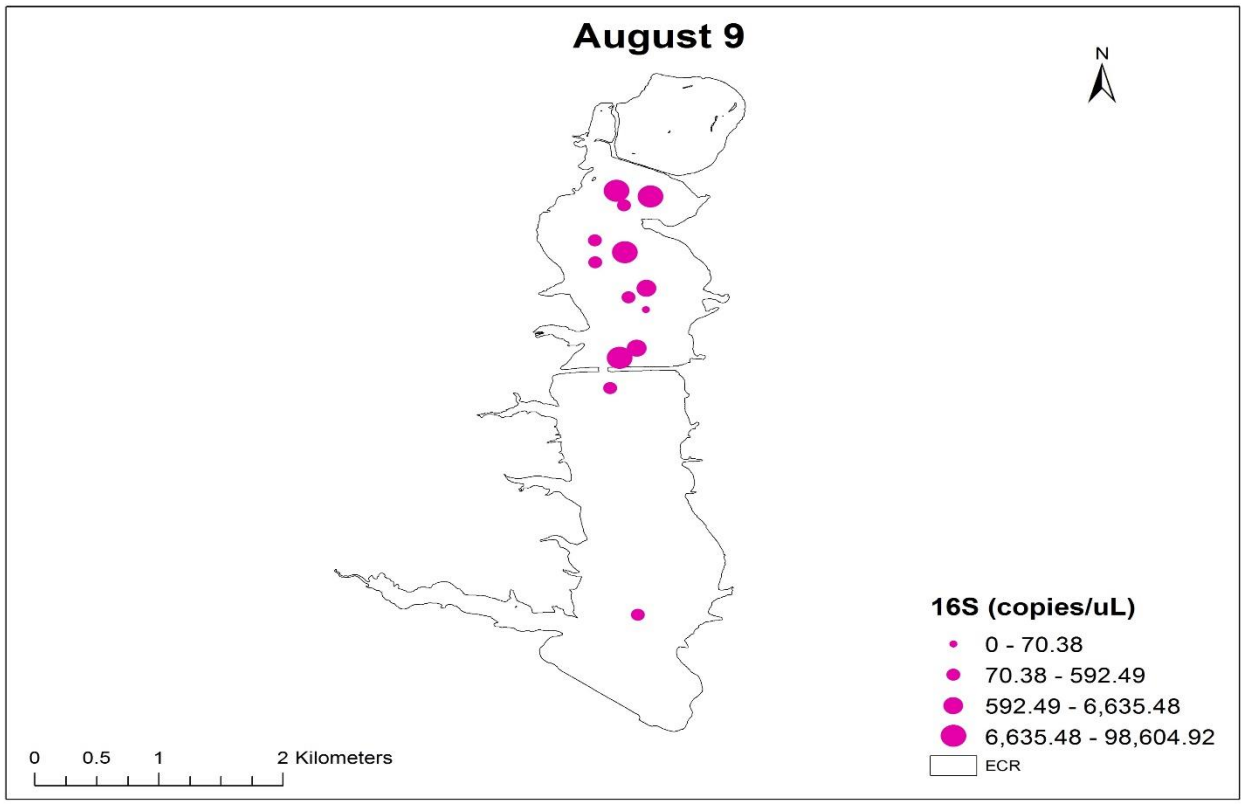


Figure 3.8 continued



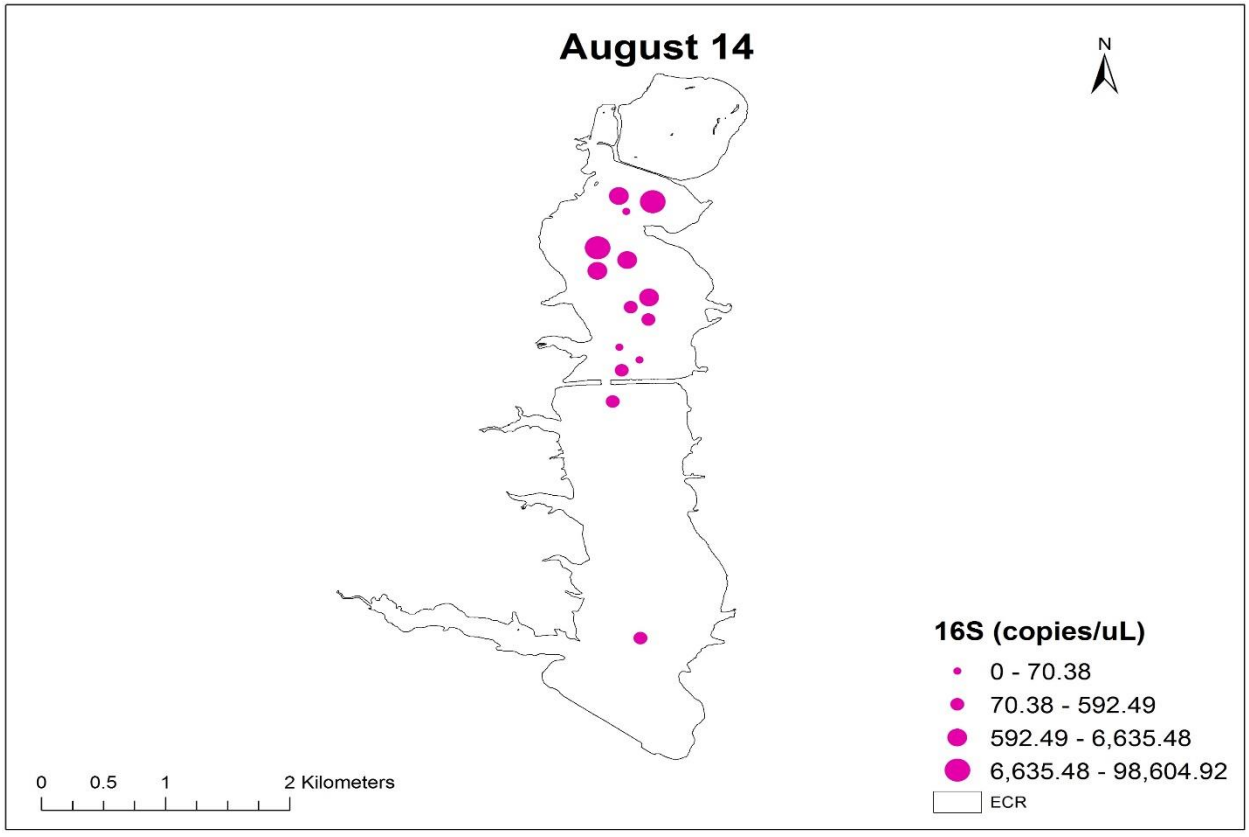
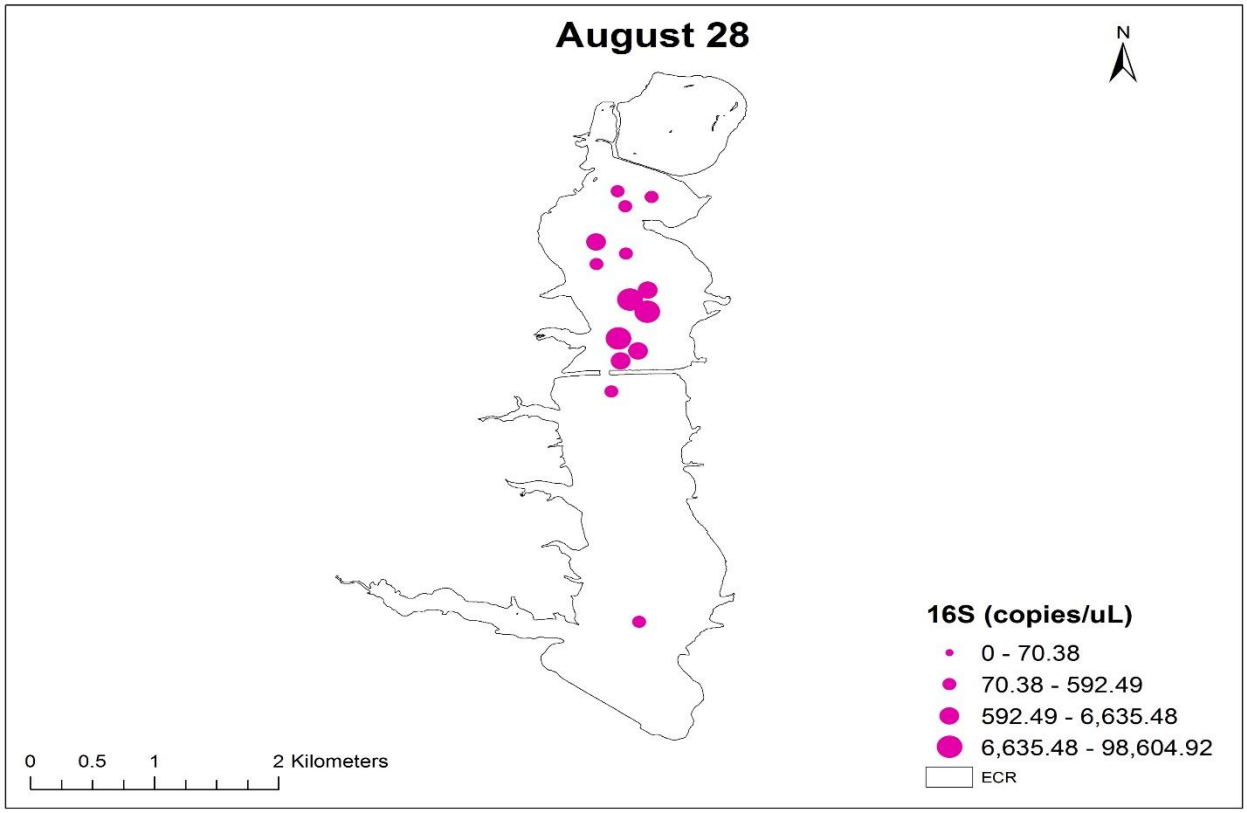


Figure 3.8 continued



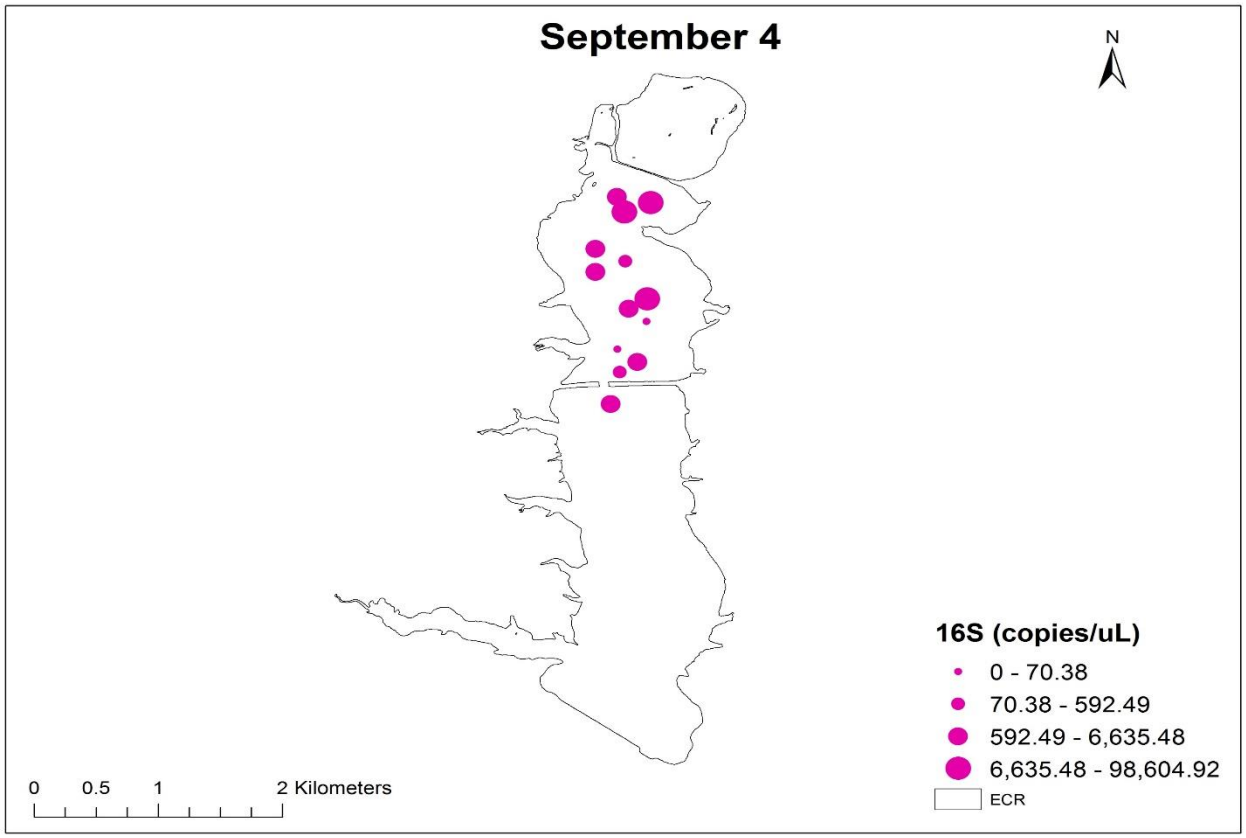
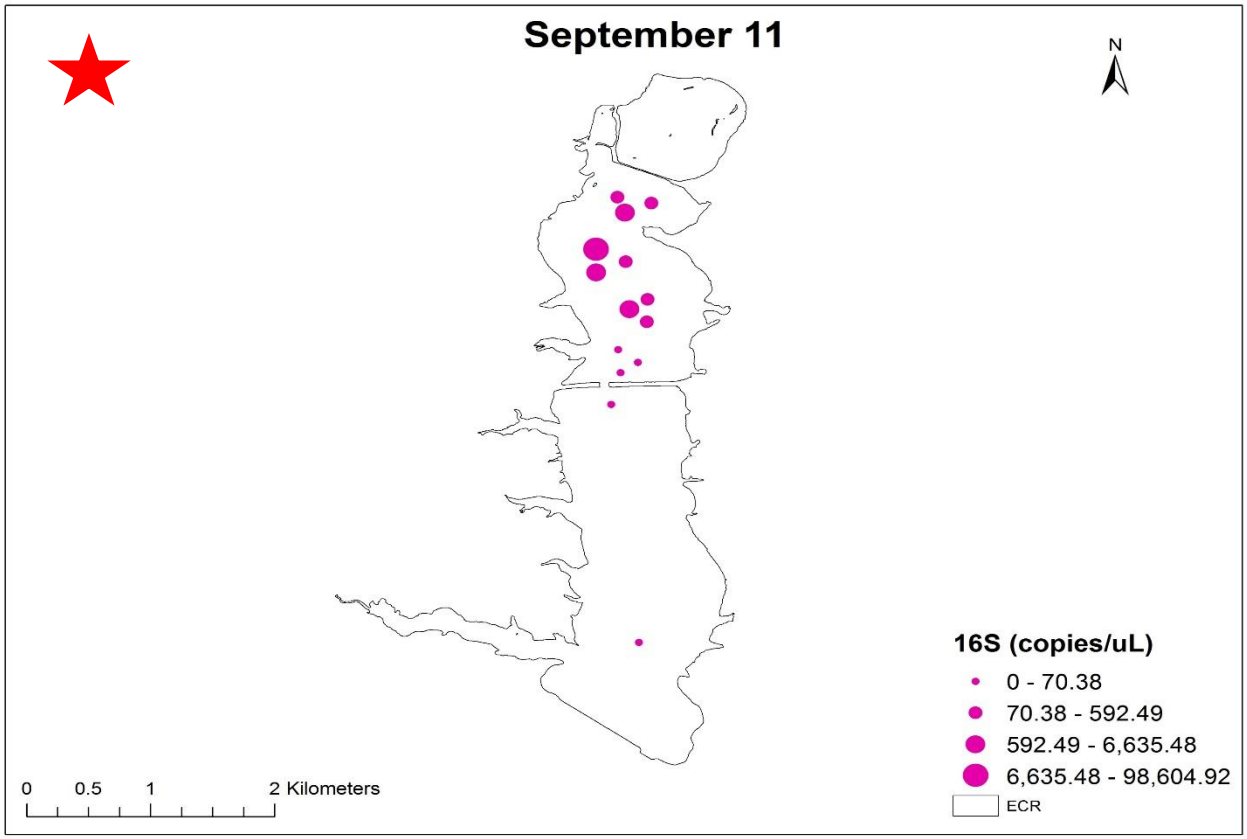


Figure 3.8 continued



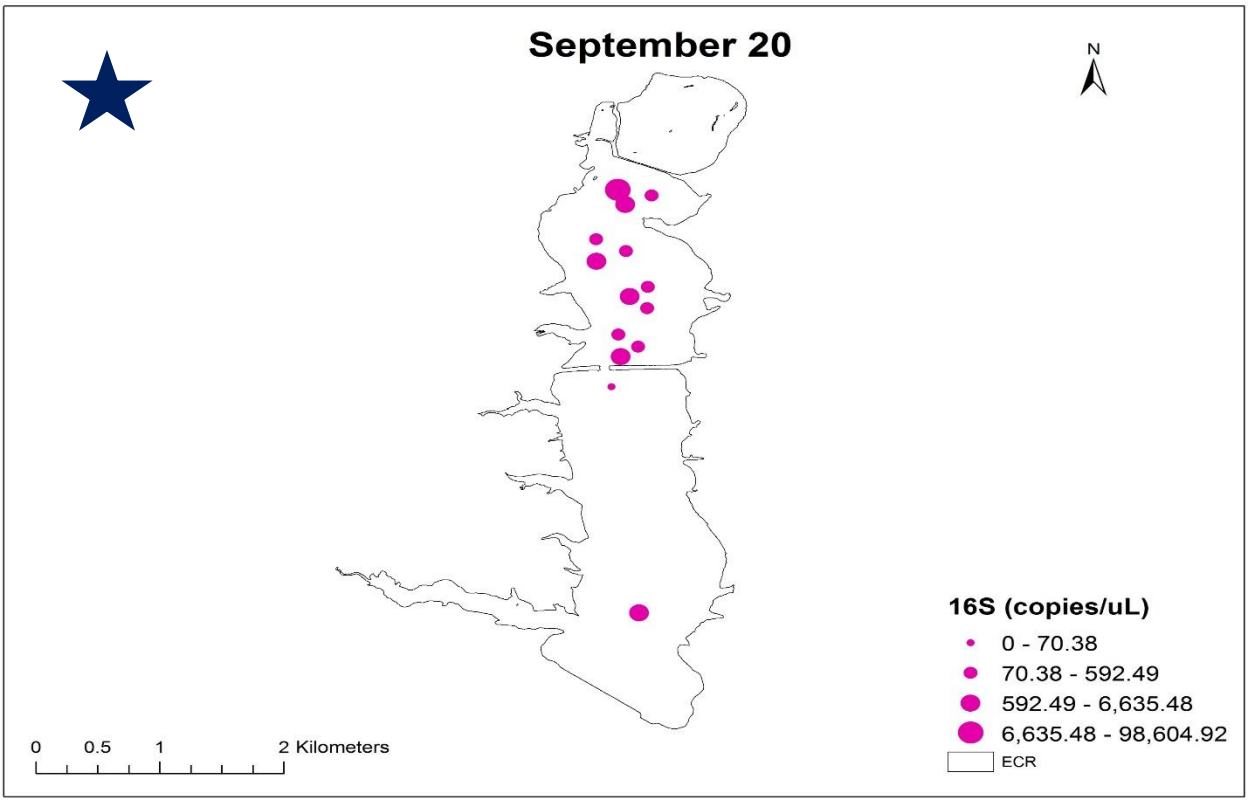
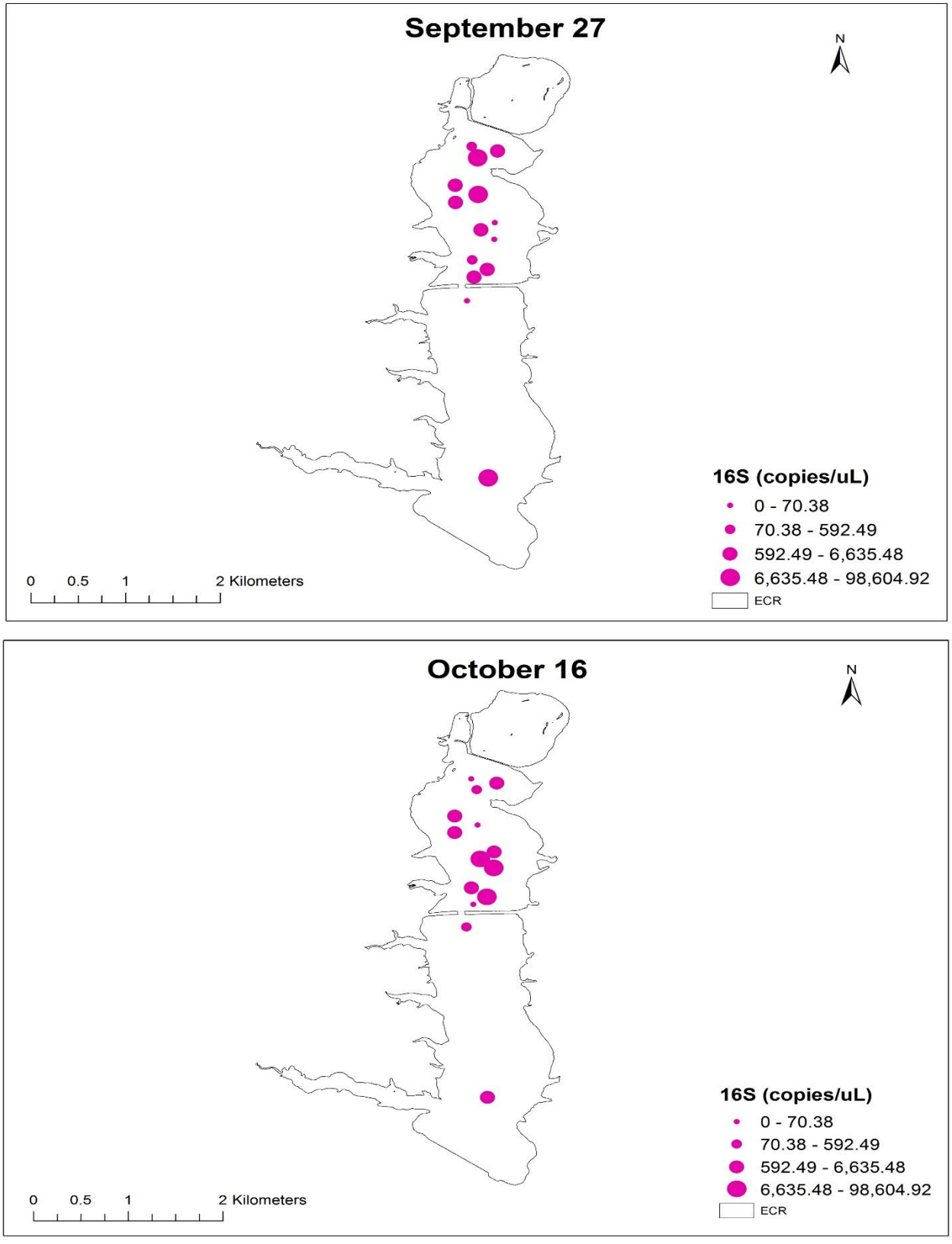


