MOLECULAR CHARACTERIZATION OF POTENTIAL GEOSMIN-PRODUCING CYANOBACTERIA FROM LAKE ONTARIO

by

Andrea Gill

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada, 2006

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the general public.

ABSTRACT

Geosmin is an odorous secondary metabolite produced by some cyanobacteria during
growth and released from the cells. Little is known about the biosynthesis of geosmin and
the gene(s) required for its production have not been characterized. During late August and
early September geosmin episodes due to planktonic cyanobacteria frequently occur in the
northwest basin of Lake Ontario waters resulting in taste and odour episodes in drinking
water that serves more than 5 million people. At high concentrations geosmin evades
traditional drinking water treatment and reaches the tap. These episodes often elicit
consumer concern and are wrongly construed to reflect impaired drinking water safety.
Water quality managers in the region have generally been unable to prevent or control taste
and odour episodes via a proactive approach due to the lack of knowledge of
cyanobacterial communities in offshore waters as well as the inability to predict when
geosmin will reach intake pipes due to downwelling, the process by which the surface
waters mix with the hypolimnion. This study evaluated denaturing gradient gel
electrophoresis (DGGE) as a molecular tool for proactive monitoring of potential taste and
odour-causing cyanobacteria in environmental samples. The 16S rRNA gene was assessed
for its ability to distinguish among geosmin-producing and non-producing strains. This
study also examined the evolutionary relationships among geosmin-producing
cyanobacteria using the full-length 16S rRNA gene and compared phylogenies with current
taxonomy.

A DGGE standard using the V3 hypervariable region of the 16S rRNA gene was
developed using geosmin-producing and non-producing isolates of cyanobacteria. Included

in the standard was the suspected primary contributor to Lake Ontario taste and odour, *Anabaena lemmermannii* Richter. This standard was then applied to various environmental collections from Lake Ontario (August 2005) to examine the cyanobacterial community composition. DGGE profiles were consistent with the presence of *An. lemmermannii* at locations with increased geosmin concentrations (determined using gas chromatography-mass spectrometry), supporting hypothesis that *An. lemmermannii* is the primary contributor to northwestern Lake Ontario taste and odour. In addition, the application of DGGE in the identification of potential geosmin-producing species of cyanobacteria was deemed to be a potentially useful approach to monitoring cyanobacterial communities in source waters. The 16S rRNA-V3 region alone did not distinguish among geosmin-producing and non-producing strains, however with additional data (actual geosmin concentration) it was showed relationships.

In the phylogenetic analyses, geosmin-producing cyanobacteria did not group monophyletically and it was not possible to state that a single evolutionary event has led to the acquisition of the geosmin-producing trait. Phylogenies also showed that the taxonomy of the Cyanobacteria is largely unresolved. All five Sections (bacteriological classification)/four orders (Komárek & Anagnostidis classification) were paraphyletic, however the heterocystous cyanobacteria (Sections IV and V/Nostocales and Stigonematales) grouped separately from the non-heterocystous cyanobacteria (Sections I, III/Chroococcales and Oscillatoriales). Although both systems of classification compared in this study were similar, nomenclature and groupings were occasionally different among the groups. This demonstrates the incongruity between bacteriologists and phycologists and emphasizes the need for a consensus system of classification for the Cyanobacteria.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

GENERAL INTRODUCTION

1.1 DRINKING WATER QUALITY AND CYANOBACTERIAL BLOOMS

Water quality is a concern for many countries worldwide. In recent years, water quality

has become an issue in Canada, particularly with the Walkerton incident in southern

Ontario (Auld *et al.*, 2004). In Ontario, the primary freshwater source for drinking water is

Lake Ontario, providing drinking water to over five million people. Accordingly, the

quality of water derived from this source must be of acceptable quality to its consumers.

The process of generating potable water from a source such as Lake Ontario is complex

and several factors can influence the quality of the final product. Environmental pollutants

such as polychlorinated biphenols (PCBs), oil, pesticides, as well as detergents, domestic

sewage, industrial wastes, and storm water runoff are a few of the common contaminants

that can have negative effects on source waters (Mason, 1996). Excess nitrogen and

phosphorus from detergents, fertilizers and sewage waste can lead to eutrophication in

source waters. General effects of eutrophication include: i) decreased species diversity in

aquatic foodwebs, ii) increased plant and animal biomass, iii) increased turbidity, iv)

shortened life span of the lake due to increased rate of sedimentation, and v) anoxic

conditions (Mason, 1996). The extreme conditions typified in a eutrophic lake often

promote the development of opportunistic taxa and blooms of nuisance algae, primarily

cyanobacteria that can severely impact water quality as well as the perception of good

water quality (Graham and Wilcox, 2000).

Blooms of cyanobacteria often result in the accumulation of unsightly, malodorous, mats of cells at the surface and shoreline of lakes or ponds (Downing *et al.*, 2001). These blooms can impair water bodies for recreational use (Watson, 2004), decrease water transparency preventing growth of plants and algae, and during their decay, result in low oxygen levels resulting in death of fish and other aquatic organisms (Graham and Wilcox, 2000). Occasionally, toxins are produced and released into the waters by some cyanobacteria and can cause unsafe waters for recreational use as well as for drinking. Cyanobacteria such as species of *Microcystis*, *Anabaena*, and *Aphanizomenon* are capable of producing hepatotoxins and neurotoxins that can cause human illness, fish kills and livestock deaths and are a growing concern in Lake Ontario (Murphy *et al.*, 2003).

1.2 *LAKE ONTARIO TASTE AND ODOUR: GEOSMIN*

Another major consequence of algal blooms dominated by cyanobacteria can be the production of secondary metabolites known as algal volatile organic compounds (AVOCs) that produce taste and odour in drinking water. One of the most frequently reported odours in northwestern Lake Ontario is an earthy, muddy, musty smell commonly caused by geosmin (*trans*-1, 10-dimethyl-*trans*-9-decalol) which is a terpenoid metabolite released by certain members of the Cyanobacteria and Eubacteria (Jüttner, 1995) (Figure 1). Geosmin is resistant to oxidation and at high concentrations it evades conventional drinking water treatment resulting in undesirable taste and odour in tap water (Nakamura *et al.*, 2001). Problematic taste and odour can also be generated in the absence of conspicuous blooms, for example by highly prolific species present at low-moderate abundances, or by benthic or subsurface populations. These episodes in particular are difficult to anticipate, trace and control.

The suspected primary contributor to northwestern Lake Ontario taste and odour is the cyanobacterium *Anabaena lemmermannii* Richter (Watson, 2004). Peak geosmin concentrations occur in offshore waters annually in late August and early September when the lake is stratified and there is a distinct thermocline in the upper 30 m. During this time some geosmin-producing cyanobacteria such as *An. lemmermannii* increase in abundance. The bulk of geosmin production is retained in the cells, but when they breakdown or are ruptured via physical processes such as grazing or treatment, intracellular geosmin is released into the water, thereby increasing the concentration of the compound significantly (Watson, 2004). The natural degradation of geosmin has been observed by co-operatively acting bacteria; *Sphingopyxis alaskensis*, *Novosphingobium stygiae*, and *Pseudomonas veroni*i, however in the absence of geosmin degrading organisms, the compound persists for a greater length of time (Hoefel *et al.*, 2006). Surface water temperature variations along the northwest shore of Lake Ontario during the summer are periodically affected by upwelling and downwelling. Studies suggest that downwelling induced by strong easterly winds results in the movement of offshore waters into the shoreline and offshore planktonic cyanobacteria and geosmin are transported to deeper inshore water treatment facility intakes (Skafel and Yerubandi, 2003) (Figure 2). Higher geosmin concentrations thus can reach intakes and often evade standard drinking water treatment such as sand sedimentation and sand filtration. These processes are largely ineffective at removing many dissolved AVOCs, additionally, carbon beds and advanced filtration technology performance is substantially reduced in the presence of high concentrations of dissolved organic material (Watson, 2004).

Figure 1. The molecular structure of geosmin (*trans*-1, 10-dimethyl-*trans*-9-decalol).
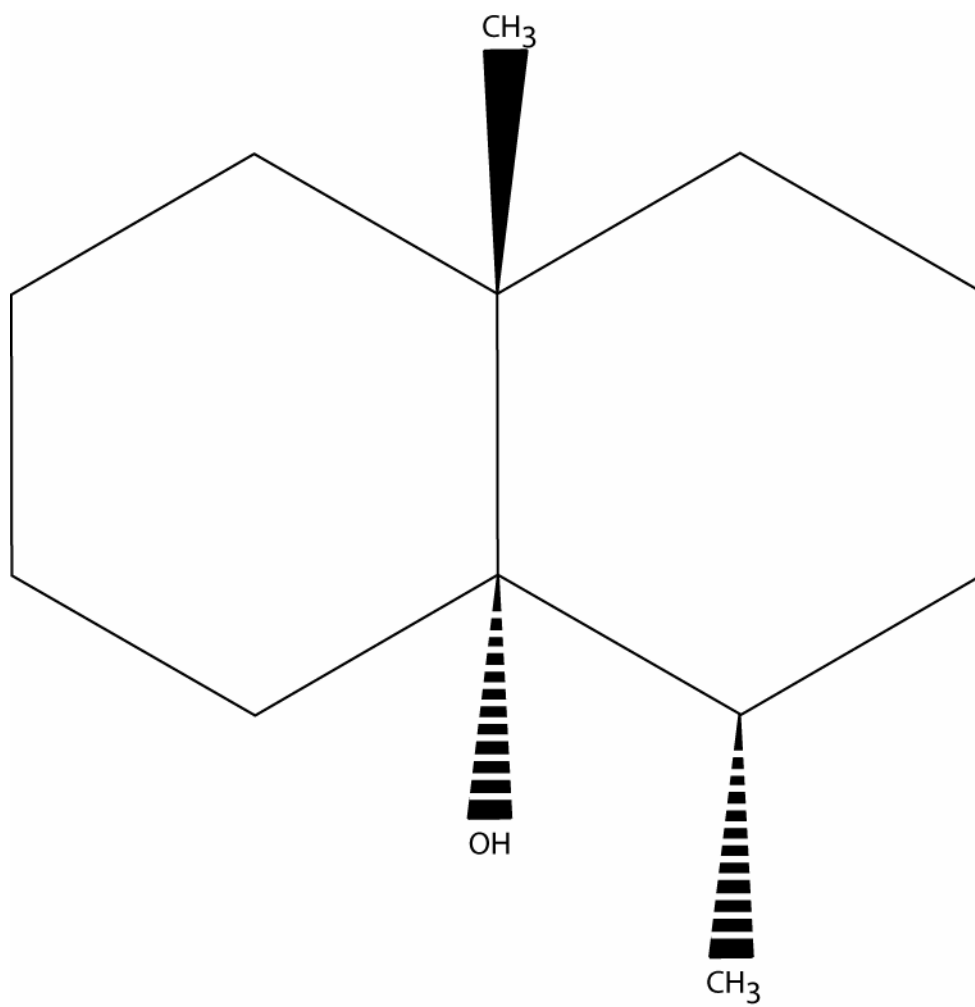
Figure 2. Downwelling caused by offshore winds brings geosmin in the warmer surface waters to water treatment facility intakes. High geosmin concentrations can evade drinking water treatment and reach consumer taps.

Onshore Wind

Warm/Geosmin

To intakes

Water treatment plant Intake
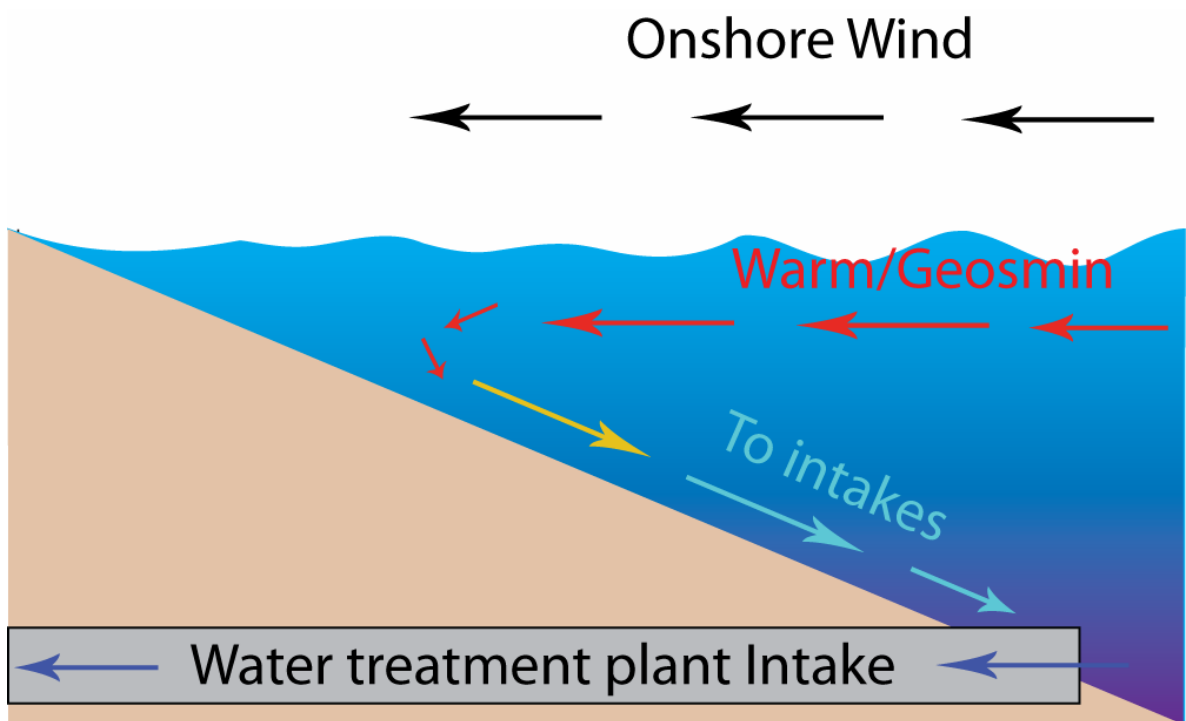
complaints and a negative effect on consumption patterns (Watson, 2004). Health is not affected by the consumption of geosmin however consumers often associate poor taste and odour in drinking water with compromised safety due to a lack of public education during these events. It is important that the public is informed on all aspects of water quality including taste and odour as well as toxins in order to prevent health concerns, complaints and loss of confidence.

Cyanobacteria are the chief geosmin producers in surface waters and are of primary concern when managing taste and odour episodes. The majority of geosmin-producing cyanobacteria are planktonic, filamentous N_2 fixers or non-fixers (Watson, 2004). Many studies suggest that geosmin production is not a species-specific trait and the only way to confirm the producing strain, is by isolation and morphological identification after each taste and odour event. The ambiguity associated with morphological identification of geosmin-producing species may originate from species misidentification due to phenotypic plasticity exhibited in lab cultures and under different environmental conditions (Watson, 2004).

Several studies suggest that geosmin production by cyanobacteria is not directly correlated with specific environmental factors. Iron (Nakashima and Yagi, 1992), temperature, nitrogen, phosphorus, light (Saadoun *et al.*, 2001) and copper (Dionigi, 1995) have shown varying relationships with geosmin production in some species under laboratory conditions. It appears that for cyanobacteria, increased light intensity and non-optimal temperatures favours geosmin production by *Anabaena* sp. (Jüttner, 1995; Saadoun *et al.*, 2001). Limited iron conditions tend to increase geosmin release by some cyanobacteria into growth media (Jüttner, 1995) and nitrogen in the form of dinitrogen (N_2)

rather than nitrate (NO_3) tends to increase geosmin production. Increased copper has resulted in reduced geosmin levels in several genera of cyanobacteria and bacteria due to its inhibitory effect on overall growth (Dionigi, 1995; Saadoun *et al.*, 2001). These experiments that have tested various environmental conditions against geosmin production have shown that geosmin has neither a constant ratio to protein nor to chlorophyll-a content suggesting that it does not directly correlate to cell biomass. The range of variation of this metabolite was limited and has not exceeded a factor of two to three, indicating that there may be some limitations to its production by the cell (Jüttner, 1995). Geosmin was hypothesized to be an overflow product of pigment production. However, Jüttner (1995) has suggested that it may be associated with cell energetics and affected by factors that influence growth. It is not known whether geosmin production is constitutive or induced and more research is needed to detemine its mode of synthesis (Jüttner, 1995).

1.3 THE BIOCHEMICAL PATHWAYS FOR GEOSMIN BIOSYNTHESIS

In addition to a lack of knowledge of the factors that influence geosmin production by cyanobacteria, little is known about the metabolism of geosmin in cyanobacteria as well because the biosynthetic pathway has not yet been fully elucidated. Geosmin is a terpenoid metabolite derived from isoprenoid biosynthetic pathways. Terpenoids are composed of an isoprene skeleton, hence they are metabolites derived from isoprenoid pathways. They are the largest known groups of natural products that are necessary for numerous biochemical functions such as: quinines in electron transport chains, prenyl-lipids and sterol components of membranes, carotenoid and side chains of chlorophyll in photosynthetic pigments as well as plant defense compounds (monoterpenes, diterpenes and sesquiterpenes) (Lange *et al.*, 2000).

Isopentenyl diphosphate (IPP) is the central intermediate of isoprenoid biosynthesis and is produced via two possible routes: the mevalonate (MVA) and the methylerythritol (MEP) (also called the non-mevalonate) pathway (Lange *et al.*, 2000). It has been observed that all organisms utilize one or both of the two pathways however the exact route to geosmin subsequent to IPP production has not been completely elucidated. In the MVA pathway, acetyl-CoA is the precursor to IPP (Figure 3A) whereas the MEP pathway uses pyruvate-derived D-1-deoxyxylulose-5-phosphate (DXP) and 2-C-methyl-D-erythritol-4-phosphate (MEP) to synthesize IPP (Figure 3B) (Cane *et al.*, 2001). An overwhelming majority of eubacteria, including the cyanobacterium *Synechocystis* sp., possess putative genes involved in the MEP pathway for isoprenoid production and are postulated to produce IPP via this route. *Synechocystis* sp. however is the only cyanobacterium that has been investigated for these genes. A review by Eisenrich (2004) summarized that *Synechocystis* sp. possesses genes only of the MEP pathway, and hypothesized that that the production of geosmin occurs via the MEP pathway. Radiolabelling experiments carried out by Ershov et al., 2002 however showed that although *Synechocystis* sp. PCC 6803 possesses genes of the MEP pathway, it is not likely that this is the route to IPP. Their data supported alternative routes of entry of pentose phosphate cycle substrates derived from photosynthesis rather than entry from glycolysis (Figure 3C) (Ershov *et al.*, 2002). More studies are needed to elucidate the exact route to IPP formation in cyanobacteria before it is possible to infer the pathway in which geosmin is produced. It seems probable that homologous genes from the MEP pathway will be discovered in more species of

Figure 3. Biosynthesis of IPP, the route to isoprenoids and subsequently geosmin via the mevalonate pathway (modified from Lange *et al.*, 2000) (A), the methylerythritol phosphate pathway (modified from Ershov *et al.*, 2002) (B), and the hypothetical Pentose Phosphate pathway (modified from Ershov *et al.*, 2002) (C). Precursor metabolites are derived from the citric acid cycle in A, glycolysis in B, and sugars from photosynthesis in C.

A	B	C
Mevalonate Pathway	*Methylerythritol phosphate Pathway*	*Pentose phosphate Pathway*

A — Mevalonate Pathway

Acetyl-CoA

↓

Acetoacetyl-CoA

↓

HMG-CoA

↓

Mevalonate

↓

Mevalonate-5-phosphate

↓

Mevalonate-5-diphosphate

B — Methylerythritol phosphate Pathway

Pyruvate + D-Glyceraldehyde-3-phosphate

↓

1-Deoxy-D-xylulose-5-phosphate

↓

2-*C*-Methyl-D-erythritol-4-phosphate

↓

4-(Cytidine 5'-diphospho)-2-*C*-methyl-D-erythritol

↓

2-Phospho-4-(cytidine 5'-diphospho)-2-*C*-methyl-D-erythritol

↓

2-C-Methyl-D-erythritol-2,4-cyclodiphosphate

C — Pentose phosphate Pathway

Sugar Phosphates from photosynthesis

↓

Ribose-5-phosphate + Xylulose-5-phosphate

↓

Diphosphocytidyl hydroxymethyl D-erythritol

↓

2,4 hydroxymethyl cyclodiphosphate

All three pathways converge →

Isopentenyl diphosphate (IPP)

↓↓

Dimethylallyl diphosphate (DMAPP)

↓

ISOPRENOID END PRODUCTS (GEOSMIN)

cyanobacteria as research continues however it does not necessarily indicate that this pathway is utilized.

To date, only a few organisms have been examined for the route of IPP to geosmin. Eubacteria, such as *Streptomyces* sp., are postulated to produce farnesyl diphosphate (FPP) from IPP and subsequently 1(10), 4-germacradien-11-ol that leads to the final geosmin product. gene that has been identified for the final conversion in the MEP or MVA pathways. A sesquiterpene cyclase (c*yc*2) that is responsible for the cyclization of farnesyl pyrophosphate (FPP) to the germacra-1(10)E,5E-dien-11-ol precursor has been characterized in *Streptomyces coelicolor* and has been determined to be essential for the production of geosmin (Gust *et al.*, 2003). Subsequent to this conversion, three more steps are involved for the final geosmin product, however these genes have not been characterized in the genome of any organism and the enzyme responsible for the final conversion is not known. Genes flanking *cyc*2 do not appear to code for enzymes likely to be involved in the final steps of geosmin production, suggesting that the remaining unexplored genes may be scattered throughout the genome (Gust *et al.*, 2003). Geosmin metabolism is very complex and many venues still need to be explored before the role of geosmin in cyanobacteria can be determined.