Figure 3.8 continued



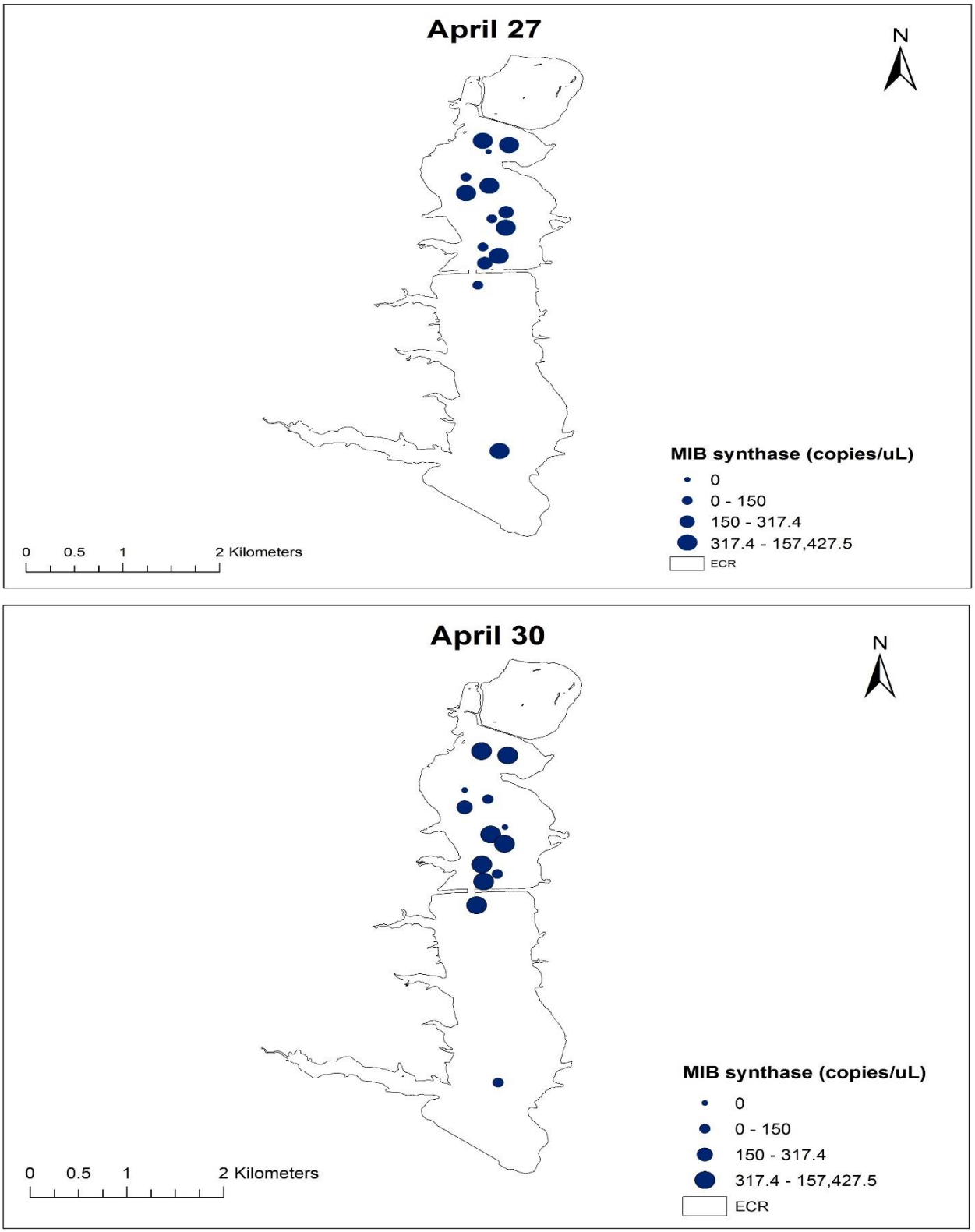
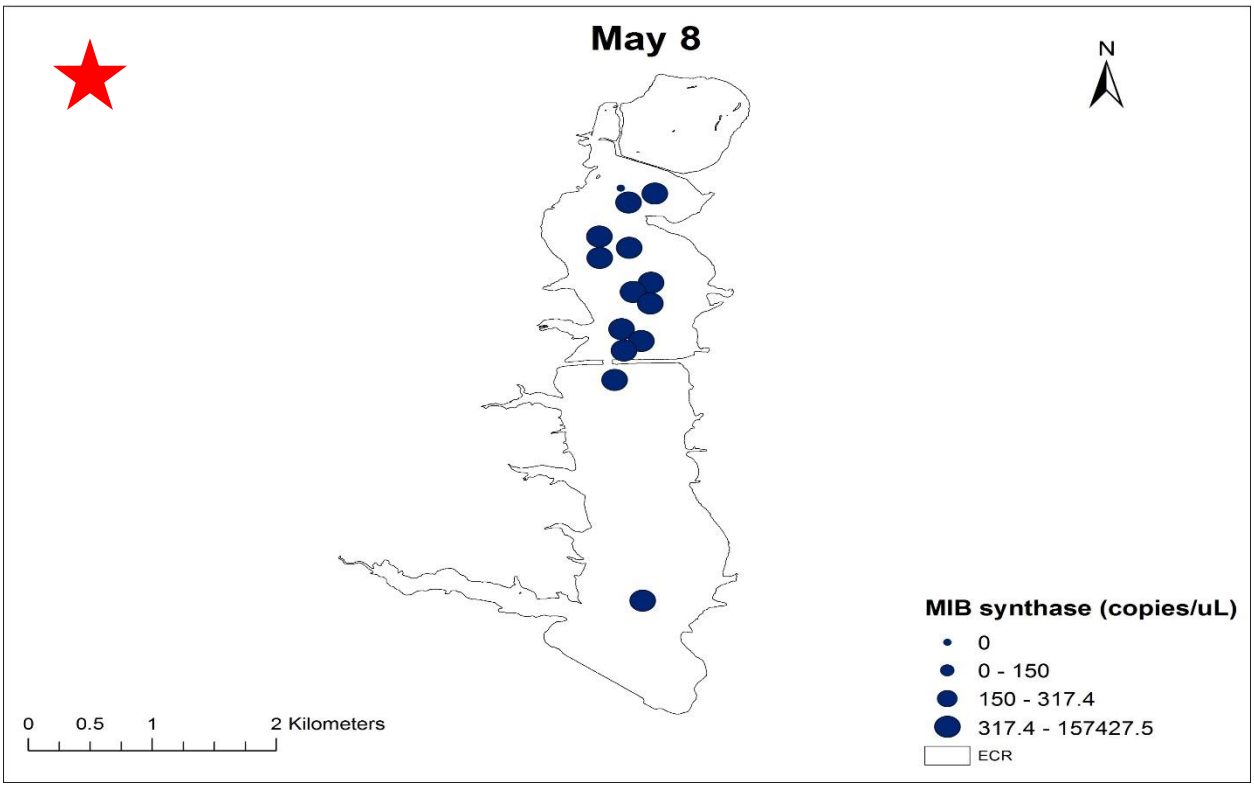


Figure 3.9 Spatial maps of *MIBS* gene quantity for each sampling date. Data are sorted on a quantile basis. Red stars indicate sampling date before the algaecide treatment and blue stars indicate the sampling date after treatment.

Figure 3.9 continued



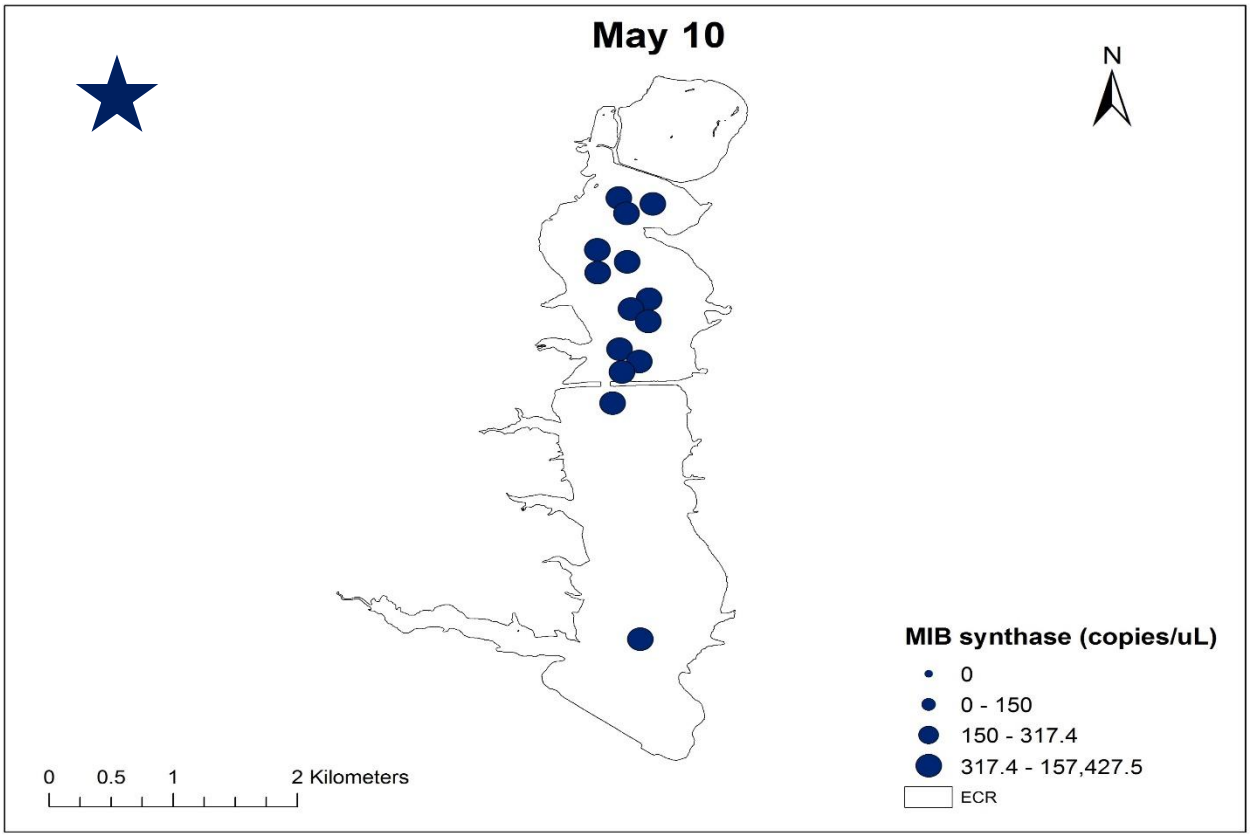
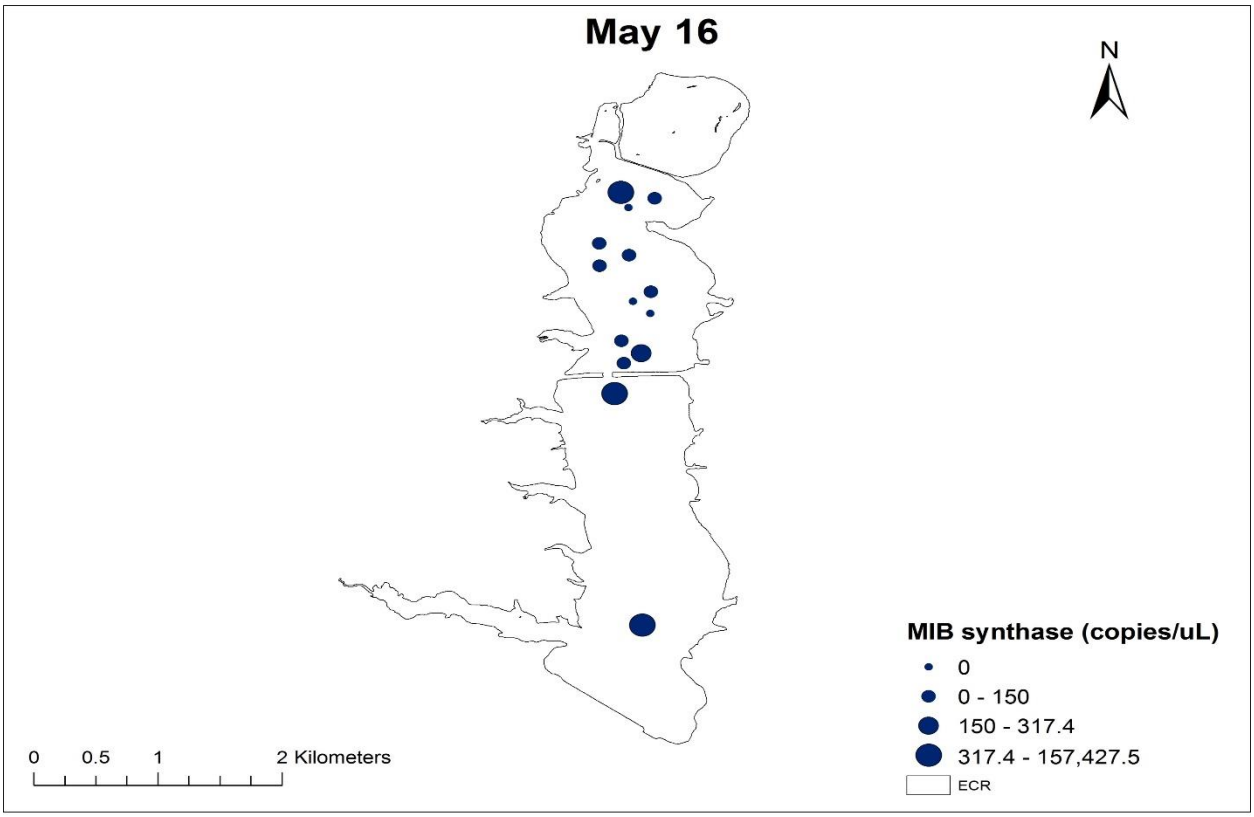


Figure 3.9 continued



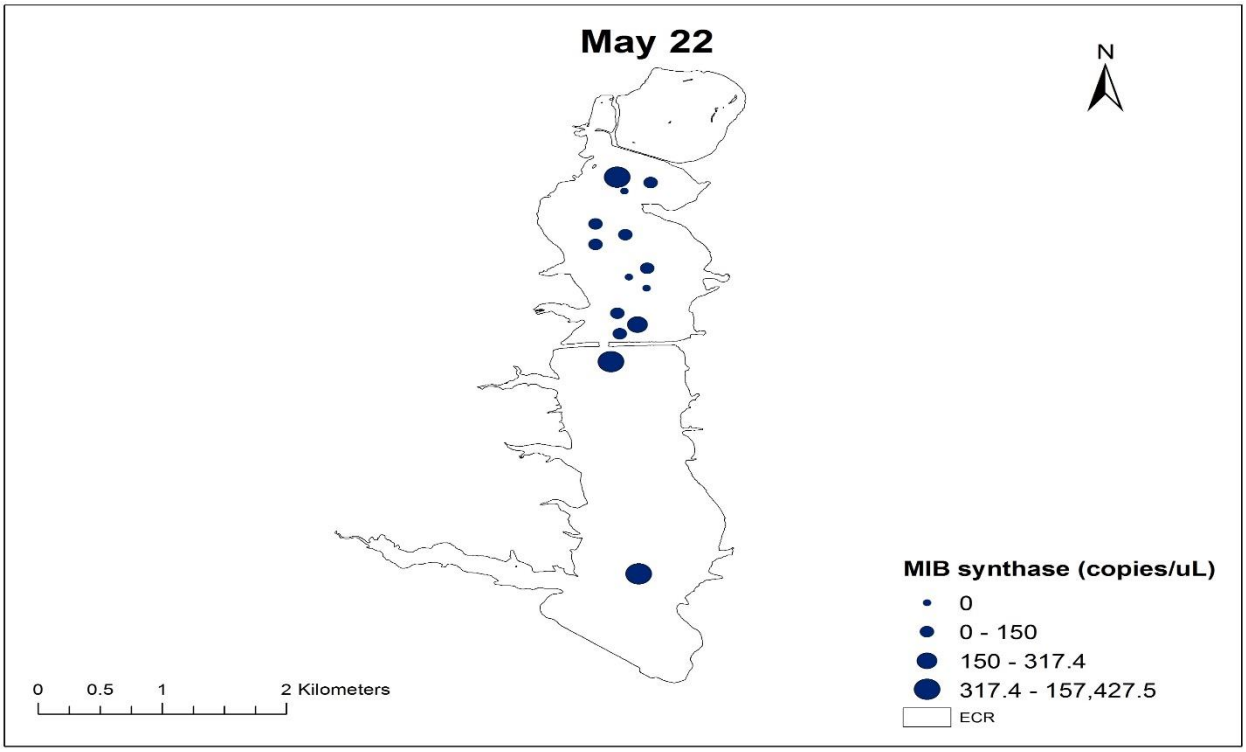
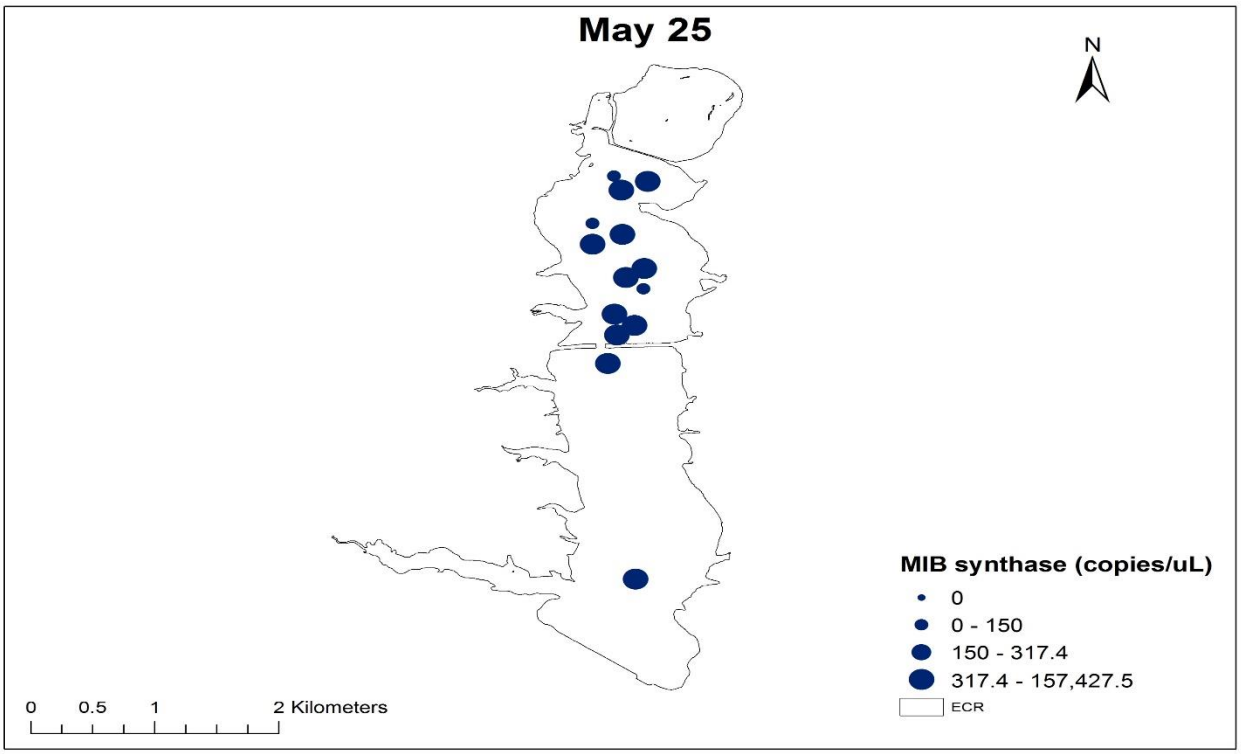