1.4 GEOSMIN DETECTION AND MANAGEMENT IN WATERS

Water quality managers have generally been unable to prevent or control taste and odour episodes of geosmin-producing cyanobacteria via proactive action prior to episodes due to a lack of knowledge of the cyanobacterial communities in the waters as well as the inability to predict when geosmin will reach intake pipes. A proactive approach to controlling taste and odour in drinking water sources by monitoring cyanobacterial species composition in

water samples for potential taste and odour producers is an option that generally has not been employed by water managers. Characterization of species composition in a water body could provide an advanced warning of the presence of potential taste and odour-causing species in source waters. This would enable water quality managers to prepare for taste and odour events and water treatment could in turn be optimized. In the long term, identification of cyanobacterial taxa that occur in source water blooms could contribute towards a more robust management of drinking water through corrective or preventative measures (Watson, 2004).

Current technology for geosmin detection includes gas chromatography-mass spectrometry (GC-MS) that allows for a highly sensitive measurement of metabolites as low as parts per trillion (ppt) levels. This method however can necessitate large sample volumes and intensive sample concentration procedures such as liquid-liquid extraction, closed-loop stripping analysis (requiring complex equipment), simultaneous distillation extraction, or purge and trap, all of which can result in low sample throughput due to lengthy protocols (Watson *et al.*, 2000). High-resolution mass spectrometers are required for detection at $ng \cdot l^{-1}$ concentrations are extremely costly (Miller *et al.*, 1999).

Sensory analysis of geosmin by human assessment is a simple method for detection but it relies upon sensing capabilities of the individual and has several limitations with the lack of quantification being paramount. Hence, sensing for geosmin is strictly qualitative and is not a suitable technique for measuring concentrations in drinking water sources. Enzyme-linked immunosorbent analysis (ELISA) uses antibodies to detect geosmin and provides a rapid field-test for geosmin detection however it is costly and its detection levels are too high to be of any practical value ($1\mu g \cdot l^{-1}$) (Chung *et al.*, 1991).

Direct control of cyanobacteria populations using algicides is another approach to preventing taste and odour in source waters however it has been very difficult to achieve due to negative impacts on surrounding ecosystems. Copper-based synthetic products have been added to waters to inhibit cyanobacterial growth in some countries (Australia and USA), however there have been several issues of concern that include: i) broad-spectrum toxicity of copper towards phytoplankton, resulting in the death of the entire community and subsequent low dissolved oxygen levels that can kill other aquatic organisms; ii) potential toxicity of copper towards non-target organisms; iii) negative public perception of the use of synthetic compounds in the environment; and vi) environmental safety issues (Schrader *et al.*, 2004). This approach is not sustainable or ecologically sound and has not been employed for controlling cyanobacterial populations in Lake Ontario and most other aquatic systems.

Algicides can be designed for a narrow ranged control measure for geosmin-producing cyanobacteria that can be effective to a species level. SeaKleen© is an example of a species-specific algicide that contains the active ingredient menadione sodium bisulfite (MSB), a water-soluble derivative of menadione (also known as vitamin K3) (Schrader *et al.*, 2004). This algicide is selective towards *Oscillatoria perornata*, a known geosmin producer and has been successful in decreasing population levels in laboratory experiments while having no significant effects on diatoms (Bacillariophyceae) or green algae (Chlorophytes) (Schrader *et al.*, 2004). This may be a useful agent for controlling certain taste and odour episodes, however it does not target all geosmin-producing cyanobacteria and at the present time, it solely affects *Oscillatoria perornata*, other susceptible strains have not been identified. More importantly, this represents a short-term measure of control,

which would need to be repeatedly applied, and the effects of these chemicals on other components of the food web and the ecosystem as a whole have not been investigated. Constant inhibition of geosmin-producing cyanobacterial populations may result in negative impacts on the surrounding ecosystem and does not represent an ecologically sound solution for taste and odour control in drinking water. There is no cost effective and rapid technique applied in water treatment management for early detection of geosmin-producing cyanobacteria and as a result, problems arising from cyanobacteria are acted upon after their occurrence.

Monitoring cyanobacterial populations in water samples to predict geosmin episodes is a proactive approach that has not been employed by most water managers. The characterization of species composition in a water body could provide a profile of potential taste and odour-causing strains present in samples. Water quality managers could prepare for taste and odour events and water treatment could in turn be optimized. In the long term, identification of problematic cyanobacterial taxa that occur in source water blooms could contribute towards a more robust management of drinking water through corrective or preventative measures (Watson, 2004).

1.5   *PCR IDENTIFICATION OF GEOSMIN-PRODUCING CYANOBACTERIA*

Identification of cyanobacteria to a species level using microscopy is extremely time-consuming and difficult due to overlapping characteristics and phenotypic plasticity exhibited by various species. Microscopy would not be a practical approach to identifying potential geosmin-producing species in drinking water sources for predicting and/or preventing taste and odour episodes, however molecular methods for early detection of geosmin-producing cyanobacteria in waters are a promising alternative. To date, no

molecular assays have been developed for characterizing geosmin-producing cyanobacteria in the environment, however some toxin-producing species have been detected using these tools (Janse *et al.*, 2003), some of which may be useful in the identification of geosmin-producing cyanobacteria.

The polymerase chain reaction (PCR) is an effective procedure that generates large quantities of a specific DNA sequence in vitro through amplification in a three-step cycling process. This technique has been used for toxin studies of microcystin, produced by some planktonic cyanobacterial species. Microcystin is encoded in a three-gene cluster present only in those species that produce the toxin. By amplification of these genes via PCR, it is possible to take an environmental sample and characterize the microcystin-producing species. If the gene amplifies then it can be concluded that a microcystin-producing strain is present in the waters (Hisbergues *et al.*, 2003). This allows for early detection and provides a means for preventative measures against toxic species. This simple and cost-effective technique would be ideal for use in determining geosmin-producing cyanobacteria, however as noted previously it is not possible since geosmin does not translate into a protein directly from a gene. It is not possible to use a genetic marker from the non-mevalonate pathway to identify geosmin-producing species of cyanobacteria as the genes involved do not solely produce geosmin, but are also responsible for the production of other vital isoprenoids and terpenoids such as sterols, phytol, and carotenoids (Disch *et al.*, 1998). As a result, gene amplification would not be indicative of a geosmin-producing strain. For this reason, a different approach must be investigated to identify geosmin-producing cyanobacterial species from environmental samples.

1.6 DGGE IDENTIFICATION OF CYANOBACTERIA STRAINS IN WATERS

A novel approach initially developed for characterizing bacterial diversity in soil, which is now used in cyanobacterial studies, employs PCR in conjunction with denaturing gradient gel electrophoresis (DGGE). DGGE is a genetic fingerprinting technique that provides a profile of the community diversity sometimes to the species level on the basis of separation of unique PCR products. It is based on the decreased electrophoretic mobility of partially melted double stranded DNA molecules in a polyacrylamide gel containing a linear gradient of denaturants in the form of formaldehyde and urea (Muyzer, 1999). When a double stranded DNA fragment migrates by electrophoresis through a denaturing gradient, it partially melts and changes form. Nucleotide "domains" that suddenly dissociate at the same time, dramatically slow down migration, resulting in denaturation of the DNA fragment (Wang *et al.*, 2005). The oligonucleotide primers used in DGGE are composed of a long, thermally stable GC-rich stretch of nucleotides at the 5' end, so that rather than dissociating into single strands, the DNA melts in a step-wise process. Slight changes in base pair composition, often as little as 1 base pair, will shift these domain boundaries, thereby altering the conditions needed for domain dissociation (Wang *et al.*, 2005). Hence, fragments of the same size but differing in one base pair will migrate to different positions in the gel, revealing specific band migration for individual strains (Figure 4) (Kolmonen *et al.*, 2004). This technique provides an immediate overview of the cyanobacterial community in the sample and removes the need for costly DNA sequencing.

In order to identify taxa of cyanobacteria in environmental samples using DGGE, a comparison between characterized isolated cultures is often required to produce a standard marker. Banding in the environmental sample that coincides with those of known species in

Figure 4. Flow diagram showing the different steps in the analysis of microbial community structure using DGGE. DNA is isolated from an environmental sample (not necessarily collected on a filter) and used as template in PCR. PCR products are then separated by DGGE resulting in a profile of the community. A standard marker (labelled S) can be used to identify strains of interest.

Cyanobacteria Bloom

Water sample collection
on filter paper

DNA Isolation

PCR amplification with GC clamp primers

Urea-Formamide Denaturant
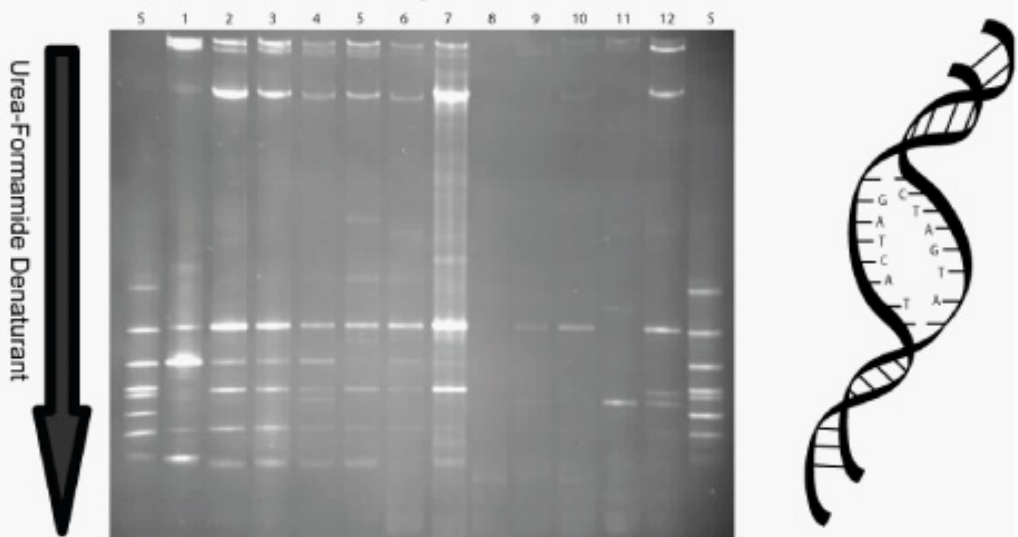
As DNA from all amplified organisms is electrophoresed through the denaturing gradient gel,
migration stops at a specific denaturant concentration/position on the gel.
The resulting pattern is the DGGE profile of the community.

the reference samples can allow for identifications. A genetic marker that can be specifically amplified from the reference organism is used in the PCR amplification and subsequent DGGE analysis. Ideally this marker must have sufficient heterogeneity for proper resolution as well as having a large sequence database that can allow for effective development of PCR primers (Janse *et al.*, 2003). DGGE has been used in a number of bacterial studies involving green sulfur bacteria, Desulfovibrionaceae, Proteobacteria (Overmann *et al.*, 1999; Zeng *et al.*, 2004), lactic acid bacteria (Temmerman *et al.*, 2004), Flavobacterium-Bacteroides-Cytophaga, and Planctomycetes (Kan and Chen, 2004) to examine community structure in various environments and under different conditions. It has also been used to characterize dinoflagellate species implicated in harmful algal blooms of *Pfisteria piscicida, Karlodinium micrum,* and *Heterocapsa rotundata* (Wang *et al.*, 2005). In addition, it has been employed to study changes in bacterial and cyanobacterial community structure after cyanophage-induced lysis of filamentous cyanobacteria (deBruin *et al.*, 2003) and the assessment of cyanobacterial community structure in Lake Loosdrecht, Netherlands composed of *Aphanizomenon, Planktothrix, Microcystis, Synechococcus, Prochlorothrix hollandica, Oscillatoria* (Zwart *et al.*, 2005). This technique therefore has proven to be useful, efficient, cost-effective, and relatively simple in studies of species composition in environmental samples and may be a useful method of characterizing geosmin-producing cyanobacteria for prevention and monitoring of taste and odour issues in drinking water quality.