Figure 3.9 continued



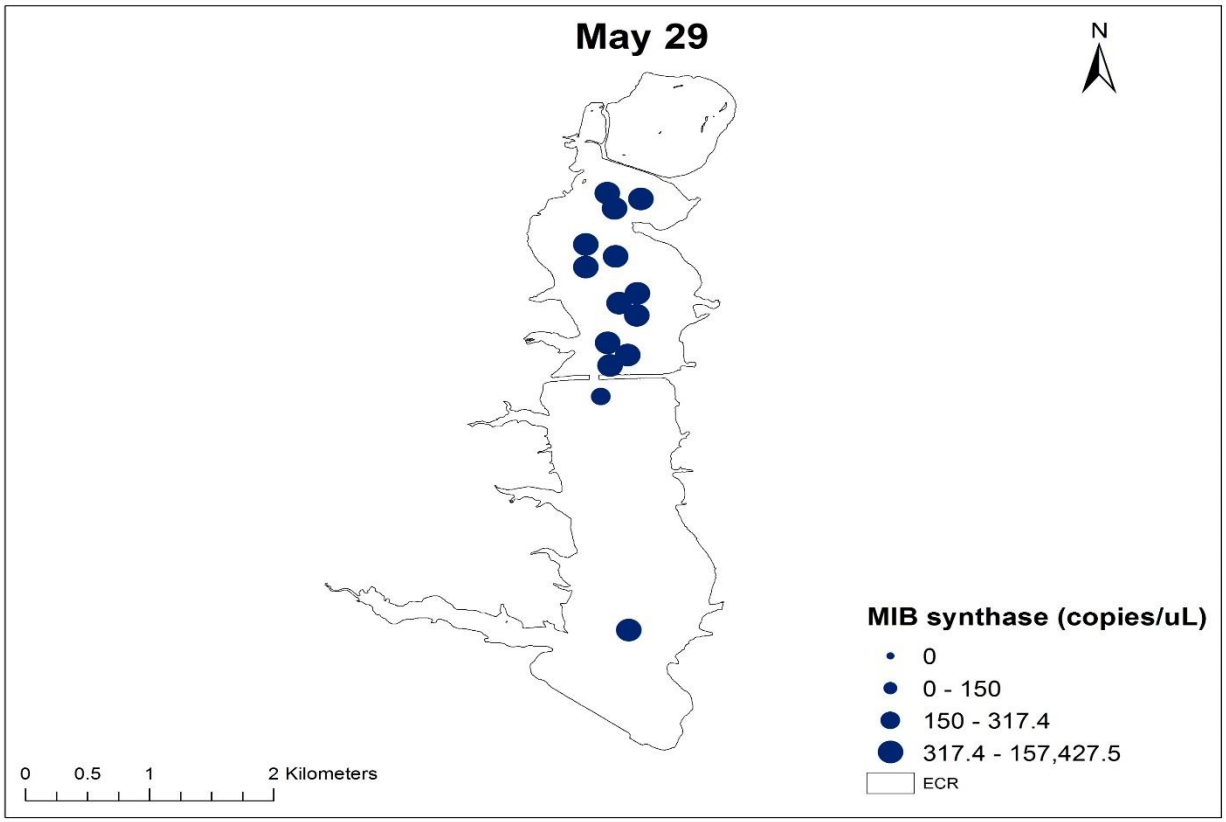
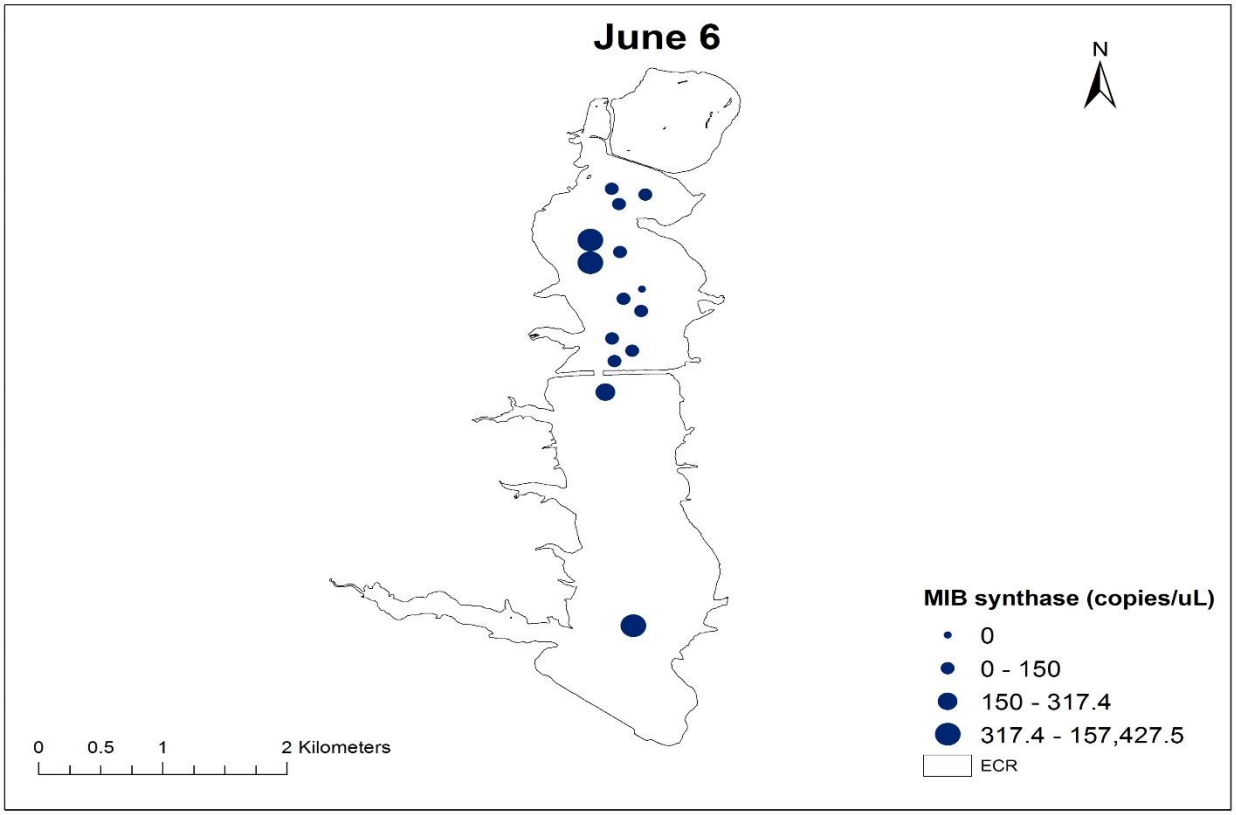


Figure 3.9 continued



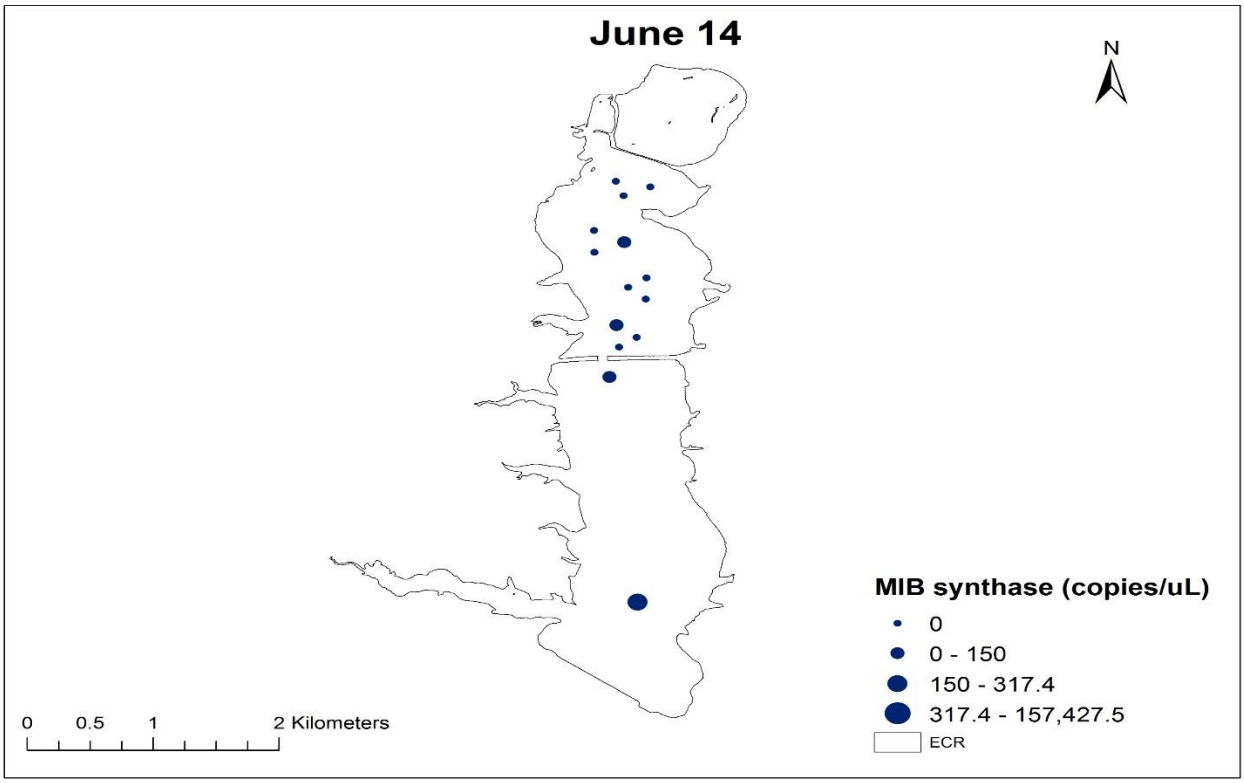


Figure 3.9 continued

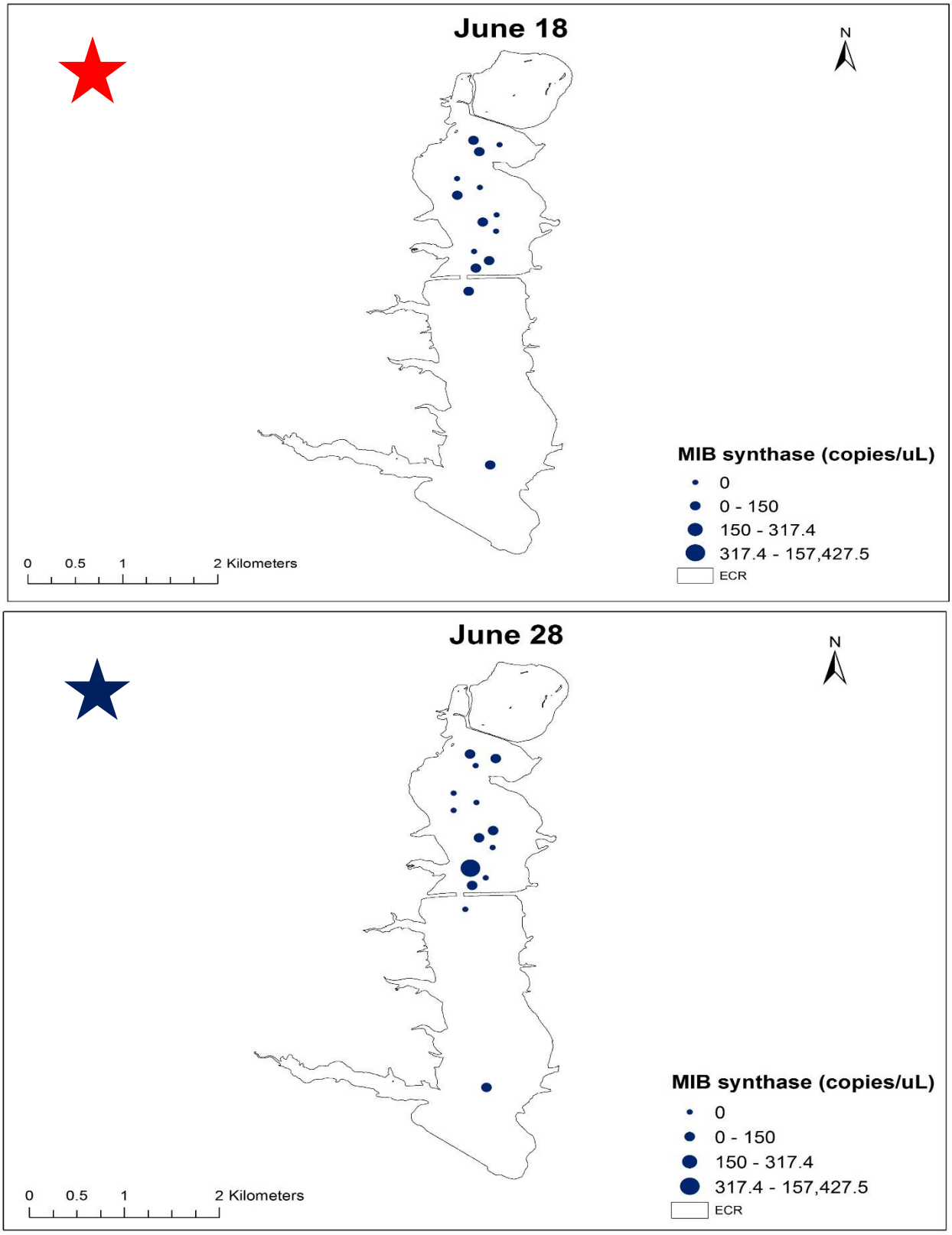
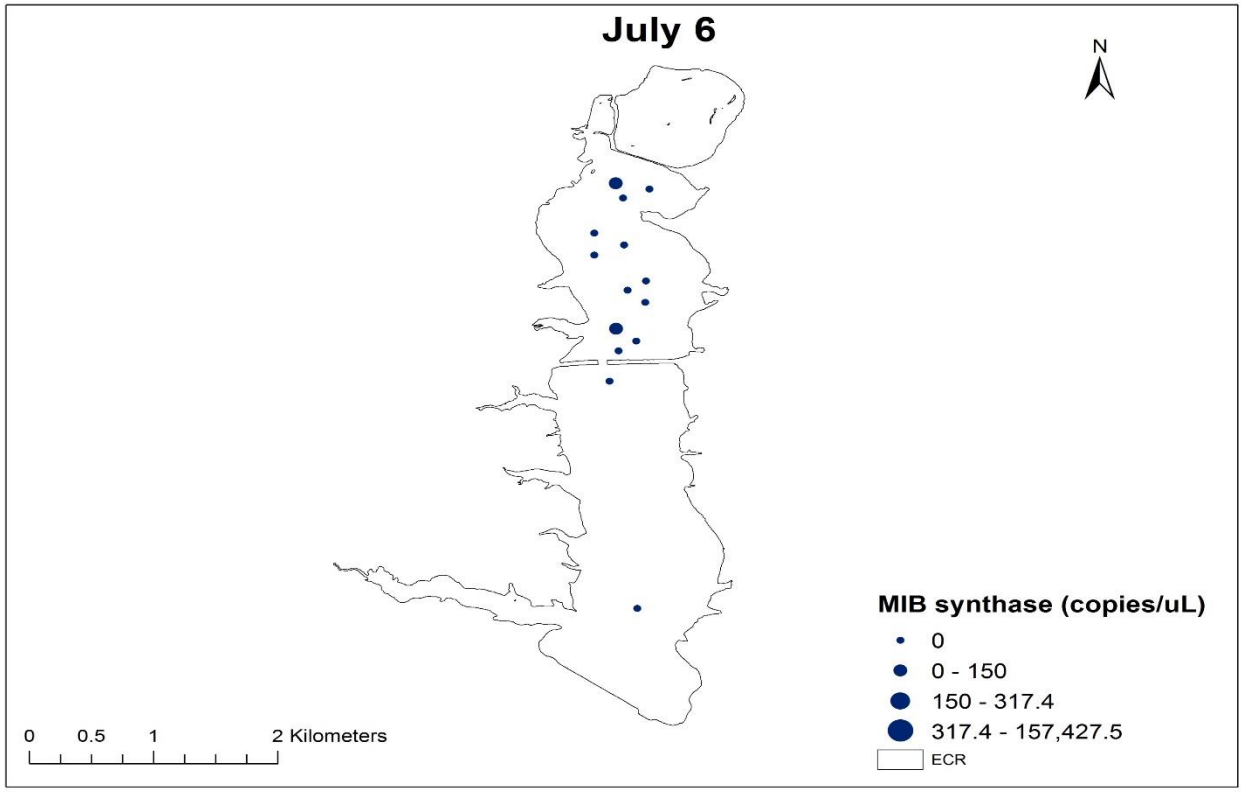


Figure 3.9 continued



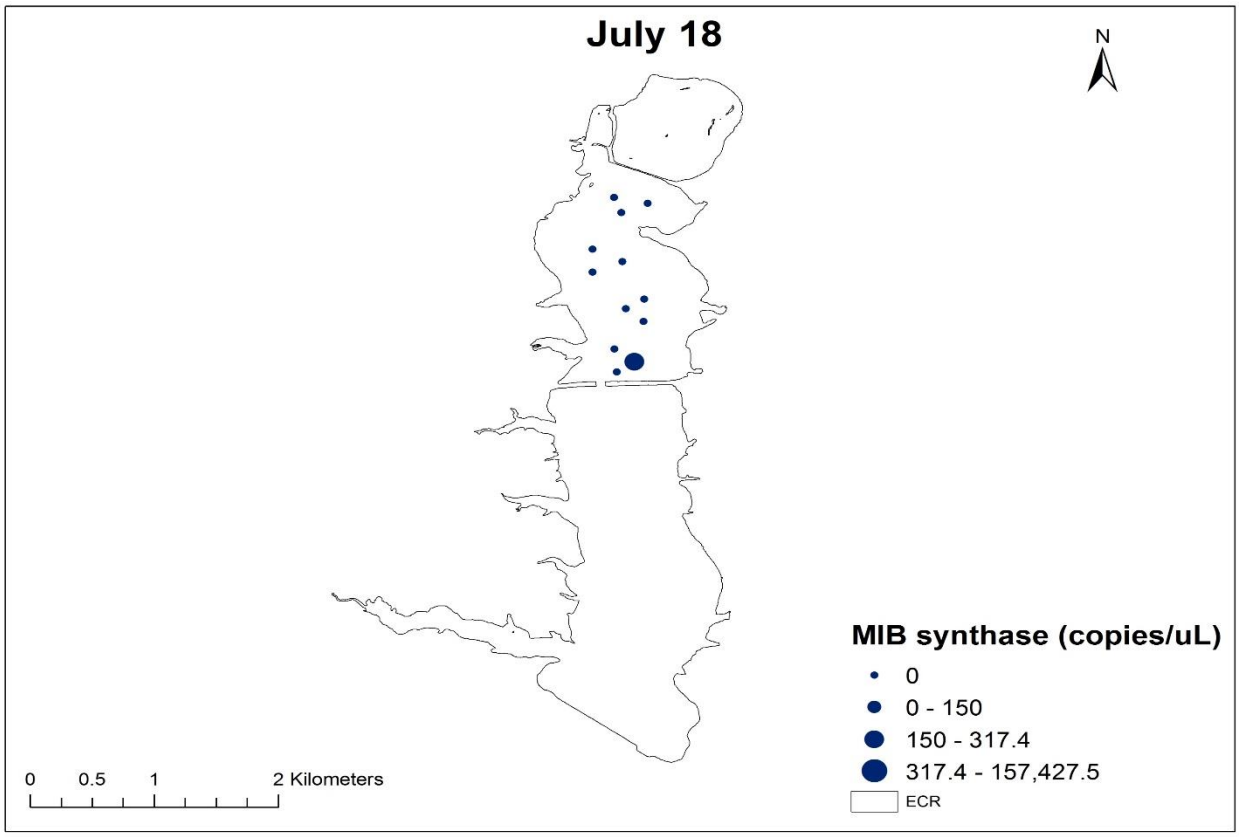
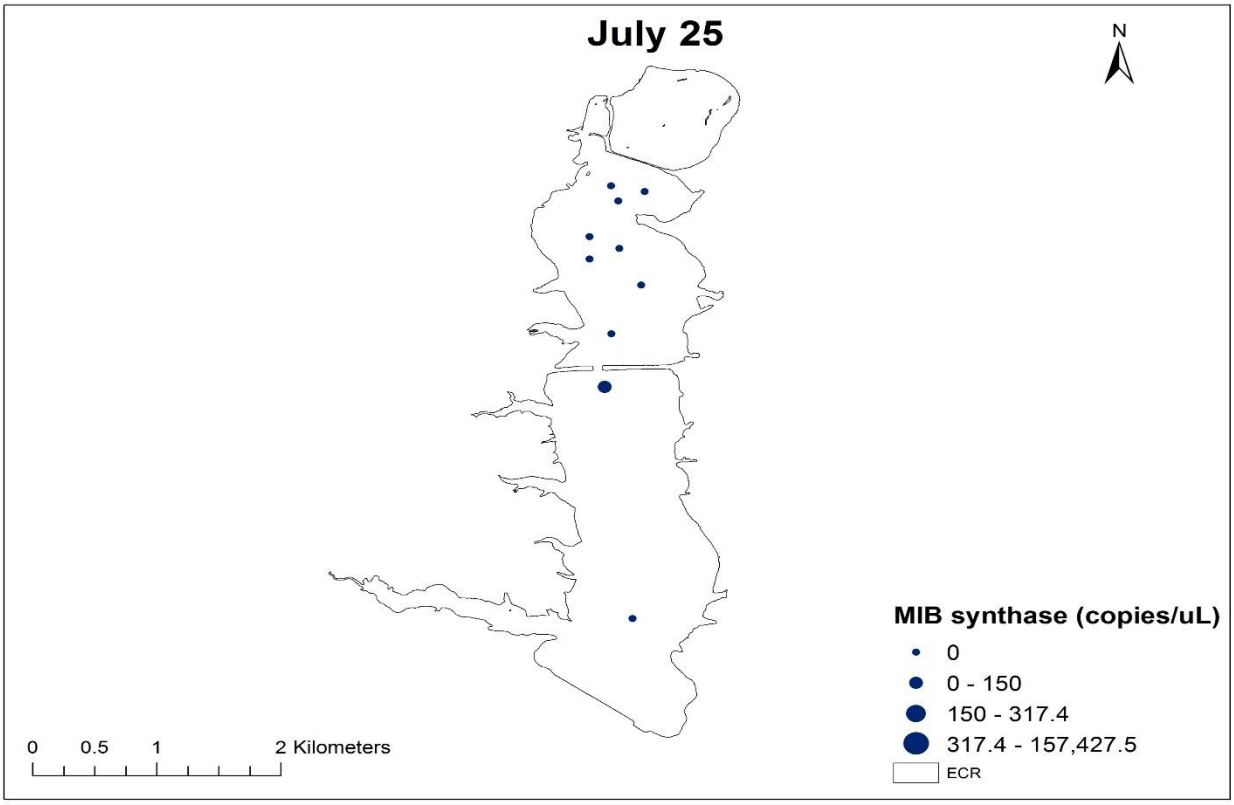


Figure 3.9 continued



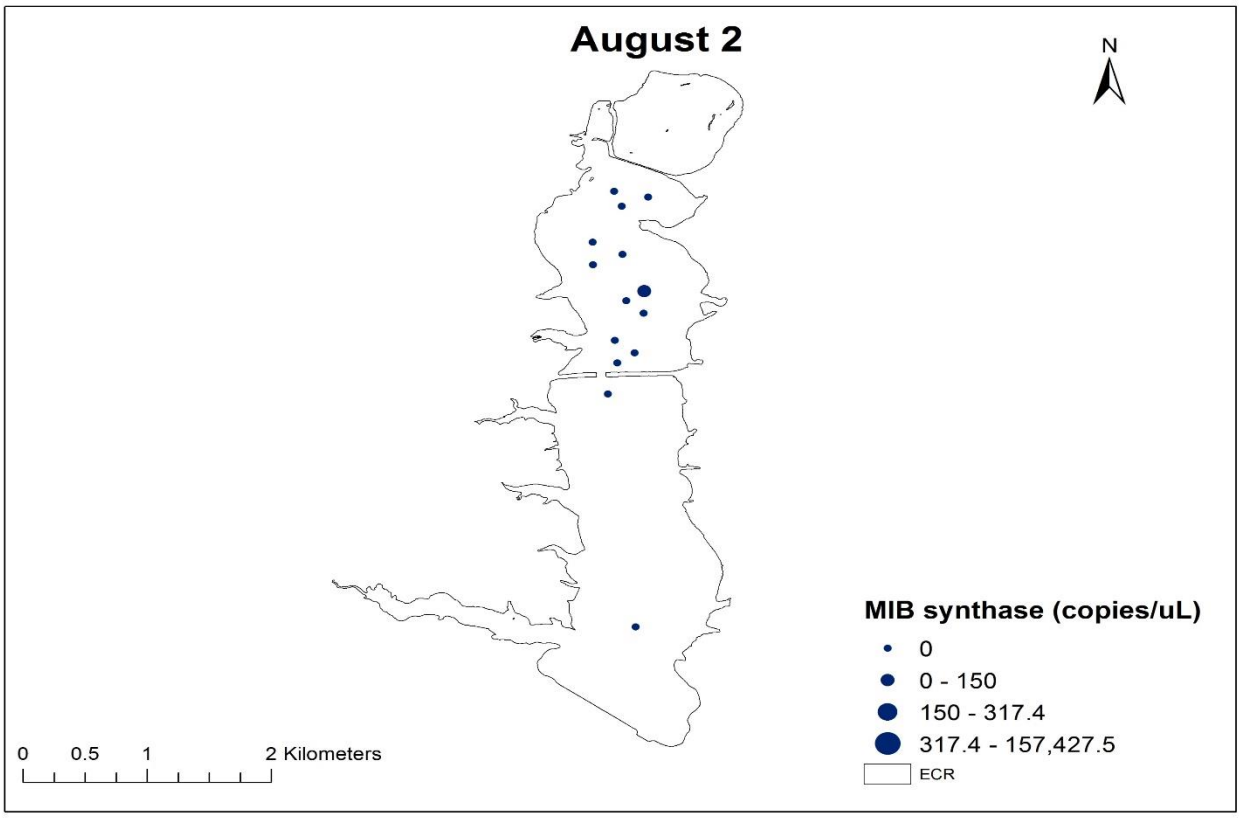
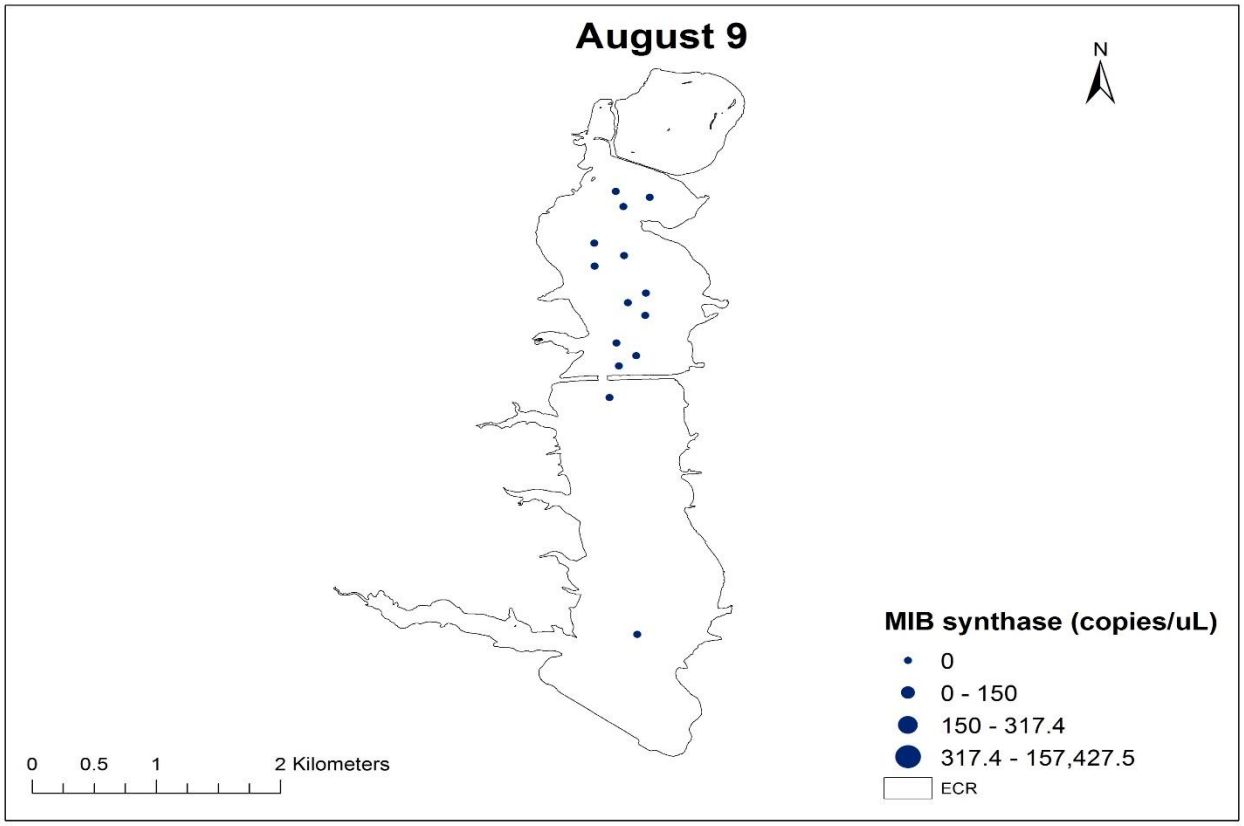


Figure 3.9 continued



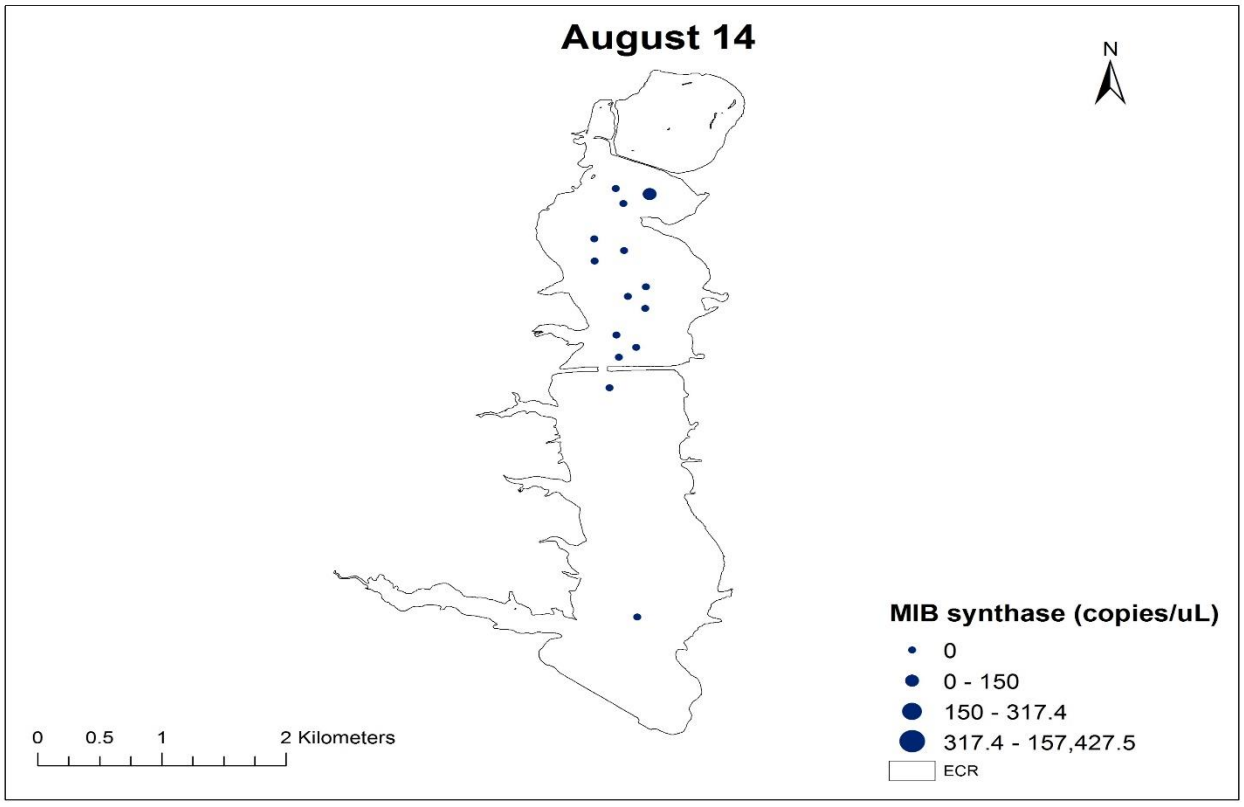


Figure 3.9 continued

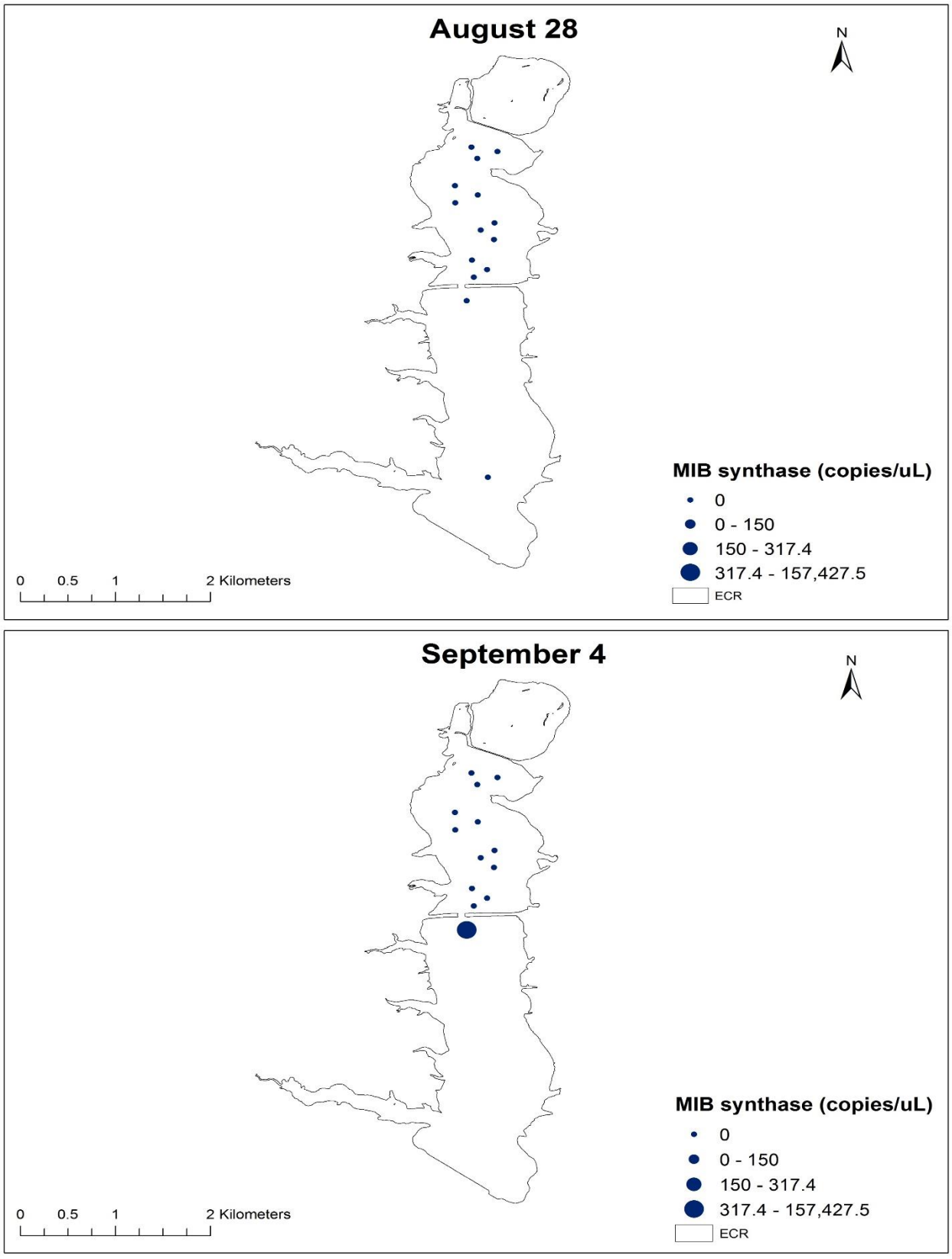
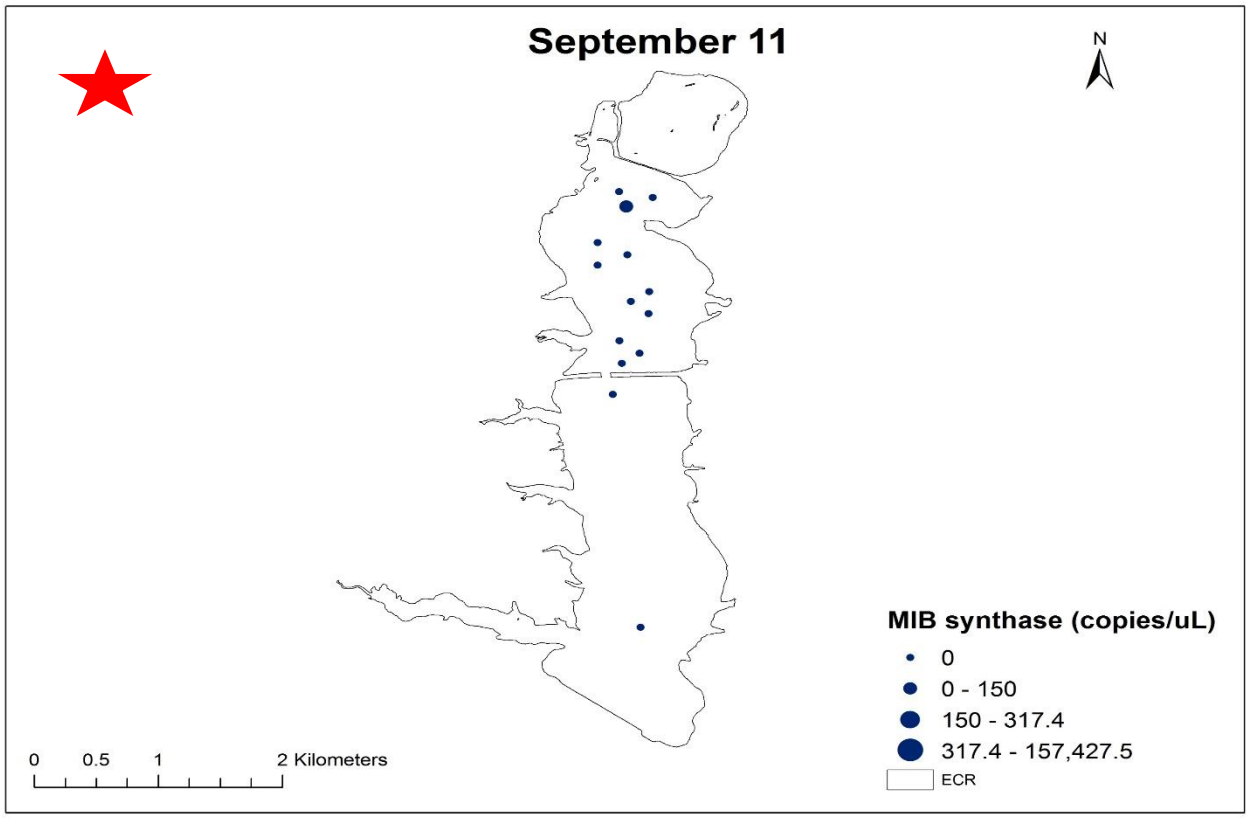