1.7   CYANOBACTERIA SYSTEMATICS

In order to accurately identify species of cyanobacteria it is necessary to understand the current systematics of this group of organisms. For some time, there has been notable

controversy over cyanobacterial taxonomy and systematics. Molecular biologists and microbiologists prefer to regard this group taxonomically as bacteria (Oren, 2004) while ecologists and phycologists tend to treat them similar to taxonomic identifications of eukaryotic algae (Graham and Wilcox, 2000). Historically, cyanobacteria have been regarded as members of the plant world and taxonomy for this group has been based on the morphological species concept under the International Code of Botanical Nomenclature (ICBN) (Oren, 2004). Because the morphological species concept relies on structural characters to classify a species, there have been several limitations to this approach. Morphological plasticity is frequently exhibited by cyanobacteria and has led to difficulties in species delineation (Komárek and Komarkova, 2004; Palinska *et al.*, 1996). Some characteristics may be present in their environmental niches while a lab culture of the organism often displays significant differences in morphology. With the advent of molecular biology, phylogenetic analyses based on the rRNA genes and other relevant genetic markers have played a vital role in the taxonomy and nomenclature for all organisms (Komárek and Komarkova, 2004).

In 1978 Stanier et al. proposed that cyanobacterial taxonomy follow the International Code of Nomenclature of Bacteria (ICNB) and in 1979 Rippka et al. published a new taxonomy of the Cyanobacteria based on physiological, morphological and some genetic characteristics of those from pure cultures. This classification organized the Cyanobacteria into five sections that generally correspond to each order with respect to previous taxonomic systems (Litvaitis, 2002). This system of classification is widely employed by scientists however a major concern is that the type specimens used were derived from lab cultures and do not necessarily take into account phenotypic plasticity that

may have been exhibited by these groups when being observed in the field. It is estimated that 50% of cyanobacterial strains existing in culture collections have been identified incorrectly or have been assigned to the wrong taxonomic group thereby hindering the reliability of this system of classification (Litvaitis, 2002).

A modern approach to the classification of cyanobacteria was taken by Komárek and Anagnostidis (1986, 1988, 1989, 1990) who altered the classification based on the ICBN by including recent data such as morphology, life cycles, and biochemistry to improve the previous system of botanical taxonomy. They felt that discussions of an exclusively bacteriological or botanical approach to this group of organisms was insolvable and decided to classify, with names, the subsections of the Bacteriological Code with some variations. This revised system is comprised of four orders rather than five subsections as in the bacteriological system of classification. These authors do not view this system as definitive and encourage updates as more knowledge becomes available from a variety of morphological, biochemical, physiological, and molecular studies (Anagnostidis and Komárek, 1988, 1990; Komárek and Anagnostidis, 1986, 1989).

Relationships among the Cyanobacteria have not been completely resolved under the current classification and researchers are increasingly using different approaches to improve the taxonomy. The phylogenetic species concept is an approach that identifies a species using monophyly. Monophyletic groups contain all of the taxa that are descendant from a common ancestor. Under this concept, a species is identified by estimating the phylogeny of closely related populations and finding the smallest monophyletic group, a clade (Freeman and Herron, 2001). Cladistic methods use techniques that identify monophyletic groups on the basis of shared derived characters. Characters derived from

molecular data such as DNA and protein sequences are commonly used to determine the evolutionary distances of one organism to another. The degree of sequence similarity is directly correlated to the relationship between the organisms (Freeman and Herron, 2001). Rationale for utilizing molecular data in conjunction with the phylogenetic species concept include: i) DNA and protein sequences are strictly heritable entities whereas morphological and biochemical characters can be affected by environmental conditions, ii) the description of molecular characters is unambiguous while morphological characters are not always clear-cut, iii) molecular data are more amenable to quantitative treatments, and iv) some molecular data allow for very distantly related organisms to be assessed (i.e. between different kingdoms and phyla) whereas morphological characters are not comparable at these distances (Graur and Li, 2000). Due to the lack of a consensus phylogeny and an unreliable system of classification, unresolved evolutionary relationships within the Cyanobacteria continue to persist. A combination of morphological and molecular data of the studied organisms will strengthen these inferred relationships.

Cyanobacteria possess a ribosomal RNA (rRNA) cistron comprised of three genes; the 16S small subunit (SSU), 23S large subunit (LSU) and the 5S subunit, each separated by an internal transcribed spacer region (ITS) (Figure 5). A genetic marker often used in phylogenetic studies is the 16S rRNA gene. Within cyanobacteria, sequence information from this gene is widely regarded as one of the most valid criterion for determining relationships between closely related groups, such as species or genera. It is the basis for systematic assignment in the latest edition of Bergey's Manual of Systematic Bacteriology and has been useful in distinguishing broad taxonomic groups as well as individual species (Casamatta *et al.*, 2005; Rajaniemi *et al.*, 2005) (Litvaitis, 2002; Svenning *et al.*, 2005).

Figure 5. Map of the rRNA cistron. The ITS region can (but does not necessarily) include tRNA$_{ile}$ and/or tRNA$_{ala}$ genes. Sizes were approximated from sequences within GenBank. The tRNA genes are approximately 100 bp in length.

(~1400 bp)   (~354-648 bp)   (~2800 bp)

16S   23S

(tRNA ile tRNA ala)

ITS
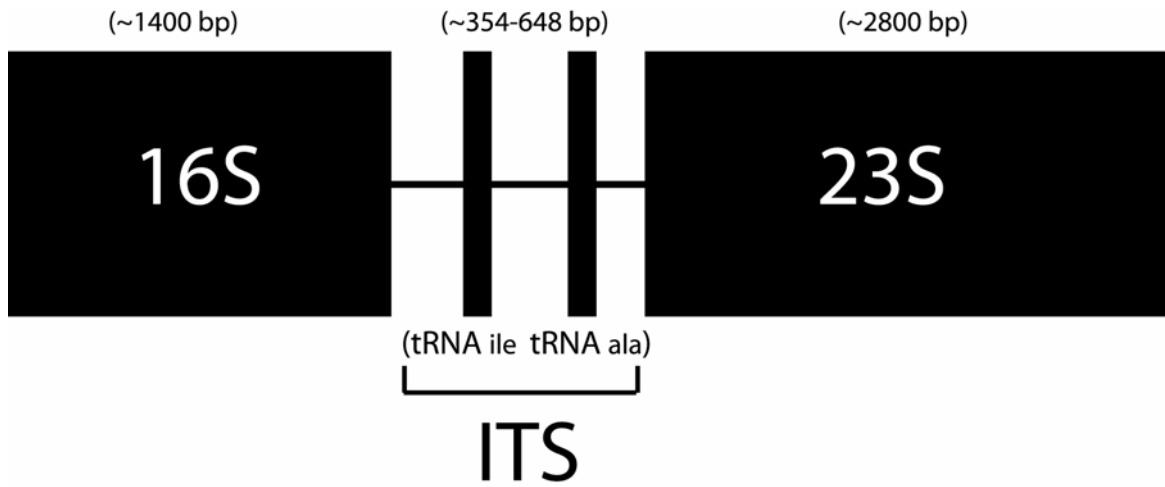
Since geosmin production is not species-specific and the genes involved in the pathway are not known, other genetic markers such as the 16S rRNA or ITS regions may be useful in discriminating among potential producers and non-producers. No phylogenetic studies have been performed on geosmin-producing strains of cyanobacteria and it is not known whether these strains have evolved from a single common ancestor or if they share more than one ancestor. Unresolved relationships among geosmin-producing cyanobacteria have not determined whether certain genetic markers will allow for discrimination among these strains. This study investigates the use of the 16S rRNA gene as a discriminative genetic marker among geosmin-producing and non-producing strains of cyanobacteria and evaluates the use of DGGE as a means for profiling communities that may lead to potential taste and odour episodes in Lake Ontario. Identification of these species using this technique may provide a means for preparation of taste and odour episodes in source waters and ultimately increase the acceptability of drinking water to consumers.

It is essential to monitor and research taste and odour events in order to develop a multi-barrier approach to drinking water safety. Furthermore, the early detection of odour producers in source waters allows managers to adopt a more proactive approach to control measures. Yet as noted above, there has been little progress towards the development of a systematic monitoring programme in most areas, in part because a reliable and cost effective method has not been developed. Taste and odour has major socioeconomic impacts and affects the trust that consumers have in the drinking water systems. Taste and odour episodes can be diagnostic of changes in biological activity or malfunction in source, treatment or distribution systems and can reflect fundamental changes in ecosystem health as a result of human related disturbance (Watson, 2004). With increased research into the

molecular, biochemical, physiological, and ecological factors that contribute to taste and odour in drinking water supplies, it is hopeful that a proactive approach to its control can be discovered.

1.8   OBJECTIVES

1. **Delineate various known geosmin-producing and non-geosmin-producing cyanobacterial isolates to the species level using the 16S rRNA gene and morphological characteristics**
2. **Develop a DGGE standard marker with reference isolates from the previous objective using a variable region of the 16S rRNA gene**
3. **Compare the DGGE standard to environmental samples collected from Lake Ontario in late August early September 2005 to identify potential geosmin producing species.**
4. **Determine the phylogenetic relationships among all isolates of geosmin-producing and non-producing cyanobacteria using 16S rRNA gene sequences to determine if geosmin-producing isolates may have evolved from a common ancestor and to compare current taxonomy with the developed phylogenies**

1.9 REFERENCES

Anagnostidis, K. and J. Komárek (1988). Modern approach to the classification system of Cyanophytes. 3. Oscillatoriales. Archiv fur Hydrobiologie Supplement 80: 327-472.