Figure 3.9 continued



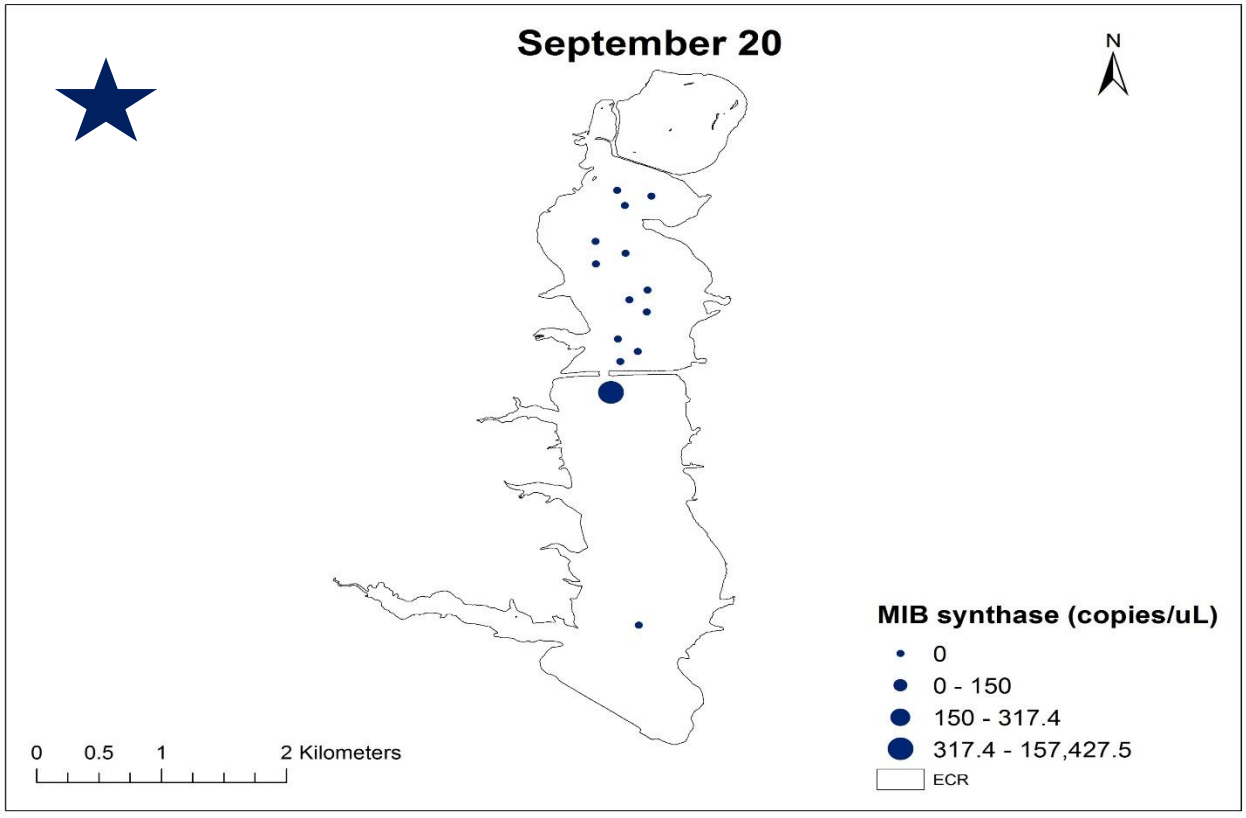
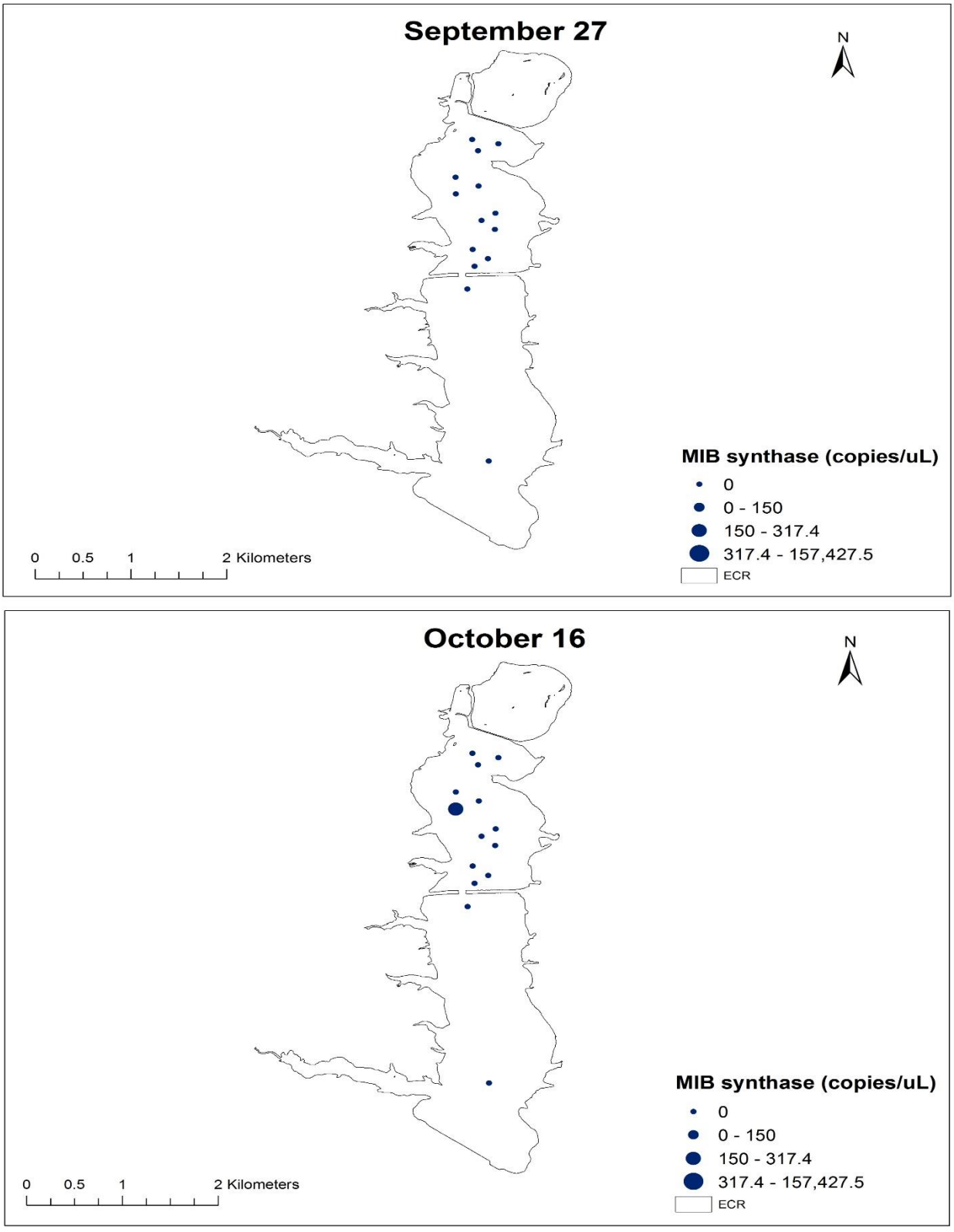


Figure 3.9 continued



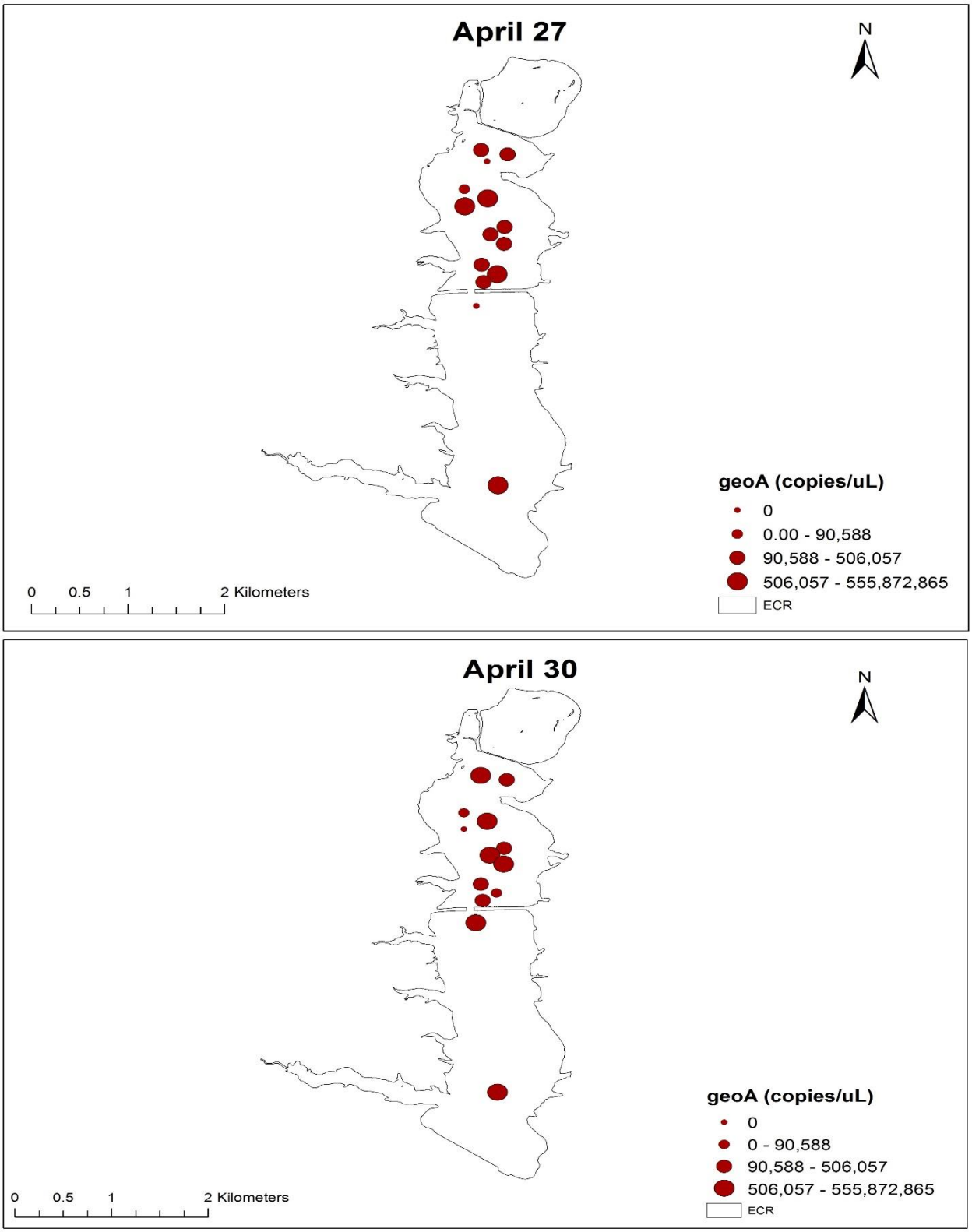
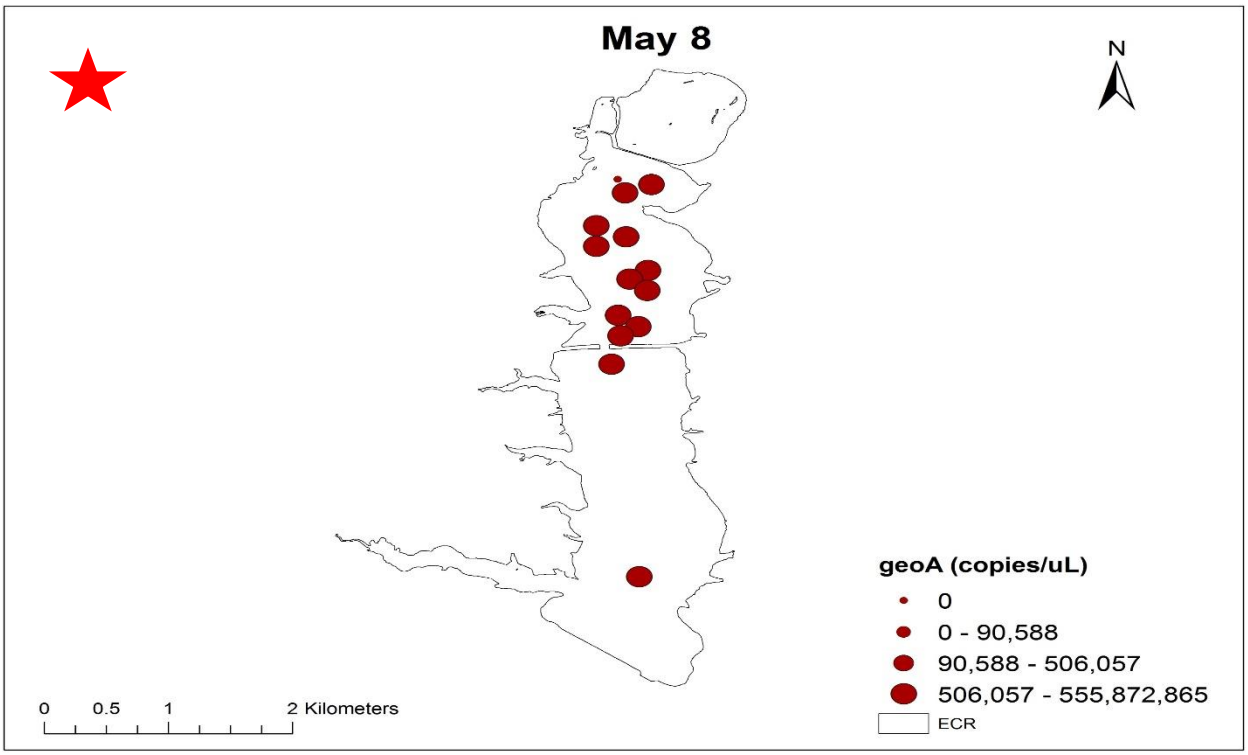


Figure 3.10 Spatial maps of *geoA* gene quantity for each sampling date. Data are sorted on a quantile basis. Red stars indicate sampling date before the algaecide treatment and blue stars indicate the sampling date after treatment.

Figure 3.10 continued



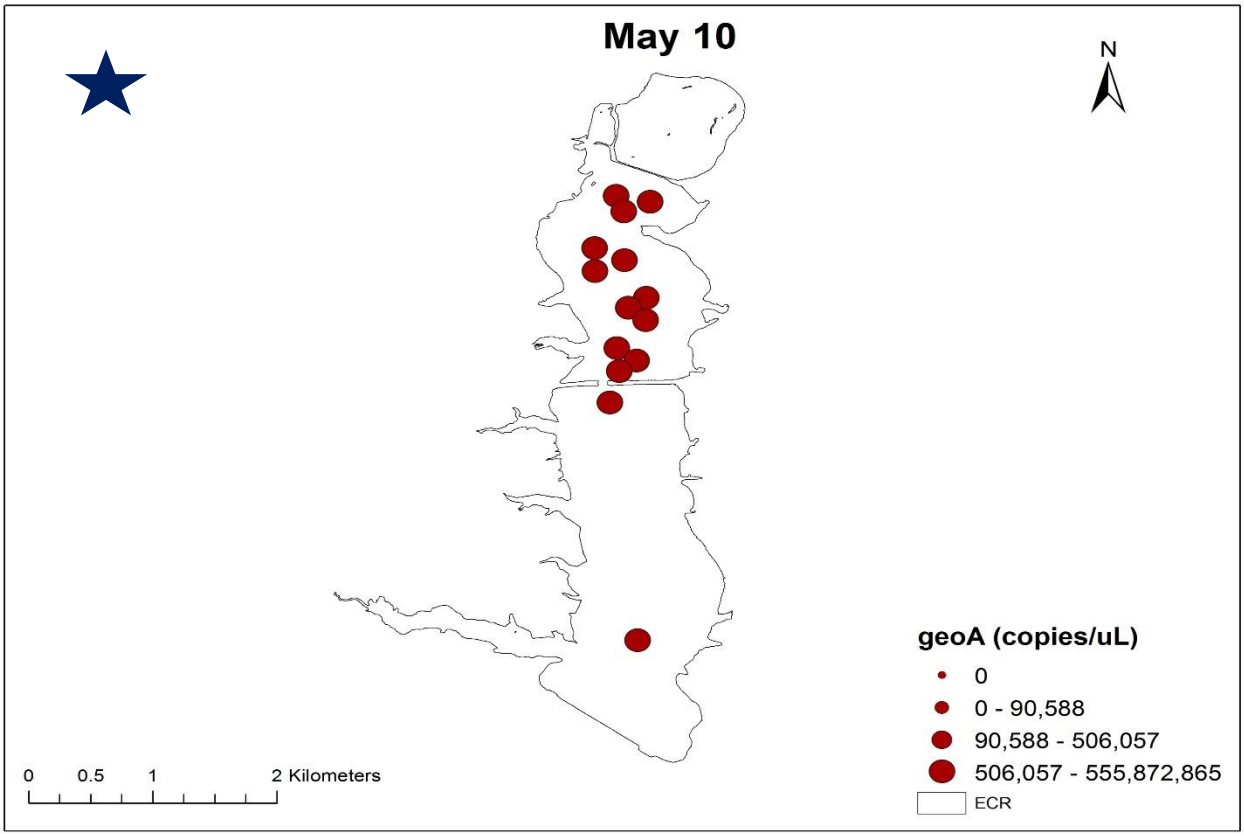
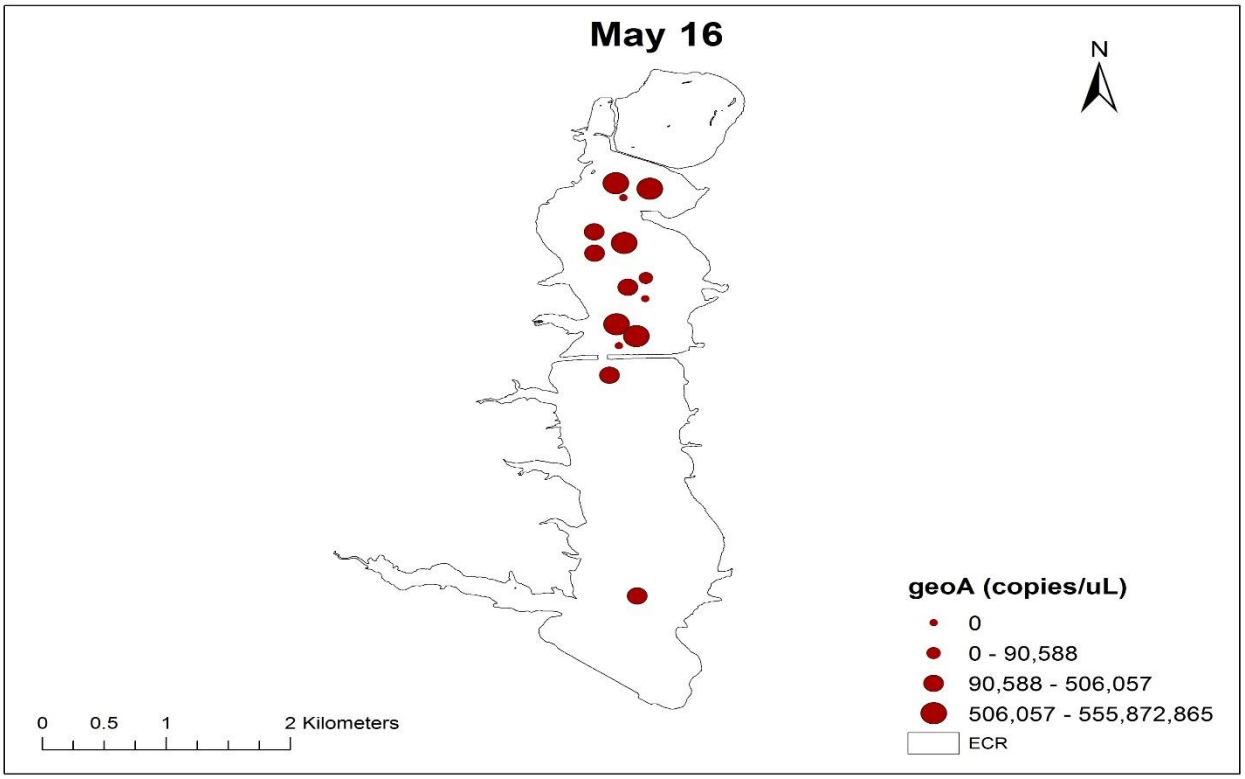


Figure 3.10 continued



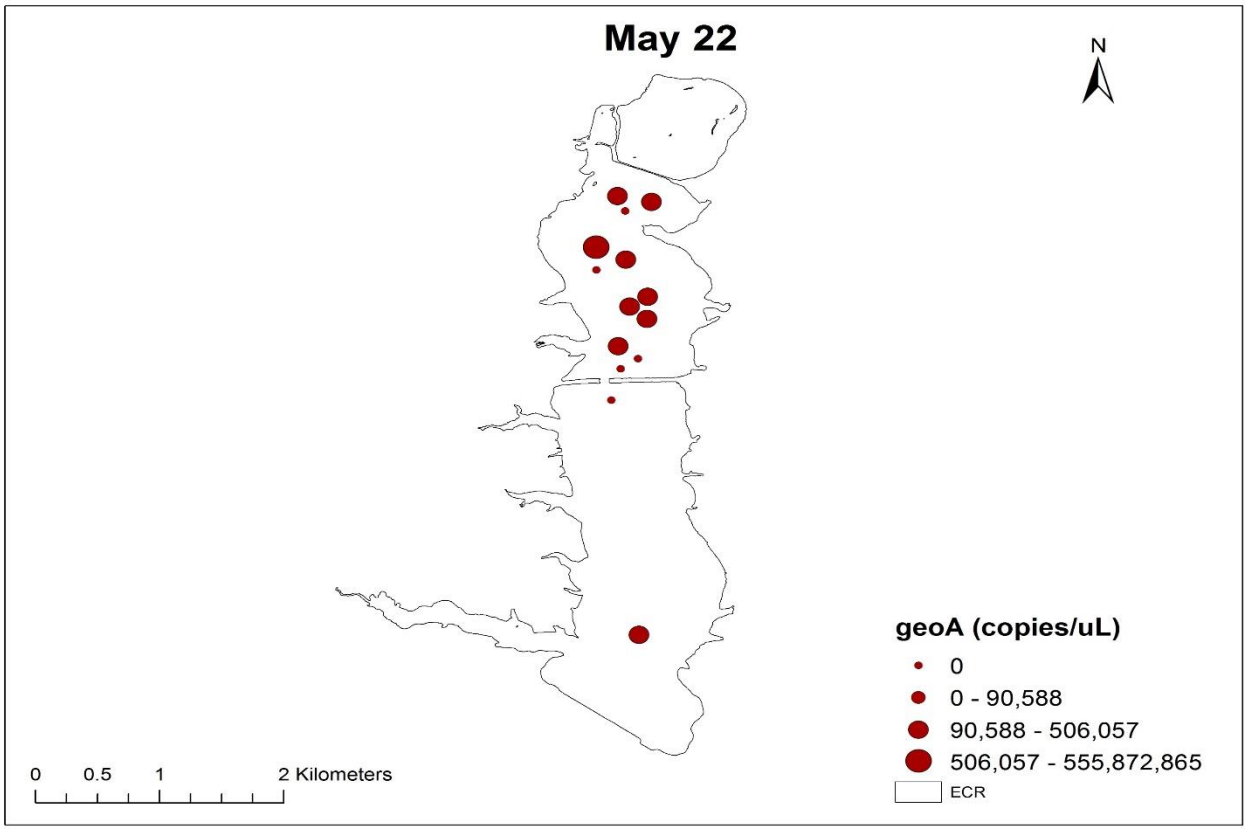


Figure 3.10 continued

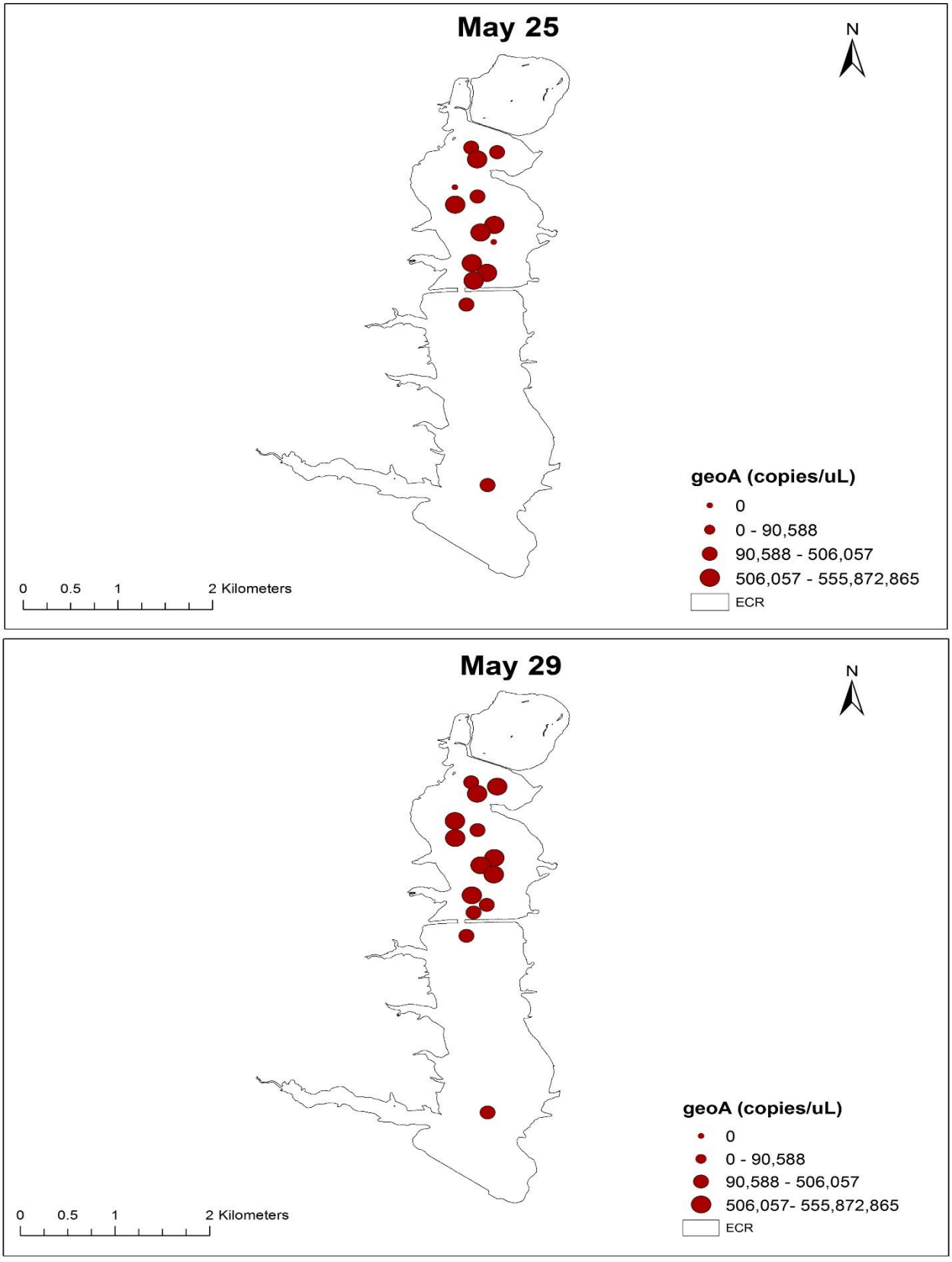
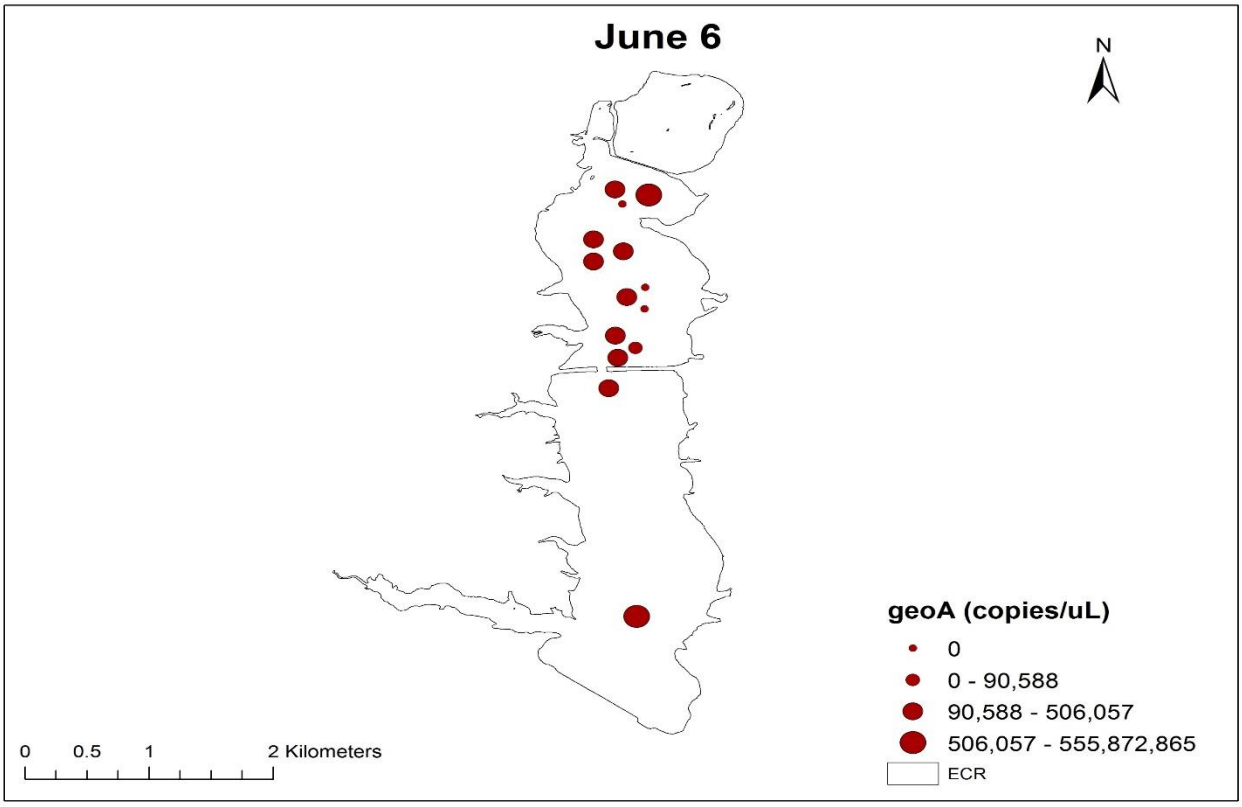


Figure 3.10 continued



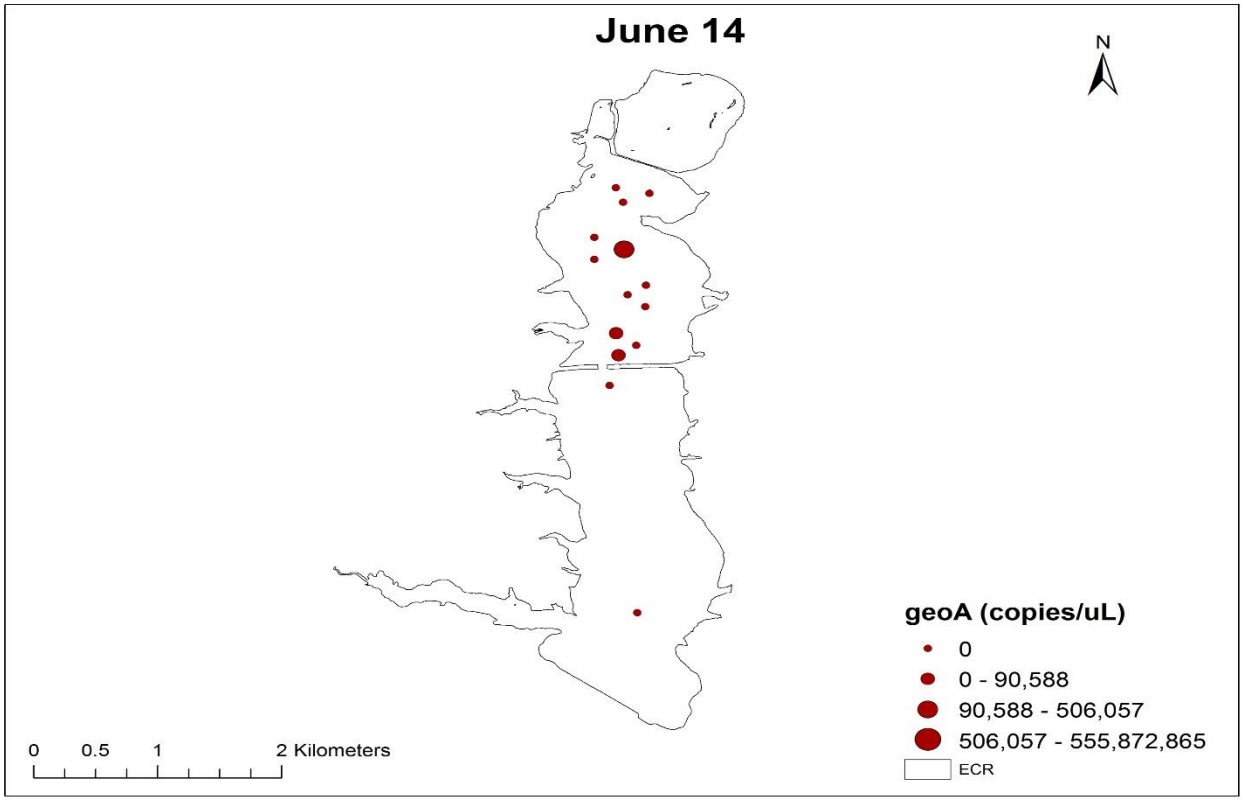
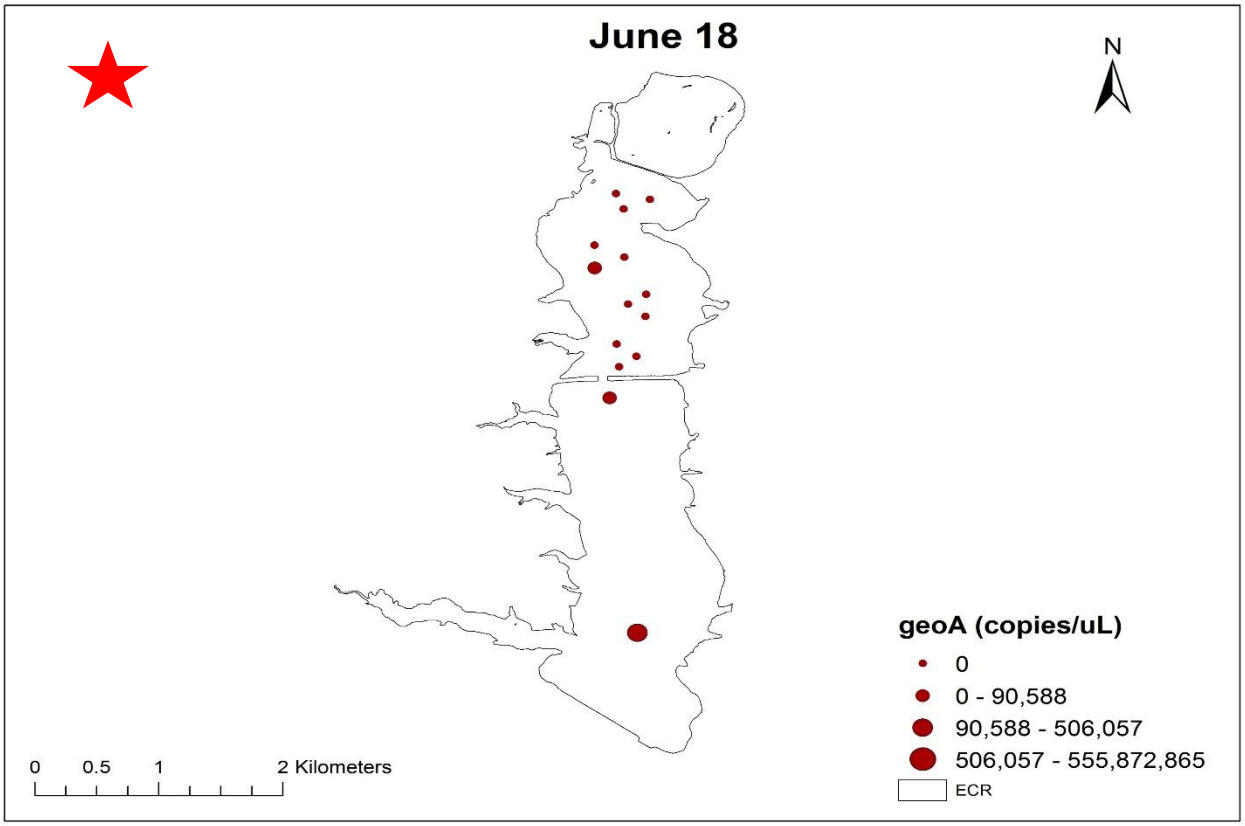


Figure 3.10 continued



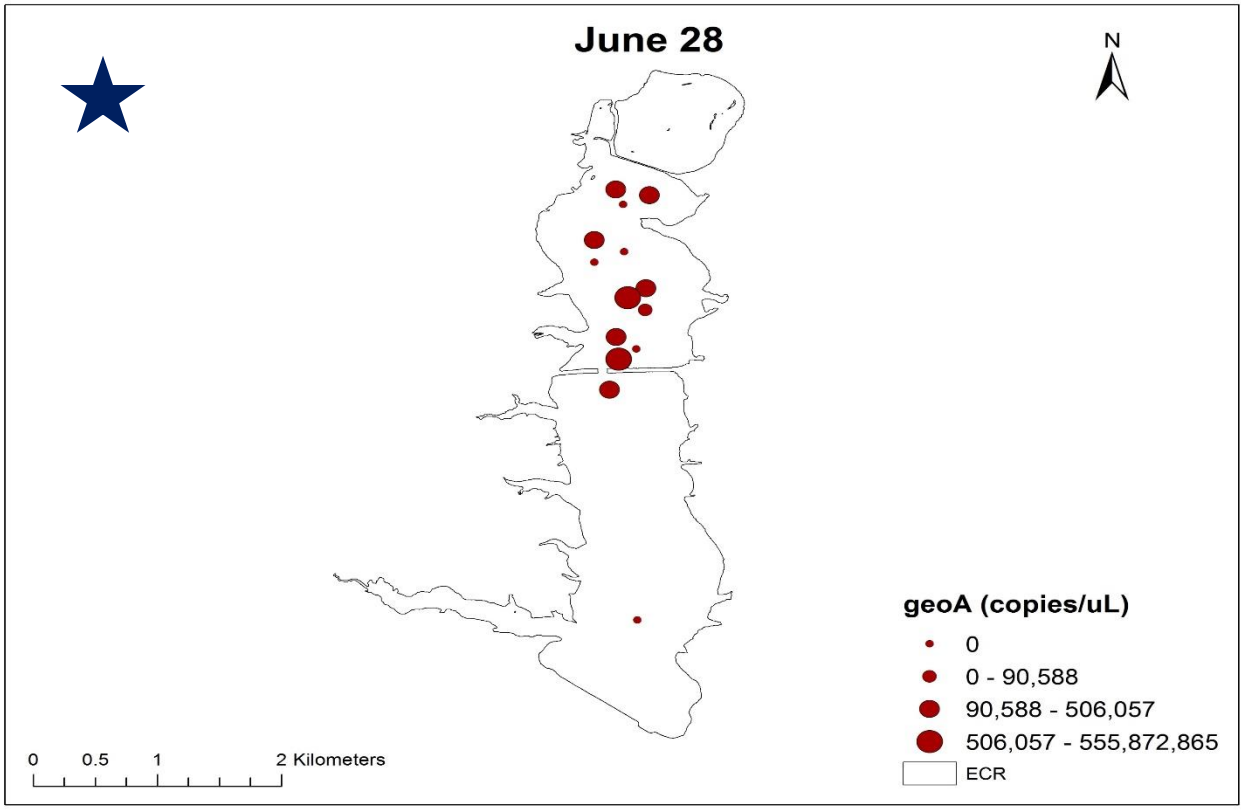


Figure 3.10 continued

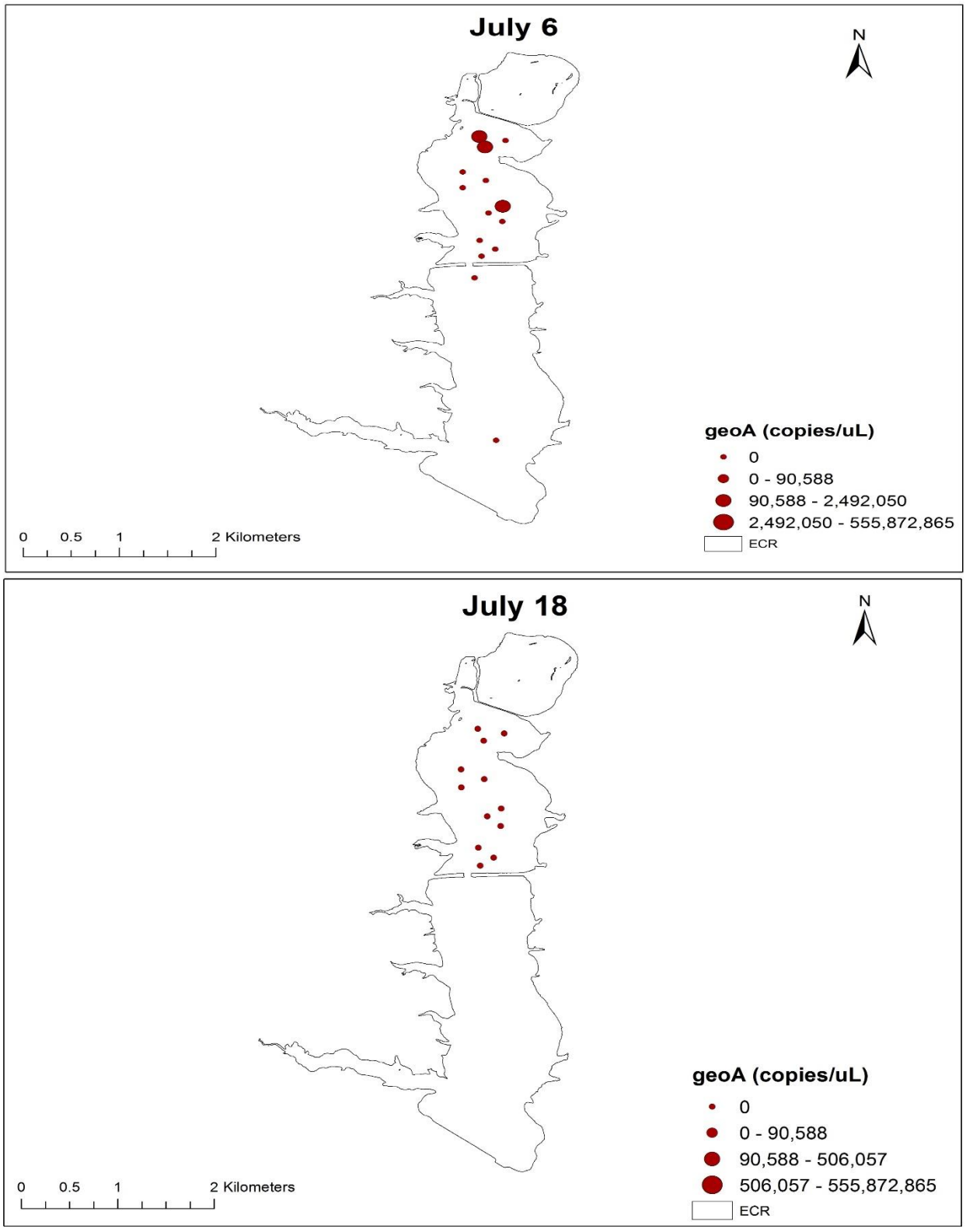
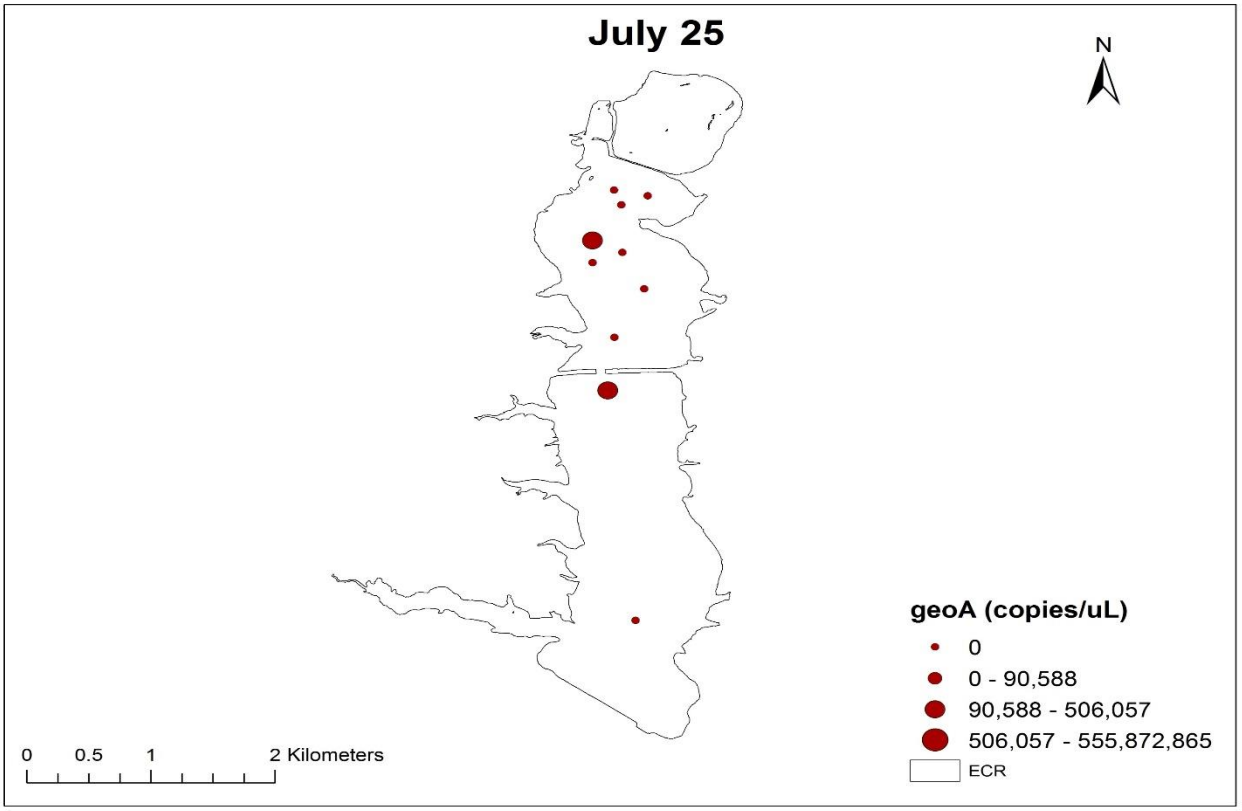