Anagnostidis, K. and J. Komárek (1990). Modern Approach to the Classification of the Cyanophytes 5 - Stigonematales. Algological Studies 59: 1-73.

Auld, H., D. MacIver and J. Klaassen (2004). Heavy rainfall and waterborne disease outbreaks: The Walkerton example. Journal of toxicology and environmental health, Part A 67(20-22): 1879-1887.

Cane, D. E., C. Chow, A. Lillio and I. Kang (2001). Molecular cloning, expression and characterization of the first three genes in the mevalonate-independant isoprenoid pathway in *Streptomyces coelicolor*. Bioorganic and Medicinal Chemistry 9: 1467-1477.

Casamatta, D. A., J. R. Johansen, M. L. Vis and S. T. Broadwater (2005). Molecular and morphological characterization of ten polar and near-polar strains within the Oscillatoriales (Cyanobacteria). Journal of Phycology 41: 421-438.

Chung, S., J. Vercelloti, P. Johnson and P. Klesius (1991). Development of an enzyme-linked immunosorbent assay for GEO. Journal of Agricultural and Food Chemistry 39: 764-769.

deBruin, A., B. W. Ibelings and E. V. Donk (2003). Molecular techniques in phytoplankton research: from allozyme electrophoresis to genomics. Hydrobiologia 491: 47-63.

Dionigi, C. P. (1995). The effects of copper sulfate on geosmin biosynthesis by *Streptomyces tendae, Streptomyces albidoflavus*, and *Penicillum expansum*. Water Science and Technology 31(11): 135-138.

Disch, A., J. Schwender, C. Muller, H. K. Lichtentaler and M. Rohmer (1998). Distribution of the mevalonate and glyceraldehydephosphate/pyruvate pathways for isoprenoid biosynthesis in unicellular algae and the cyanobacterium *Synechocystis* PCC 6714. Journal of Biochemistry 333: 381-388.

Downing, J. A., S. B. Watson and E. McCauley (2001). Predicting Cyanobacteria dominance in lakes. Canadian Journal of Fisheries and Aquatic Sciences 58: 1905-1908.

Ershov, Y. V., R. R. Gantt, J. Francis X. Cunningham and E. Gantt (2002). Isoprenoid Biosynthesis in *Synechocystis* sp. Strain PCC6803 Is Stimulated by Compounds of the Pentose Phosphate Cycle but Not by Pyruvate or Deoxyxylulose-5-Phosphate. Journal of Bacteriology 184(18): 5045-5051.

Freeman, S. and J. C. Herron (2001). Evolutionary Analysis. Upper Saddle River, Prentice Hall.

Graham, L. E. and L. W. Wilcox (2000). Algae. Upper Saddle River, Prentice-Hall.

Graur, D. and W. H. Li (2000). Fundamentals of Molecular Evolution. Sunderland, MA, Sinauer Associates, Inc.

Gust, B., G. L. Challis, K. Fowler, T. Kieser and K. Chater (2003). PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proceedings of the National Academy of Sciences 100(4): 1541-1546.

Hisbergues, M., G. Christiansen, L. Rouhiainen, K. Sivonen and T. Borner (2003). PCR-based identification of microcystin-producing genotypes of different cyanobacterial genera. Archives of Microbiology 180: 402-410.

Hoefel, D., L. Ho, W. Aunkofer, P. Monis, A. Keegan, G. Newcombe and C. Saint (2006). Cooperative biodegradation of geosmin by a consortium comprising three gram-negative bacteria isolated from the biofilm of a sand filter column. Letters of Applied Microbiology 43(3): 417-423.

Janse, I., M. Meima, W. E. A. Kardinaal and G. Zwart (2003). High-Resolution Differentiation of Cyanobacteria by Using rRNA-Internal Transcribed Spacer Denaturing Gradient Gel Electrophoresis. Applied and Environmental Microbiology 69(11): 6634-6643.

Jüttner, F. (1995). Physiology and biochemistry of odorous compounds from freshwater Cyanobacteria and algae. Water Science and Technology 31(11): 69-78.

Kan, J. and F. Chen (2004). Co-monitoning bacterial and dinoflagellates communities by denaturing gradient gel electrophoresis (DGGE) and SSU rRNA sequencing during a dinoflagellates bloom. Acta Oceanologica Sinica 23(3): 483-492.

Kolmonen, E., K. Sivonen, J. Rapala and K. Haukka (2004). Diversity of Cyanobacteria and heterotrophic bacteria in cyanobacterial blooms in Lake Joutikas, Finland. Aquatic Microbial Ecology 36: 201-211.

Komárek, J. and K. Anagnostidis (1986). Modern Approach to the Classification System of Cyanophytes 2. Chroococcales. Archiv fur Hydrobiologie Supplement 73: 157-226.

Komárek, J. and K. Anagnostidis (1989). Modern Approach to the Classification System of Cyanophytes 4. Nostocales. Archiv fur Hydrobiologie Supplement 82: 247-345.

Komárek, J. and J. Komarkova (2004). Taxonomic review of the cyanoprokaryotic genera *Planktothrix* and *Planktothricoides*. Czech Phycology 4: 1-18.

Lange, B. M., T. Rujan, W. Martin and R. Croteau (2000). Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. Proceedings of the National Academy of Sciences 97(24): 13172-13177.

Litvaitis, M. K. (2002). A molecular test of cyanobacterial phylogeny: inferences from constraint analyses. Hydrobiologa 468: 135-145.

Mason, C. F. (1996). Biology of freshwater pollution. Essex, Longman Group Limited.

Miller, D., E. Conte, C. Shen and P. Perschbacher (1999). Colorimetric approach to cyanobacterial off-flavor detection. Water Science and Technology 40(6): 165-169.

Murphy, T. P., K. Irvine, J. Guo, J. Davies, H. Murkin, M. Charlton and S. B. Watson (2003). New microcystin concerns in the lower Great Lakes. Water Quality Research Journal of Canada 38(1): 127-140.

Muyzer, G. (1999). DGGE/TGGE a method for identifying genes from natural ecosystems. Current Opinion in Microbiology 2: 317-322.

Nakamura, S., N. Nakamura and S. Ito (2001). Determination of 2-methylisoborneol and geosmin in water by gas chromatography-mass spectrometry using stir bar sorptive extraction. Journal of Separation Science 28(8): 674-677.

Nakashima, S. and M. Yagi (1992). Iron forms that influence the growth and musty odor production of selected Cyanobacteria. Water science and technology 25: 207-216.

Oren, A. (2004). A proposal for further integration of the Cyanobacteria under the Bacteriological Code. International Journal of Systematic and Evolutionary Microbiology 54: 1895-1902.

Overmann, J., M. J. L. Coolen and C. Tuschak (1999). Specific detection of different phylogenetic groups of chemocline bacteria based on PCR and denaturing gradient gel electrophoresis of 16S rRNA gene fragments. Archives of Microbiology 172: 83-94.

Palinska, K. A., W. Liesack, E. Rhiel and W. E. Krumbien (1996). Phenotype variability of identical genotypes: the need for a combined approach in cyanobacterial taxonomy demonstrated on *Merismopedia*-like isolates. Archives of Microbiology 166: 224-233.

Rajaniemi, P., P. Hrouzek, K. Kastovska, R. Williame, A. Rantala, L. Hoffman, J. Komarek and K. Sivonen (2005). Phylogenetic and morphological evaluation of the genera *Anabaena*, *Aphanizomenon*, *Trichormus* and *Nostoc* (Nostocales,

Cyanobacteria). International Journal of Systematic and Evolutionary Microbiology 55: 11-26.

Saadoun, I., K. K. Schrader and W. T. Blevins (2001). Environmental and nutritional factors affecting geosmin synthesis by *Anabaena* sp. Water Research 35(5): 1209-1218.

Schrader, K. K., A. M. Rimano, C. S. Tucker, J. Glinski, S. J. Cutler and H. G. Cutler (2004). Evaluation of the natural product SeaKleen for controlling the musty-odor-producing cyanobacterium *Oscillatoria perornata* in catfish ponds. North American Journal of Aquaculture 66: 20-28.

Skafel, M. and R. Yerubandi (2003). Physical Processes in Western Lake Ontario during a taste and odour episode. Canadian Coastal Conference 2003, NWRI, Environment Canada.

Svenning, M. M., T. Eriksson and U. Rasmussen (2005). Phylogeny of symbiotic cyanobacteria within the genus *Nostoc* based on 16S rDNA sequence analyses. Archives of Microbiology 183: 19-26.

Temmerman, R., G. Huys and J. Swings (2004). Identification of lactic acid bacteria: culture-dependent and culture-independent methods. Trends in Food Science & Technology 15: 348-359.

Wang, Q., J. R. Deeds, A. R. Place and R. Belas (2005). Dinoflagellate community analysis of a fish kill using denaturing gradient gel electrophoresis. Harmful Algae 4: 151-162.

Watson, S. B. (2004). Aquatic taste and odour: a primary signal of drinking water integrity. Journal of Toxicological and Environmental Health, Part A 67: 1779-1795.

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

Zeng, Y., N. Jiao, H. Cai, XHChen and C. Wei (2004). Phylogenetic diversity of ribulose-1, 5-bisphosphate carboxylase/oxygenase large subunit genes of bacterioplankton in the East China Sea. Acta Oceanologica Sinica 23(4): 673-685.

Zwart, G., M. P. K.-v. Agterveld, I. v. Werff-Staverman, F. Hagen, H. L. Hoogveld and H. J. Gons (2005). Molecular characterization of cyanobacterial diversity in a shallow eutrophic lake. Environmental Microbiology 7(3): 365-377.

CHAPTER 2

DEVELOPMENT OF A DGGE STANDARD MARKER FOR USE IN IDENTIFICATION OF POTENTIAL GEOSMIN-PRODUCING SPECIES OF CYANOBACTERIA

2.0 INTRODUCTION

2.1.1 GEOSMIN

Geosmin (*trans*-1, 10-dimethyl-*trans*-9-decalol) is a terpenoid metabolite produced by a number of bloom-forming cyanobacteria and bacteria. This volatile organic compound (VOC) has an earthy, musty, muddy scent and is detected by humans at extremely low levels (~4-10 ng $\cdot$ l^{-1}) (Nakamura *et al.*, 2001). Geosmin is the most frequently reported undesirable taste and odour compound in Lake Ontario drinking water. It is not only present subsequent to a bloom but can also be generated in the absence of conspicuous blooms by highly prolific species present at low-moderate abundances, or by benthic or subsurface populations. The biosynthesis of geosmin is not fully elucidated and has resulted in difficulties when anticipating, tracing and controlling taste and odour episodes in drinking water.

Geosmin is resistant to oxidation and oxidants such as Cl_2, ClO_2, chloramines and $KMNO_4$ that are commonly applied in water purification have been largely ineffective at removing the compound due to a tertiary alcohol structure (Hu *et al.*, 2003). Granulated activated carbon (GAC) is typically used in the water purification process and will absorb geosmin from source water. However, at high levels geosmin often evades standard GAC treatment and reaches consumer taps thereby eliciting concern for the quality and safety of the water (Watson, 2004). Although geosmin is not a health hazard (Blaha *et al.*, 2004), the

lack of public education in the awareness of taste and odour has resulted in numerous complaints and loss of consumer confidence in the system.

A proactive approach to controlling taste and odour episodes in drinking water by monitoring cyanobacterial species composition in water samples is an option that has not been employed by the majority of water managers. Characterization of species diversity in a water body could provide a profile of potential taste and odour-causing species present in samples and may help predict when these events occur. Water quality managers could in turn prepare for taste and odour events thereby optimizing water quality. In the long term, identification of potentially problematic cyanobacteria taxa that occur in source water could contribute towards a more robust management of drinking water through corrective or preventative measures (Watson, 2004).

2.1.2 *IDENTIFICATION OF GEOSMIN-PRODUCING CYANOBACTERIA*

The majority of geosmin-producing cyanobacteria are planktonic, filamentous N_2 fixers or non-fixers. Some benthic forms are also known to produce geosmin and there are many uncharacterized cyanobacterial species that may have the ability to produce this compound (Watson, 2004). Studies suggest that geosmin production is not a species-specific trait and the only way to confirm the producing strain is by isolation and morphological identification after each taste and odour event (Rashash *et al.*, 1995; Watson *et al.*, 2000). Due to the difficulty in isolation and maintenance of pure strains of cyanobacteria, there has been immense difficulty in their characterization from environmental samples. Selective enrichment cultures often fail to mimic the growth conditions organisms require for proliferation in their natural environment (Muyzer *et al.*, 1993). It is estimated that far less than 5% of all cyanobacterial species have been established in cultures, partly because

certain species or strains are impossible to grow on specified media (Castenholz, 1992). In addition, successful isolation and maintenance of strains under laboratory conditions for prolonged periods can result in an alteration of cell characteristics. Morphological characters initially observed in the field and used to classify each strain have been reported to change over time in culture, sometimes resulting in strain re-assignment (Palinska *et al.*, 1996). This phenotypic plasticity observed in cultured strains of cyanobacteria has hindered the development of "type" specimens used as controls for morphological identification of species (Oren, 2004) and as a result, molecular biology techniques have been increasingly used to overcome these difficulties.

2.1.3 *MOLECULAR CHARACTERIZATION OF CYANOBACTERIA*

The 16S rRNA gene is the basis for systematic assignment in the latest edition of Bergey's Manual of Systematic Bacteriology and has been useful in distinguishing broad taxonomic groups as well as individual species and is frequently used in the identification of cyanobacteria (Casamatta *et al.*, 2005; Litvaitis, 2002; Rajaniemi *et al.*, 2005; Svenning *et al.*, 2005). Direct sequence analysis of this region has been fundamental in phylogenetic studies of cyanobacteria and has resulted in a large expansion of the molecular data that is available in public databases such as GenBank (Svenning *et al.*, 2005). Although the analysis of 16S rRNA gene sequences remains the most resolved method for distinguishing taxonomic groups, it nevertheless is a lengthy process. To characterize species of cyanobacteria in environmental samples using sequence analysis of the full 16S rRNA gene, time-consuming procedures such as cloning or isolation and culturing of each strain would be necessary. Although a substantial amount of data can be gained through sequence analysis of genes such as the 16S rRNA, this technique is neither rapid nor simple and

would not be ideal for use in the identification of geosmin-producing cyanobacteria to facilitate the prediction of taste and odour episodes.

Denaturing gradient gel electrophoresis (DGGE) is a molecular technique that has been used to assess the genotypic diversity in environmental samples as well as to evaluate the purity of cultures without the need for strain isolation or lengthy DNA sequence analysis. This technique works through sequence-dependant separation of PCR products and is based on the decreased electrophoretic mobility of partially melted double stranded DNA molecules in a polyacrylamide gel containing a linear gradient of denaturants in the form of formaldehyde and urea (Muyzer, 1999). When a double stranded DNA fragment migrates by electrophoresis through a denaturing gradient, it partially melts and changes form. Nucleotide "domains" that suddenly dissociate at the same time, dramatically slow down migration, resulting in denaturation of the DNA fragment (Chapter 1, Figure 4) (Wang *et al.*, 2005). Slight changes in base pair (bp) composition, often as little as 1 bp, will shift domain boundaries, thereby altering the conditions needed for domain dissociation. Accordingly, the denaturing concentrations required for domain dissociation for all of the DNA fragments in an environmental sample will vary. Each species within an environmental sample will correlate to a DNA fragment and migration of fragments to different positions in the gel, will result in a DGGE profile (Kolmonen *et al.*, 2004). DGGE profiles from environmental samples can give insight into the diversity of certain groups of organisms and has been used in a number of cyanobacterial studies to assess community structure (Zwart *et al.*, 2005), monitor diversity over time (Kolmonen *et al.*, 2004), and identify characteristic strains of cyanobacteria such as toxin-producing *Microcystis* (Janse *et al.*, 2003). This technique has proven to be efficient, cost-effective, and relatively simple

in studies of species composition in environmental samples and is applied in this study for the first time in the identification of potential geosmin-producing cyanobacteria in Lake Ontario.

2.1.4 DGGE STANDARD MARKER

In many DGGE studies a standard marker is usually developed that is composed of DNA fragments of known origin and can be used for identification of strains within communities. Standard markers enable comparisons between DGGE gels and can be used as to assess the quality of a run (Janse *et al.*, 2004). They can be produced from cultured and characterized isolates in which a precise identity (often to a species level) is known (Janse et al. 2004) or from excised and sequenced DGGE bands from environmental samples (Kan *et al.*, 2006). Maximum fragment sizes used in DGGE are approximately 1000 nucleotides. With an increased number of nucleotides and melting domains however, the mobility shifts decrease, thus the fraction of mutations detected decreases. A greater resolution is obtained with fewer melting domains, however it is important to use an adequate number of nucleotides to observe differences and separation (Boutte *et al.*, 2006). A completed standard marker comprises a mixture of isolate fragments or cut bands and is electrophoresed alongside other samples resulting in a profile in which each band is of known source and can be used to make comparisons (Zwart *et al.*, 2005).

Development of a DGGE marker from characterized isolates ideally requires a pure culture yielding a single amplicon representative of that strain. The purity of an isolated or clonal organism can be judged using DGGE when a single band is observed on the gel, indicative of a single amplicon (Janse *et al.*, 2003). Occasionally, more than one amplicon from a pure culture may be present in a PCR reaction if the gene is present in more than

one variation for the same species or due to heteroduplex formations (Crosbie *et al.*, 2003). Heteroduplex formations are chimeric DNA molecules that form when an incompletely extended PCR product acts as a primer on a heterologous sequence. These chimeras are separated from the homoduplex molecule under denaturing conditions and result in an overestimation of sequence variants (Thompson *et al.*, 2002). Samples with more than one band due to multiple homoduplexes or heteroduplexes are not ideal for use in a DGGE marker simply because more than one band is representative of a single isolate.

When applying DGGE, it is important to use a genetic marker that provides enough sequence variability that it can distinguish among the taxonomic groups of interest. In addition, it should possess conserved regions that will allow for primers to anneal and selectively amplify groups of interest. The V3 hypervariable region spanning the 357-518 nucleotide region of the 16S rRNA gene has been identified as the most effective in providing superior band separation and the highest species richness among bacterial community samples (Yu and Morrison, 2004). Cyanobacteria specific V3 primers have been developed by Zwart et al., 2005 to selectively amplify DNA from cyanobacteria. Use of these primers in PCR amplification has allowed for effective analysis of cyanobacterial communities using DGGE without interference from contaminating bacterial DNA (Zwart *et al.*, 2005).

In order to effectively apply DGGE to the identification of potential geosmin-producing cyanobacteria in Lake Ontario samples, it is first necessary to develop a marker using DNA from characterized geosmin-producing species. In this study, development of a marker using isolates from both geosmin-producing and non-producing cyanobacteria originating from Lake Ontario and other locations in Europe and the United States was a

primary objective. This chapter evaluates the use of the 16S rRNA-V3 region for distinguishing among geosmin-producing and non-producing strains of cyanobacteria and discusses its potential for application in characterizing geosmin-producing strains of cyanobacteria in Lake Ontario.

2.2 MATERIALS AND METHODS

2.2.1 CULTURE CONDITIONS OF CYANOBACTERIA ISOLATES

Twenty-three geosmin-producing and non-producing cyanobacterial isolates (Table 1) were obtained from Dr. Susan Watson (Environment Canada, University of Calgary), Dr Freiderich Jüttner (University of Zurich), and Mr. George Izaguirre (Metropolitan District of Southern California). Isolates from various global locations as well as from three culture collections; the Pasteur Culture Collection (PCC), the University of Toronto Culture Collection (UTCC), and the University of Texas Culture Collection (UTEX) were amongst the studied cyanobacteria. The resulting collection used in this study was comprised of 20 planktonic strains of which four were geosmin-producing and three benthic strains of which two were geosmin-producing. Isolates of particular focus originated from Lake Ontario and included five *Anabaena lemmermannii* (two geosmin-producing), one *Anabaena flos-aquae* (non geosmin-producing), and one *Aphanizomenon flos-aquae* (non geosmin-producing). These cultures were grown at 23ºC in Chu-10 (Stein, 1973) or cyano media (Jüttner *et al.*, 1983) in a sterile 30 ml borosilicate glass test tube at an irradiance of approximately 25 $\mu mol \cdot m^{-2} \cdot s^{-1}$ for a cycle of light-to-dark ratio of 16:8 h. To increase the volume of each culture and allow for better growth conditions, a small amount of cells (approximately 1 ml) were transferred to a larger sterile 150 ml Erlenmeyer flask containing approximately 60 ml of media during the period between mid-exponential growth phase and stationary phase of the culture. To maintain constant cell growth, subcultures of each isolate were performed every one to two months depending on the growth rate of each culture. During this time, approximately 1-5 ml of each culture was transferred to a new sterile 150 ml or 250 ml Erlenmeyer flask and grown under the same

conditions as the initial culture. The isolates were monitored microscopically to eliminate cross-contamination and dominance of bacteria. All steps were performed in a laminar flow hood using aseptic technique and sterile equipment.

2.2.2 ISOLATE MORPHOLOGY

For each of the 23 isolates, averages of triplicate length and width measurements were taken on vegetative cells, heterocysts (when observed), and akinetes (when observed). The cells were viewed under the Olympus BX51 System compound light microscope and images were photographed at 400x, 600x, and 1000x magnification. Cell measurements were calculated using the ImagePro© Express ver. 4.0 imaging program for Windows (Cybernetics, 1999). Geosmin production by each isolate was determined in previous studies by Dr. Susan Watson (Environment Canada, University of Calgary), Dr Frederich Juttner (University of Zurich), or Mr. George Izaguirre (Metropolitan District of Southern California) using gas-chromatography-mass-spectroscopy (GC-MS), other chromatographic methods, and/or sensory analysis by human perception by the culturing scientists.