Figure 3.10 continued



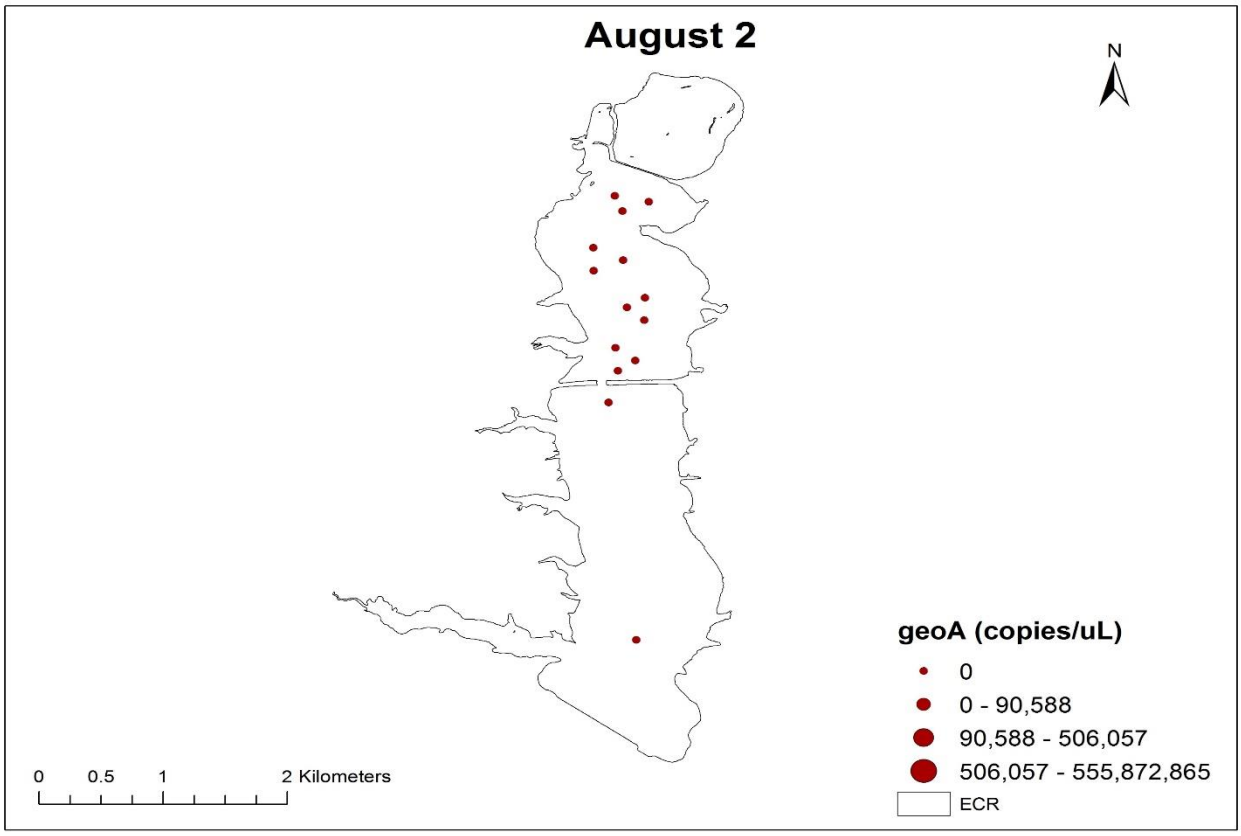


Figure 3.10 continued

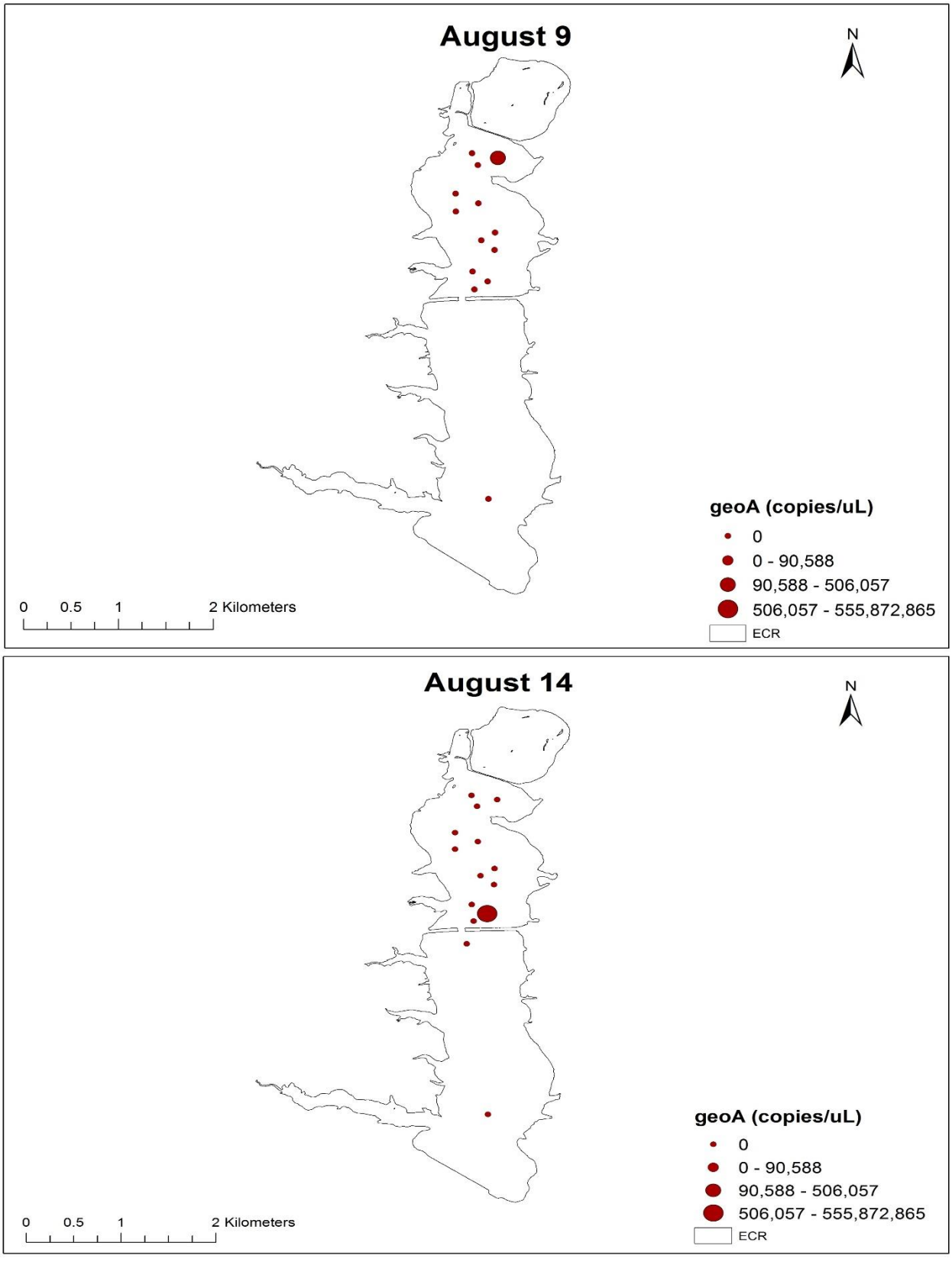


Figure 3.10 continued

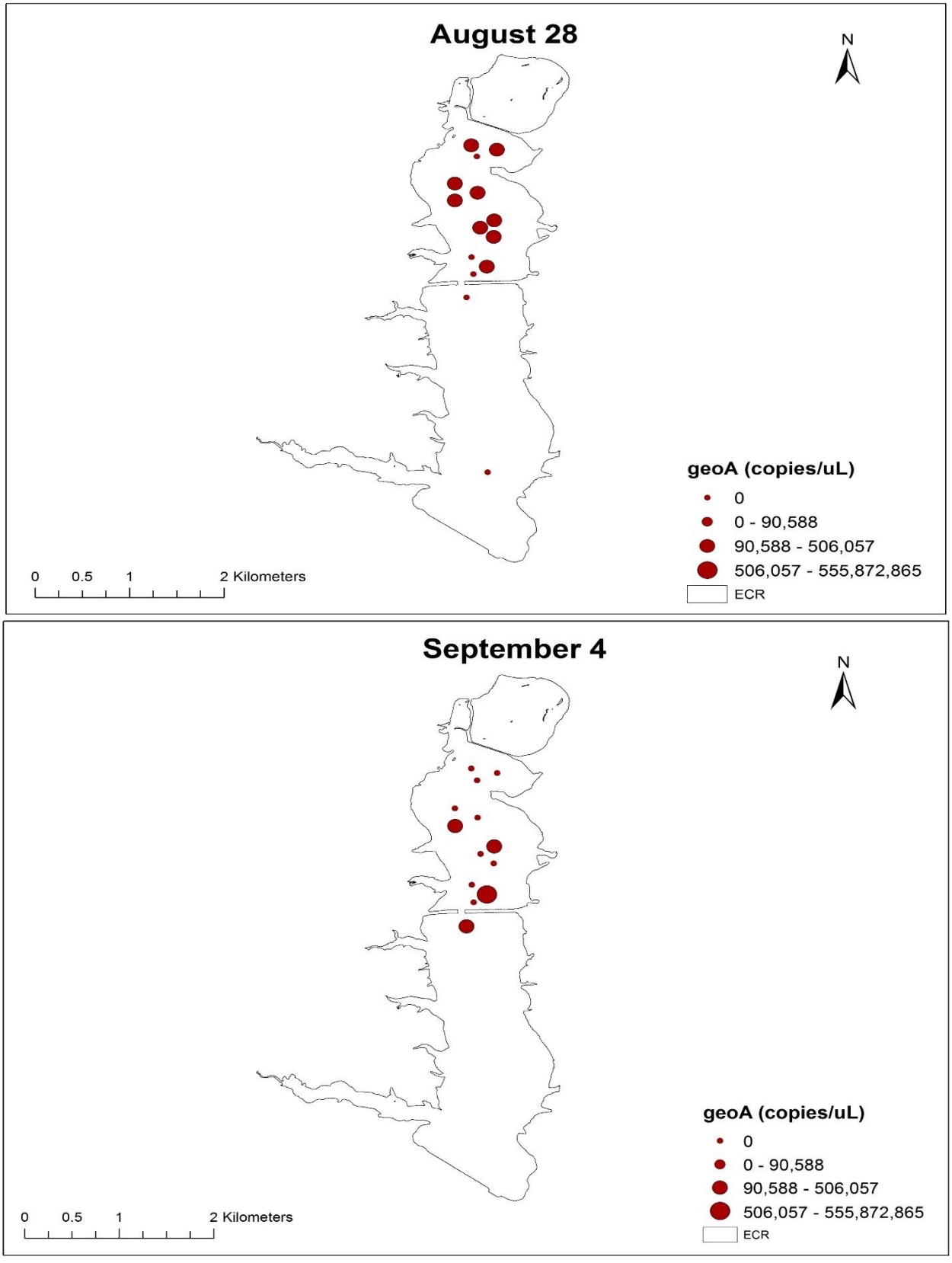
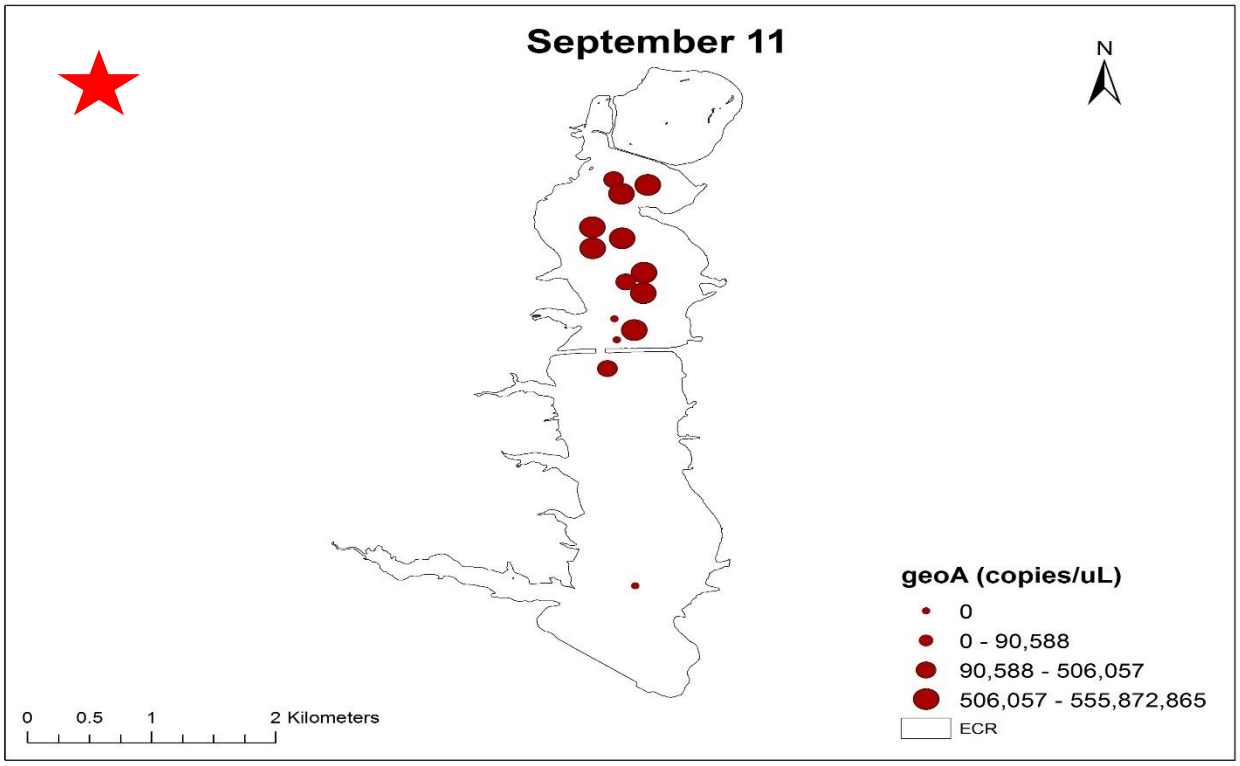


Figure 3.10 continued



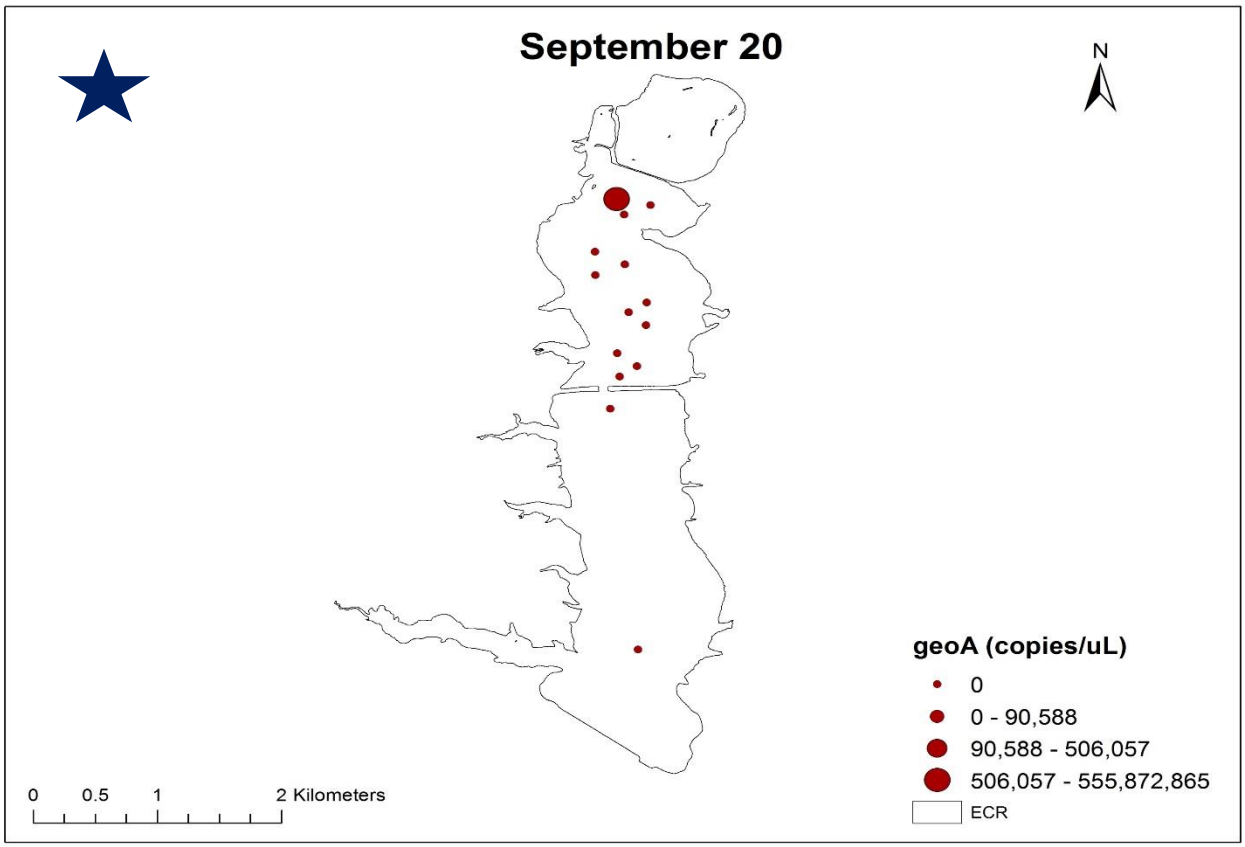
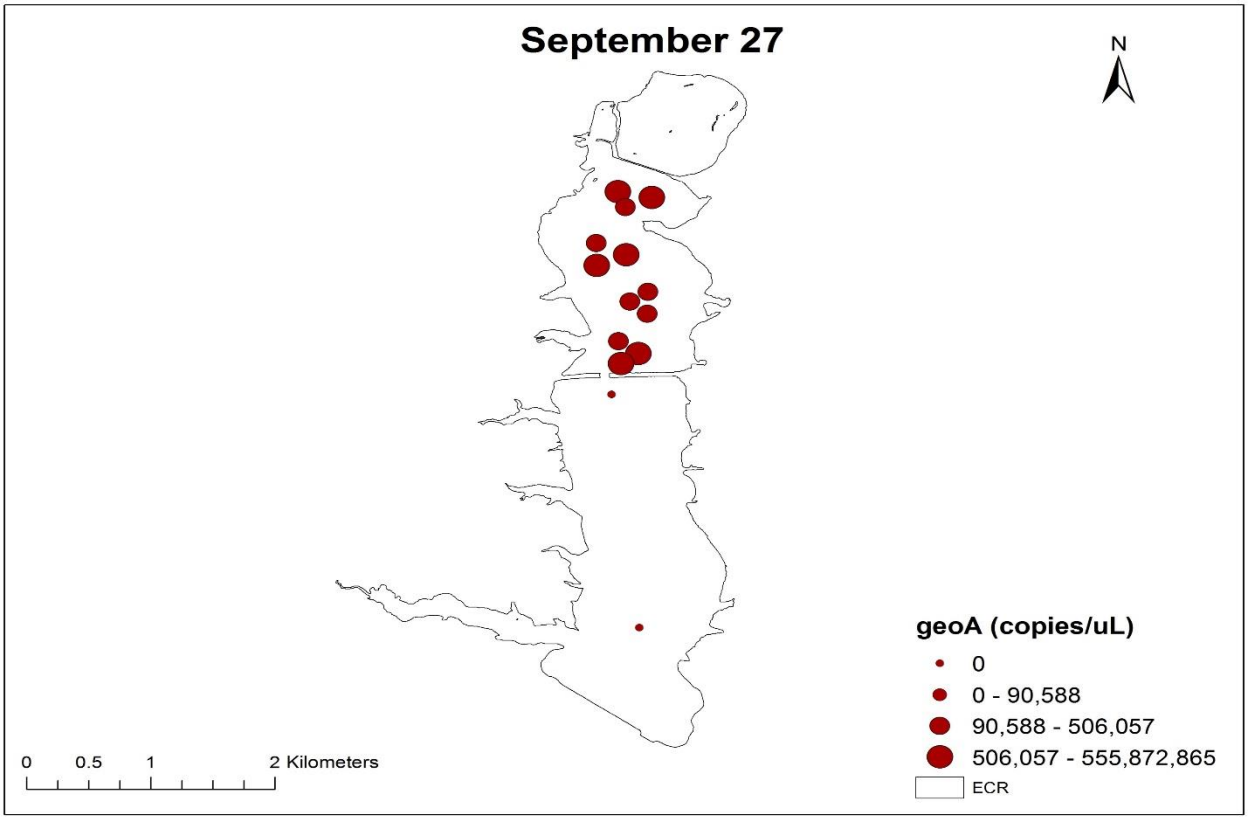
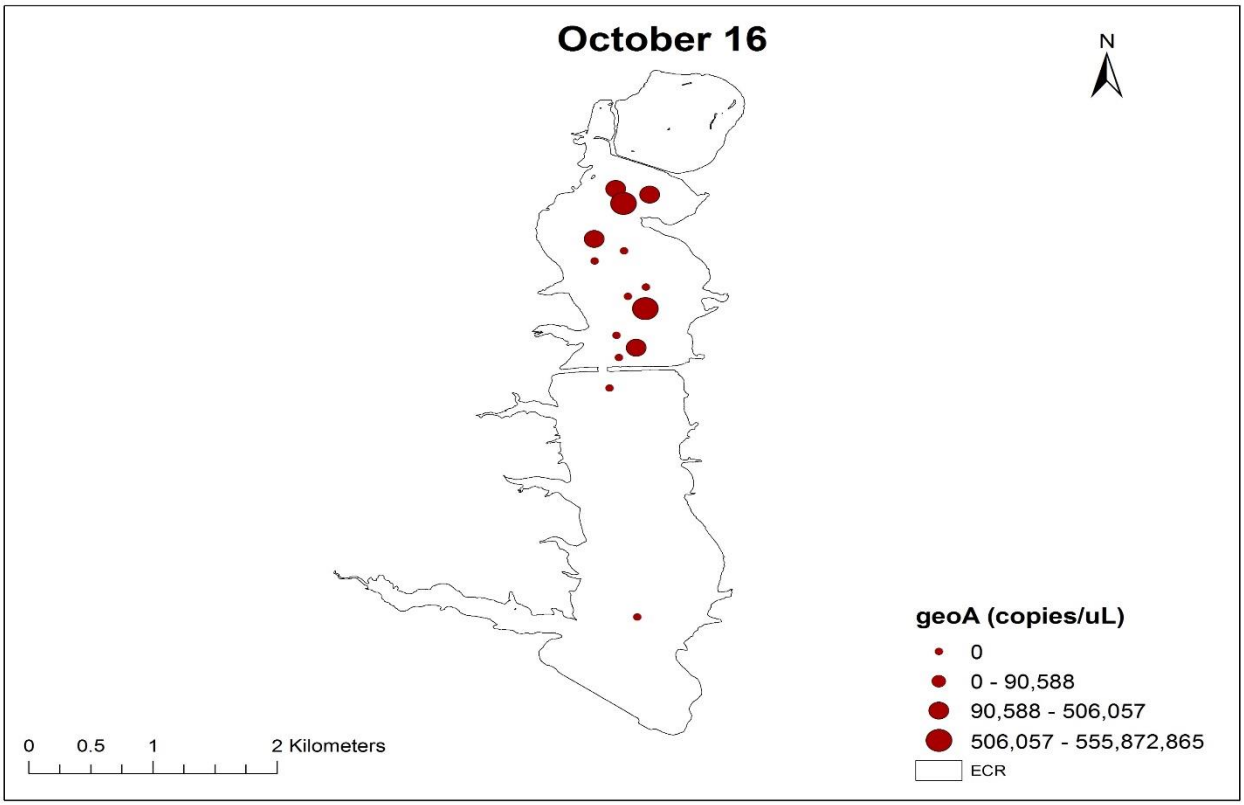