2.2.3 DNA ISOLATION AND PCR AMPLIFICATION OF THE 16S RRNA GENE

Approximately 250 μl of healthy cells (assessed by buoyancy, colour, and abundance in culture) were transferred from each culture into a sterile 1.5 ml microfuge tube and centrifuged at 8000 rpm for 5 min to pellet the cells. For cell lysis, the supernatant was removed and the pelleted cells were freeze-thawed by immersing each tube in liquid nitrogen for 10 s immediately followed by immersion in a 65˚C water bath for 15 s. DNA

was isolated using the Qiagen DNeasy Plant Mini Kit (Qiagen, Canada) and DNA was eluted in 50-150 µl of either AE buffer or sterile water.

The 16S rRNA gene (approximately 1400 bp) was PCR amplified using the Eppendorf Mastercycler® Gradient 5331 (Eppendorf, USA) and one program with several modifications and four primers 27F1 (Svenning *et al.*, 2005), 781Fa, 781Fb, 781Ra, and 781Rb (Nubel *et al.*, 1997), and 2000R (Boyer *et al.*, 2001) that were used depending on the efficiency of amplification. A schematic representation of the primers positions relative to the rRNA cistron is shown in Figure 6. A volume of 1-3 µl of DNA, 2.5 µl 10x taq polymerase buffer (500 mM KCl, 100 mM Tris-HCl, pH 9.0 (at 25°C), 1.0% Triton® X-100, and 15 mM MgCl$_2$) (Promega, Canada), 0.625 U Taq DNA polymerase (Promega, Canada), 200 µM each dNTP (Promega, Canada), 0.5 mM of each primer of a set in a total reaction volume of 25 µl. Primer sets were composed of one forward and one reverse primer and were chosen based on the region requiring sequencing and amplification efficiency. The Cyano2 program (modified from (Casamatta *et al.*, 2003)) used in the amplification consisted of an initial denturation step at 95˚C for 1 min 30 s, followed by 30 cycles performed at 93˚C for 1 min, 55˚C for 1min, and 72˚C for 1 min 10 s. A final extension step at 72°C for 10 min completed the amplification program. Modifications to this program were attempted if reaction products were not satisfactory in the following manner:

No or little amplification   - decreased annealing temperature to within the range of 50°C-53°C

- increased number of cycles to 32

Table 1. Cyanobacteria isolates of the Chroococcales, Nostocales, and Oscillatoriales used in the sequence analysis of the 16S rRNA gene and in the development of the16S rRNA-V3 region DGGE standard marker. The asterisk denotes a geosmin-producing strain in culture.

Taxon		Strain	Origin	Source Collection
Nostocales	*Anabaena flos-aquae* Breb ex Born. et Flah.	UTEX LB 2383	Burton Lake, Saskatchewan, Ontario, Canada	University of Texas Culture Collection c/o Dr. Susan Watson (Environment Canada, University of Calgary)
	An. flos-aquae	UTCC64	Western Lake Ontario, Canada	University of Toronto Culture Collection c/o Dr. Susan Watson (Environment Canada, University of Calgary)
	Anabaena lemmermannii Richter	AL4	Western Lake Ontario, Canada	Dr. Susan Watson (Environment Canada, University of Calgary)
	An. lemmermannii	AL5	Western Lake Ontario, Canada	Dr. Susan Watson (Environment Canada, University of Calgary)
	An. lemmermannii	AL7	Western Lake Ontario, Canada	Dr. Susan Watson (Environment Canada, University of Calgary)
	*An. lemmermannii**	LONT2	Western Lake Ontario, Canada	Dr. Susan Watson (Environment Canada, University of Calgary)
	*An. lemmernannii**	LONT5	Western Lake Ontario, Canada	Dr. Susan Watson (Environment Canada, University of Calgary)
	*An.lemmermannii**	GIOL8	California, USA	Mr. George Izaguirre (Metropolitain District of Southern California)
	Anabaena planktonica Brunneth		Unknown	Dr Freiderich Juttner (University of Zurich)
	Anabaena sp.	LOW	Lake of the Woods, Manitoba	Dr. Susan Watson (Environment Canada, University of Calgary)
	Anabaena spiroides Klebhan		Unknown	Dr Freiderich Juttner (University of Zurich)
	Anabaena viguieri Denis et Frémy		Unknown	Dr Freiderich Juttner (University of Zurich)
	Anabaena cylindrica Lemmermann		Liverpool, England	Prof.N.G.Carr c/o Dr Freiderich Juttner (University of Zurich)

Taxon		Strain	Origin	Source Collection
	Trichormus variabilis Ralfs ex Bornet et Flahault		Unknown	Dr Freiderich Juttner (University of Zurich)
	Aphanizomenon gracile* Lemmermann		Unknown	Dr Freiderich Juttner (University of Zurich)
	Aphanizomenon flos-aquae (Linneaus) Ralfs	HHAFA	Hamilton, Harbour, Lake Ontario, Canada	Dr. Susan Watson (Environment Canada, University of Calgary)
	Calothrix parietina (Thuret) Bornet et Flahault	PCC 6303	Wisconsin, USA	Pasteur Culture Collection c/o Dr Freiderich Juttner (University of Zurich)
	Calothrix sp.*	PCC 7507	Vierwaldstättersee, Switzerland	Pasteur Culture Collection c/o Dr Freiderich Juttner (University of Zurich)
Chroococcales	*Microcystis aeruginosa* Kutzing		Unknown	Dr. Susan Watson (Environment Canada, University of Calgary)
	Microcystis wesenbergii Komárek		Hamilton, Harbour, Lake Ontario, Canada	Dr. Susan Watson (Environment Canada, University of Calgary)
Oscillatoriales	*Planktothrix aghardii* (Gomont) Anagnostidis & Komárek		Zurich, Switzerland	Dr Freiderich Juttner (University of Zurich)
	Oscillatoria limosa*		Unknown	Dr Freiderich Juttner (University of Zurich)
	Pseudanabaena sp. UTCC 593	UTCC 593	Pulp and paper effluent, Windish, PQ, Canada	University of Toronto Culture Collection c/o Dr. Susan Watson (Environment Canada, University of Calgary)

Figure 6. PCR primers used in the amplification of the 16S rRNA gene (27F1, 781F (a&b), 781R (a&b), and 2000R) and the 16S rRNA-V3 region for use in DGGE (cya-b-F371, F357GC, cyaR783, and R518). One reverse and one forward primer was used in a set, arrows indicate direction of extension.
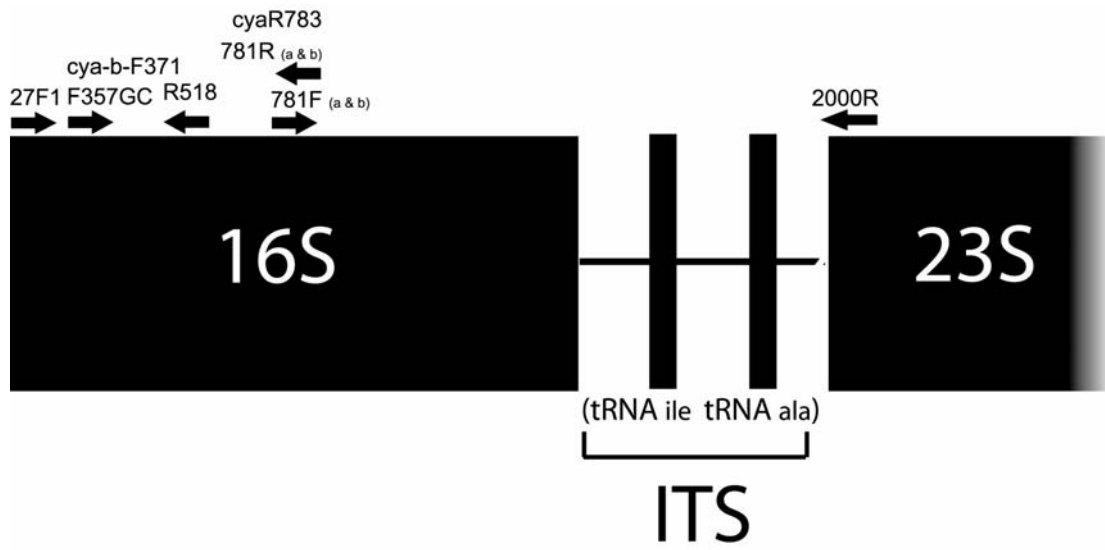
> *Non-specific binding*   - increased annealing temperature to within the range
>
> of 57°C-63°C

PCR products were electrophoresed through a 1.5% agarose gel in 1X TBE buffer at 130 V for 45 min. Products were visualized using the Syngene Bioimaging System (Synoptics Ltd., United Kingdom).

2.2.4 16S rRNA GENE SEQUENCING AND ANALYSIS

PCR amplification products were purified using the QIAquick PCR Purification kit© (Qiagen, Canada) and eluted in 30-50 µl of biotech grade water (Fisher, Canada). DNA concentrations were estimated with comparison to a *Hae* III DNA standard and approximately 10 ng of the purified product was sequenced at the University of Waterloo molecular core facility using the Applied Biosystems 3130XL Genetic Analyzer. Because sequencing of this region was difficult likely due to development of secondary structures, first attempts in sequencing used the forward primer of the original PCR reaction while subsequent sequencing applied different primers based on sequence quality from previous attempts (Appendix A). In order to obtain the nearly completed 16S rRNA gene sequence, numerous sections of the full-length gene were combined to produce a consensus sequence. Sequence reaction products were analyzed using Bioedit sequence alignment editor and analysis program (Hall, 1999). The resulting nucleotide sequences from each isolate were aligned with sequences deposited in the GenBank database using the program BLAST (Altschul *et al.*, 1990) (www.ncbi.nlm.nih.gov/blast/).

2.2.5 *PCR* AMPLIFICATION OF *16S rRNA-V3* REGION

Using DNA extracts of the isolates, a 161 base pair region of the 16S rRNA gene was amplified using a nested PCR protocol with cyanobacterial specific primers (Zwart *et al.*, 2005) to eliminate possible bacterial DNA interference from the non-axenic isolates. A schematic representation of the primer positions relative to the full 16S rRNA gene is shown in Figure 6. The first cyanobacteria-selective round of amplification was performed with 3μl DNA, 2.5 μl 10x taq polymerase buffer (500 mM KCl, 100 mM Tris-HCl, pH 9.0 (at 25°C), 1.0% Triton® X-100, and 15 mM MgCl$_2$) (Promega, Canada), 0.625 U Taq DNA polymerase (Promega, Canada), 200 μM each dNTP (Promega, Canada), 0.5 mM of each cyanobacterial specific primer Cya-b-F371 (5'-CCTACGGGAGGCAGCAGTGGGGAATTTTCCG-3') and Cya-R783 (5'-GACTACWGGGGTATCTAATCCW-3') (Zwart *et al.*, 2005) in a total reaction volume of 25 μl. A 20-cycle touchdown procedure was followed using the Eppendorf Mastercycler® Gradient 5331 (Eppendorf, USA). After an initial denaturation step at 95°C for 1 min 30 s, a 20 cycle touchdown procedure was performed at 94°C for 1 min, 65°C for 1min, and 72°C for 1 min, in which the annealing temperature decreased by 0.5°C each cycle to end at 55°C in the final cycle. A final extension step at 72°C for 10 min concluded the end of the preamplification.