Figure 3.10 continued





3.4 Discussion

For the quantification of T&O bacterial producers in ECR, two qPCR assays were developed targeting the synthesis genes of 2-MIB and geosmin in Cyanobacteria and were applied on our water samples along with a universal 16S rRNA assay. The qPCR results were based on the generation of a standard curve, which unfortunately did not have the optimal efficiency of 90-105% [*Seashols-Williams et al.,* 2018; *Simmons et al.,* 2007]. Several different reasons can affect DNA amplification and the resulting qPCR efficiency, including primer mismatches, annealing temperatures, PCR inhibitory substances, and steric hindrance [*Brankatschk et al.,* 2012]. In our analysis, the use of degenerate primers that cause suboptimal primer-template kinetics could be one of the main reasons for the lower efficiencies [*Brankatschk et al.,* 2012; *Seashols-Williams et al.,* 2018]. In the future, qPCR assay optimization in regards of amplification conditions, concentration of input components and primer design should be made.

As expected, the highest levels of MIB and geosmin concentrations were observed during May when discharge events happen and the water column is fully mixed, and at the beginning of Fall when the reservoir destratifies. The quantities of the *geoA* and *MIBS* genes, did not always follow the fluctuation pattern of T&O compound concentrations and in some cases, they were undetectable even though geosmin and MIB levels were present. This suggests that other bacterial producers that are not detected by the qPCR assays described in this study might also contribute to the synthesis of MIB and geosmin in ECR. Spearman's correlation analysis indicated that there is a statistically significant positive correlation between the *MIBS* genes and MIB concentration. If long-term sample collection and analysis occurs, a good correlation between gene copy and concentration could be established and serve as an effective tool for monitoring emerging T&O episodes. Statistically significant positive correlation between the *geoA* genes and the corresponding geosmin concentration existed only for 2 of our sampling locations. The association between those two variables could have been affected by several factors, such as the production rate of geosmin from different organisms, their growth phase and gene copy numbers of the producing species [*Chiu et al.,* 2016]. Furthermore, the lack of correlation between the geosmin synthesis genes detected by our qPCR assay and the respective concentration, could be partially explained by the fact that although *geoA* genes might be present, they may not be metabolically active. Also, there is the possibility that our qPCR assay underestimated the *geoA* gene quantities, since the designed primers only targeted a certain range of cyanobacterial geosmin producers.

The qPCR assay that was used to estimate 16S rRNA gene quantities showed statistically significant correlation with geosmin and MIB concentrations only for a few locations and it was negative. Total cell counts did not correlate with neither MIB nor geosmin concentrations. Those results are in accordance with other studies that have demonstrated the lack of linear correlation between the synthesis genes and the T&O compound concentrations [*Wang et al.,* 2016].This highlights that even though traditional cell counting techniques are widely used by the water industry for the monitoring of T&O episodes, they are not able to distinguish between geosmin and MIB bacterial producers and non-producers as they do not take into consideration the synthetic capabilities of microorganisms [*Jüttner and Watson,* 2007; *Su et al.,* 2013].

Algaecide treatments effectively decreased the quantities of the genes detected by qPCR, with the exception of some cases where the levels appeared to increase immediately after their application. However, based on our sampling scheme the Wilcoxon Signed Rank Test showed that there was not a statistically significant median difference between the gene quantities pre and post treatment. In the future, sample collection should be made the dates right before and after each treatment application for an accurate evaluation of its impact. In regards of spatial distribution, based on the results from the Kruskal-Wallis test, which indicated that there was no statistically significant difference in the genes detected by our qPCR assays between sampling locations, over the course of the study period, we cannot identify a specific point at the reservoir where T&O problems tend to originate from and consequently guide algaecide application.

Even though the qPCR assays developed in this study successfully identified geosmin and MIB cyanobacterial producers in ECR, they do not provide information regarding the active synthetic capability of the cells. Since DNA could originate from either dead or living microorganisms, incorporation of transcriptomic approaches could provide a better understanding on which cells are metabolically active. Future research efforts should also focus on the development of qPCR assays targeting the synthesis genes of geosmin and MIB in Actinobacteria, in order to investigate their role in contributing to T&O problems. Finally, assuming that the qPCR assays will be furtherly optimized, establishment of a multiplex technique would be more cost and time effective for analysis of large sample volumes.

3.5 Conclusions

In Eagle Creek Reservoir, frequent and sometimes severe T&O episodes are observed during Spring and Fall while fewer of them are detected during the Summer months. The development of qPCR assays targeting the synthesis genes of geosmin and MIB in Cyanobacteria, successfully quantified potential T&O producers and highlighted possible associations between gene quantities and the respective concentrations. A statistically significant positive correlation was found between *MIBS* gene quantity and MIB concentration for all sampling locations, implying that this assay could potentially be used as a tool for the early prediction of upcoming T&O episodes. The *geoA* gene detection assay, did not correlate well with geosmin concentrations, suggesting that even though the gene might be present, this does not necessarily mean that it is metabolically active. As expected, correlation between the 16S rRNA gene quantity and geosmin and MIB concentrations existed only for certain locations and it was negative. This highlights that approaches used on the basis of total bacterial biomass evaluation, are not good indicators of T&O problems. Further optimization of the qPCR assays developed in this study, will strengthen their applicability in quantifying geosmin or 2-MIB producing bacteria in ECR.

3.6 References

Auffret, M., Pilote, A., Proulx, É., Proulx, D., Vandenberg, G., & Villemur, R. (2011). Establishment of a real-time PCR method for quantification of geosmin-producing Streptomyces spp. in recirculating aquaculture systems. *Water research*, *45*(20), 6753-6762.

Brankatschk, R., Bodenhausen, N., Zeyer, J., & Bürgmann, H. (2012). Simple absolute quantification method correcting for quantitative PCR efficiency variations for microbial community samples. *Appl. Environ. Microbiol.*, *78*(12), 4481-4489.

Chiu, Y. T., Yen, H. K., & Lin, T. F. (2016). An alternative method to quantify 2-MIB producing cyanobacteria in drinking water reservoirs: Method development and field applications. *Environmental research*, *151*, 618-627.

Clercin, N., (2018). Origin and fate of odorous metabolites, 2-methylisoborneol and geosmin in a eutrophic reservoir.

Du, H., Lu, H., Xu, Y., & Du, X. (2013). Community of environmental Streptomyces related to geosmin development in Chinese liquors. *Journal of agricultural and food chemistry*, *61*(6), 1343-1348.

ESRI Inc., 2019. ArcGIS 10.7. ESRI Inc., Redlands, CA.

Giglio, S., Jiang, J., Saint, C. P., Cane, D. E., & Monis, P. T. (2008). Isolation and characterization of the gene associated with geosmin production in cyanobacteria. *Environmental science & technology*, *42*(21), 8027-8032.

Giglio, S., Chou, W. K. W., Ikeda, H., Cane, D. E., & Monis, P. T. (2010). Biosynthesis of 2-methylisoborneol in cyanobacteria. *Environmental science & technology*, *45*(3), 992-998.

Harris, T. D., Smith, V. H., Graham, J. L., Van de Waal, D. B., Tedesco, L. P., & Clercin, N. (2016). Combined effects of nitrogen to phosphorus and nitrate to ammonia ratios on cyanobacterial metabolite concentrations in eutrophic Midwestern USA reservoirs. *Inland Waters*, *6*(2), 199-210.

Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M., & Visser, P. M. (2018). Cyanobacterial blooms. *Nature Reviews Microbiology*, *16*(8), 471.

John, N., Koehler, A. V., Ansell, B. R., Baker, L., Crosbie, N. D., & Jex, A. R. (2018). An improved method for PCR-based detection and routine monitoring of geosmin-producing cyanobacterial blooms. *Water research*, *136*, 34-40.

Jüttner, F., & Watson, S. B. (2007). Biochemical and ecological control of geosmin and 2-methylisoborneol in source waters. *Appl. Environ. Microbiol.*, *73*(14), 4395-4406.

Lee, J., Rai, P. K., Jeon, Y. J., Kim, K. H., & Kwon, E. E. (2017). The role of algae and cyanobacteria in the production and release of odorants in water. *Environmental pollution*, *227*, 252-262.

Li, L., Pascual, D. L., Tedesco, L. P., Randolph, K. L., Sengpiel, R. E., Hall, B. E., & Wilson, J. S. (2006). Developing a survey tool for the rapid assessment of blue-green algae in central Indiana's reservoirs. *IUPUI. Indianapolis: Central Indiana Resources Partnership/CEES*, *7*(11).

Liato, V., & Aïder, M. (2017). Geosmin as a source of the earthy-musty smell in fruits, vegetables and water: Origins, impact on foods and water, and review of the removing techniques. *Chemosphere*, *181*, 9-18.

Lylloff, J. E., Mogensen, M. H., Burford, M. A., Schlüter, L., & Jørgensen, N. O. (2012). Detection of aquatic streptomycetes by quantitative PCR for prediction of taste-and-odour episodes in water reservoirs. *Journal of Water Supply: Research and Technology-Aqua*, *61*(5), 272-282.

Nadkarni, M. A., Martin, F. E., Jacques, N. A., & Hunter, N. (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology*, *148*(1), 257-266.

Nielsen, J. L., Klausen, C., Nielsen, P. H., Burford, M., & Jørgensen, N. O. (2006). Detection of activity among uncultured Actinobacteria in a drinking water reservoir. *FEMS microbiology ecology*, *55*(3), 432-438.

R Development Core Team, R. F. F. S. C. "R: A language and environment for statistical computing." (2011): 409.

Seashols-Williams, S., Green, R., Wohlfahrt, D., Brand, A., Tan-Torres, A. L., Nogales, F., ... & Singh, B. (2018). An accurate bacterial DNA quantification assay for HTS library preparation of human biological samples. *Electrophoresis*, *39*(21), 2824-2832.

Simmons, S. L., Bazylinski, D. A., & Edwards, K. J. (2007). Population dynamics of marine magnetotactic bacteria in a meromictic salt pond described with qPCR. *Environmental microbiology*, *9*(9), 2162-2174.

Song, K., Li, L., Li, S., Tedesco, L., Hall, B., & Li, L. (2012). Hyperspectral remote sensing of total phosphorus (TP) in three central Indiana water supply reservoirs. *Water, Air, & Soil Pollution*, *223*(4), 1481-1502.

Su, M., Gaget, V., Giglio, S., Burch, M., An, W., & Yang, M. (2013). Establishment of quantitative PCR methods for the quantification of geosmin-producing potential and Anabaena sp. in freshwater systems. *Water research*, *47*(10), 3444-3454.

Suurnäkki, S., Gomez-Saez, G. V., Rantala-Ylinen, A., Jokela, J., Fewer, D. P., & Sivonen, K. (2015). Identification of geosmin and 2-methylisoborneol in cyanobacteria and molecular detection methods for the producers of these compounds. *Water research*, *68*, 56-66.

Tsao, H. W., Michinaka, A., Yen, H. K., Giglio, S., Hobson, P., Monis, P., & Lin, T. F. (2014). Monitoring of geosmin producing Anabaena circinalis using quantitative PCR. *Water research*, *49*, 416-425.

Wang, Z., Song, G., Shao, J., Tan, W., Li, Y., & Li, R. (2016). Establishment and field applications of real-time PCR methods for the quantification of potential MIB-producing cyanobacteria in aquatic systems. *Journal of applied phycology*, *28*(1), 325-333.

Watson, S. B., Brownlee, B., Satchwill, T., & Hargesheimer, E. E. (2000). Quantitative analysis of trace levels of geosmin and MIB in source and drinking water using headspace SPME. *Water Research*, *34*(10), 2818-2828.

Watson, S. B., Ridal, J., & Boyer, G. L. (2008). Taste and odour and cyanobacterial toxins: impairment, prediction, and management in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, *65*(8), 1779-1796.

CHAPTER 4. CONCLUSIONS & PERSPECTIVES

Occurrences of T&O episodes are a common phenomenon in freshwater environments, resulting in the deterioration of drinking water quality and the loss of consumers trust, who perceive water as unsafe to drink. It is hard to trace the biological origins and predict those events, as they pose a complex problem which has different intensity and frequency every year depending on varying environmental conditions and the composition of microbial community. In Eagle Creek Reservoir those episodes are usually observed during May when the reservoir receives most of its water from snow melt and rainfall and then at the beginning of Fall after thermal stratification is over. However, T&O outbreaks do not follow the same trend every year, as it was observed in years 2017 and 2018 where major peaks of T&O compound concentrations occurred during May and June respectively. The terpenoid bacterial metabolites geosmin and 2-MIB are the most common substances in T&O incidents, and the microbial capability of their production in ECR was confirmed in this study. Mapping analysis, using a metagenomic dataset from samples collected in 2013, successfully retrieved the genes of interest *geoA* and *MIBS*, responsible for the synthesis of those compounds. In addition, taxonomic analysis of samples collected in 2017 that were sequenced targeting the 16S rRNA gene, confirmed the presence of known cyanobacterial producers of geosmin and 2-MIB.

Management strategies of T&O episodes in ECR include the application of a copper-based algaecide treatment. Based on the 16S sequencing results (2017) presented in this study, the treatment was only effective towards Cyanobacteria as it drastically decreased their relative abundance in samples, and not towards Actinobacteria which remained at relatively constant levels and are considered to be major producers of geosmin and 2-MIB as well. Looking at the gene quantities detected by the qPCR assays, even though the initial hypothesis was that they decrease after the algaecide treatment, something like that was not supported by the Wilcoxon Signed Rank Test which indicated that the median difference of paired measurements before and after treatment was not significant. The only exception was the median difference for the pre/post treatment measurement of the *geoA* gene, which was statistically significant for the second and third treatment, but did not show a consistent trend, making the drawing of conclusions hard. Since sample collection was not performed the day before and right after each treatment, both biotic and abiotic factors could have interfered, impacting the gene levels and preventing an accurate

evaluation of the pre and post algaecide treatment behaviour. In the future, sample collection should be arranged right before and after the algaecide application so that we can better evaluate the microbial population response to it. As an alternative treatment approach, the incorporation of biodegradation methods could be an option, since the presence of bacteria having the ability to breakdown geosmin and 2-MIB was confirmed in this study.

The argument that a small fraction of the total bacterial population contributes to the production of T&O compounds in ECR was supported by several methods of analysis for samples collected during different years. Based on the mapping analysis of the 2013 metagenomic dataset, less that 1% of the sample reads mapped to reference sequences representing the *geoA* and *MIBS* genes. Furthermore, the relative abundance of Cyanobacteria and Actinobacteria in samples that were collected during 2017 and analysed through 16S sequencing, did not correlate with the concentrations of geosmin and 2-MIB. Likewise, the 16S rRNA gene quantities detected by qPCR and Cyanobacterial cell counts did not correlate with those concentrations either, highlighting the importance of designing diagnostic tools targeting specific T&O bacterial producers and not the microbial community as a whole.

In consideration of this, qPCR assays were developed in this study targeting the *geoA* and *MIBS* genes in Cyanobacteria. Correlation analysis showed a positive correlation between the *MIBS* gene quantities and the respective 2-MIB concentration for all sampling locations, indicating that this assay could be used as an accurate diagnostic tool for the early detection of emerging T&O events. In contrast, a positive correlation between the *geoA* gene quantities and geosmin concentrations existed only for 2 of our sampling locations, suggesting that even though gene detection was successful, they might not be expressed in those organisms. Since the efficiency of those assays was lower than the ideal range of 90-105%, further optimization in terms of primer design, annealing temperatures and the concentration of each reaction component should be made in the future to achieve better quantification of those genes. Another limitation includes the high detection limits of the *geoA* assay, that were probably caused by the use of DNA extracted straight from the *Nostoc sp.* culture, which could have multiple copies of the same gene, resulting in an incorrect calculation of the gene copy number used to generate the standard curve. Furthermore, horizontal gene transfer between bacteria in freshwater environments should be investigated and taken into consideration in the future, since the presence of identical genetic material in species

belonging to different bacterial classes, can lead to false positive results and inaccurate identification of the biological source of the problem.

In terms of spatial distribution, the initial expectation was that T&O problems in ECR tend to originate from the north side where the input of minor streams is located, and then spreading to the rest of the reservoir. This trend was not depicted by the spatial maps constructed for the gene quantities measured by qPCR that indicated a random pattern of fluctuation, and it was not supported by the statistical analysis either which showed that there is no statistically significant difference between sampling locations. As a result, a specific point acting as source of the problem, which could potentially guide treatment application, could not be identified. This does not necessarily exclude the possibility that there are certain locations in the reservoir where T&O episodes tend to originate from, since our sampling scheme might not have captured this pattern.

In addition to qPCR assay optimization, the development of a multiplex approach and the design of primers targeting the genes of interest in Actinobacteria, should become an objective for future research. This will enable as to get a better understanding of the biological origin of geosmin and 2-MIB production, and also develop accurate diagnostic tools with shorter analytical time which could be used for on-site monitoring of T&O episodes. Furthermore, since DNA does not indicate whether genes are expressed or not, incorporation of transcriptomic approaches is essential for the evaluation of active synthetic capabilities of the targeted bacterial groups. Finally, strain isolation of Cyanobacterial and Actinobacterial species could also contribute to the identification of microbial T&O producers and consequently to the development of diagnostic tools targeting those producers.