The second round of amplification was performed with similar reagent concentrations as the first round but with the following changes: 3 μl template DNA, 0.5mM of each general bacterial primer F357GC (5'-CCTACGGGAGGCAGCAG-3') and R518 (5'-CCAGCAGCCGCGGTAAT-3') (Zwart *et al.*, 2005) in a total reaction volume of 25 μl. A 40 bp GC-clamp 5'-CGC CCG CCG CGC CCC GCG CCC GGC CCG CCG

49

CCC CCG CCC-3' was added to the 5' end of primer F357GC to increase the separation of DNA bands in DGGE gel (Muyzer *et al.*, 1993). Following the initial denaturation step at 95ºC for 1 min 30 s, a touchdown procedure consisting of 20 cycles at 94˚C for 1 min, 65˚C for 1min, and 72˚C for 1 min, in which the annealing temperature decreased by 0.5˚C each cycle was performed. Subsequently five additional cycles were performed at 94˚C for 1 min, 55˚C for 1min, and 72˚C for 1 min. A final extension step at 72˚C for 10 min concluded the nested round of the procedure. Each PCR product was electrophoresed through a 1.5% agarose gel in 1X TBE buffer at 130 V for 45 min. Products were visualized using the Syngene Bioimaging System (Synoptics Ltd., United Kingdom). DNA concentrations were estimated with comparison to a *Hae* III DNA standard.

2.2.6 DGGE

Optimal acrylamide percentage and denaturant concentrations for parallel DGGE were determined with electrophoresis of 50 μl of two PCR products from *Anabaena lemmermannii* AL5 and *Anabaena planktonica* in a parallel denaturing gradient gel. The 16 mm x 16 mm gel was composed of 8% (wt/vol) polyacrylamide (at an acrylamide-to-bisacrylamide ratio of 37.5:1) and a linear gradient of the denaturants urea and formamide, increasing from 20% on the left to 70% on the right. The denaturing solutions were made from 8% polyacrylamide stocks of 100% denaturant and 0% denaturant with the 100% denaturant defined as 7 M urea and 40% formamide as described by Zwart and Bok, 2004. The gel was polymerized with the addition of 4.5% ammonium persulfate (APS) and 3.4% TEMED.

Electrophoresis was performed using the D-Code Universal Mutation Detection System (BioRad, Canada) in a buffer containing 0.04 M Tris-acetate and 0.001 M EDTA

(pH 7.6) (0.5X TAE buffer) for 16 h at 60 V and a constant temperature of 60˚C. The gel was stained in 0.5 μg ethidium bromide ml^{-1} for 1 h and subsequently visualized using the Syngene Bioimaging System (Synoptics Ltd., United Kingdom). The optimal denaturing range for perpendicular DGGE was determined based on the migration trend of the PCR products.

After optimization experiments, perpendicular DGGE was performed using the D-Code Universal Mutation Detection System (BioRad, Canada) 25 μl of PCR product from the V3 region from each isolate was electrophoresed in an 8% polyacrylamide gel containing a gradient of denaturant ranging from 35-55% (100% denaturant is 7 M urea and 40% deionized formamide). Electrophoresis was performed using the D-Code Universal Mutation Detection System (BioRad, Canada) in a buffer containing 0.04 M Tris-acetate and 0.001 M EDTA (pH 7.6) (0.5X TAE buffer) for 16 h at 60 V and a constant temperature of 60˚C. The gel was stained in 0.5 μg ethidium bromide ml^{-1} for 1 h and subsequently visualized using the Syngene Bioimaging System (Synoptics Ltd., United Kingdom).

2.2.7 *SEQUENCING OF DNA FROM DGGE PCR PRODUCTS AND BANDS*

Single band PCR products observed on the DGGE gel were purified using the QIAquick PCR Purification kit© (Qiagen, Canada). For samples exhibiting more than one band in DGGE, the most intense band was cut out using a sterile blade and incubated in 50 μl sterile for 24 h at 4˚C. 0.5 μl of the eluant was re-amplified using the second round of amplification protocol and PCR products were purified as in previously noted procedures. All purified PCR products were sequenced using the Applied Biosystems 3130XL Genetic Analyzer and either the F357GC primer or F357 primer (Zwart *et al.*, 2005) without a GC

clamp. Sequence reaction products were visualized using Bioedit sequence alignment editor and analysis program (Hall, 1999) and similarity to sequences deposited in the GenBank, EMBL, and DDBJ databases was verified by using the program BLAST (Altschul *et al.*, 1990) (www.ncbi.nlm.nih.gov/blast/). The dataset was analyzed using Bayesian inference of phylogeny (Lewis, 2001), a model-based phylogenetic method. The evolutionary model used was selected by applying PAUP* (Swofford, 2001) and Modeltest (Posada and Crandall, 1998). The general time-reversible model (GTR) with gamma distribution of rates and a proportion of variable sites was selected. All model parameters were estimated in the Bayesian inference analysis using the program Mr Bayes (Huelsenbeck *et al.*, 2005).

2.2.8 DGGE STANDARD

Estimated from agarose gel electrophoresis, approximately 100 ng · μl^{-1} of PCR product from 7 well resolved single band isolates was mixed to produce a DGGE standard. When electrophoresed in DGGE, each band in the mixture was representative of one DNA sequence at a single position on the gel and associated with one or more of the 23 cyanobacterial isolates used in this study

2.3 RESULTS

2.3.1 ISOLATE MORPHOLOGY AND PHYSIOLOGY

Cell sizes of vegetative, heterocyst, and akinete cell types were measured and compared to expected size ranges as described by their authorities in Wehr and Sheath, 2003 and John *et al.*, 2002. In addition, filament morphology and mucilage characteristics were observed and compared. All isolates had characteristics that were consistent with published morphological data for each species with most differences being observed in cell sizes and filament morphology (Figure 7 A-W).

Descriptions of cyanobacteria Isolates

Anabaena flos-aquae **Breb ex. Bornet et Flah.UTEX LB 2383**

Vegetative cells spherical to barrel-shaped, green in colour, some gas vacuoles present, 5.7 μm long, and 4.1 μm wide. Heterocysts spherical, smaller than vegetative cells, 4.3 μm long, 4.2 μm wide. No akinetes or mucilage observed. Filaments mostly short (4-10 cells) and tangled with other filaments. No geosmin production. (Figure 7A)

Anabaena flos-aquae **Breb ex. Bornet et Flah. UTCC 64**

Vegetative cells spherical to barrel-shaped, dark green in colour, some gas vacuoles present, 5.8 μm long, and 3.3 μm wide. Heterocysts spherical, smaller than vegetative cells, 4.5 μm long, 4.3 μm wide. No akinetes or mucilage observed. Filaments long (>20 cells) and tangled with other filaments. No geosmin production. Vegetative cells appeared to change drastically in morphology throughout the culture period. Some cells grew very large, contorted and very irregular in shape compared to the remaining filaments in the culture. There was a large degree of phenotypic plasticity in this culture. (Figure 7B)

Anabaena lemmermannii **Richter AL4**

Vegetative cells cylindrical, green in colour, gas vacuoles present, 8.6 µm long, and 5.5 µm wide. Heterocysts cylindrical, slightly smaller than vegetative cells, browner in colour than vegetative cells, 8.3 µm long, 5.4 µm wide. No akinetes or mucilage observed. Filaments ranged from short (<10 cells) to long (>20 cells) and were straight, solitary and free-floating. No geosmin production (Figure 7C).

Anabaena lemmermannii **Richter AL5**

Vegetative cells cylindrical, green in colour, gas vacuoles present, 6.0 µm long, and 4.3 µm wide. Heterocysts longer than vegetative cells, slightly browner in colour, 8.1 µm long, 4.3 µm wide. No mucilage or akinetes observed. Filaments short (<10 cells) to intermediate (10-20 cells) in length and were straight, solitary and free-floating. No geosmin production (Figure 7D).

Anabaena lemmermannii **Richter AL7**

Vegetative cells cylindrical, green in colour, gas vacuoles present, 7.6 µm long, and 6.0 µm wide. No heterocysts, mucilage or akinetes observed. Filaments short (<10 cells) to intermediate (10-20 cells) in length and were straight, solitary and free-floating. No geosmin production (Figure 7E).

Anabaena lemmermannii **Richter LONT2**

Vegetative cells cylindrical, green in colour, gas vacuoles present, 7.0 µm long, and 6.0 µm wide. Heterocysts slightly smaller than vegetative cells, slightly browner in colour, 6.3 µm long, 5.8 µm wide. No mucilage or akinetes observed. Filaments short (<10 cells) to intermediate (10-20 cells) in length and were straight, solitary and free-floating. Geosmin producing (Figure 7F).

***Anabaena lemmermannii* Richter LONT5**

Vegetative cells cylindrical, green in colour, gas vacuoles present, 8.4 μm long, and 5.9 μm wide. Heterocysts slightly larger than vegetative cells, slightly browner in colour, 9.4 μm long, 6.2 μm wide. No mucilage or akinetes observed. Filaments short (<10 cells) to intermediate (10-20 cells) in length and were straight, solitary and free-floating. Geosmin producing (Figure 7G).

***Anabaena lemmermannii* Richter GIOL8**

Vegetative cells cylindrical, green in colour, gas vacuoles present, 5.7 μm long, and 5.1 μm wide. No heterocysts, mucilage or akinetes observed. Filaments very long (>20 cells), helical in shape, and were solitary and free-floating. Geosmin producing (Figure 7H).

***Anabaena* sp. LOW**

Vegetative cells barrel-shaped, dark green in colour, somewhat bent, gas vacuoles present, 12.0 μm long, and 4.6 μm wide. Heterocysts smaller than vegetative cells, slightly browner in colour, 8.0 μm long, 5.6 μm wide. Akinetes much larger than vegetative cells, 19.5 μm long, 6.2 μm wide. No mucilage observed. Filaments short (<10 cells) to intermediate (10-20 cells) in length and were bent and tangled with other filaments. No geosmin production (Figure 7I).

***Anabaena planktonica* Brunneth**

Vegetative cells spherical to barrel-shaped, green in colour, some gas vacuoles present, 6.3 μm long, and 6.1 μm wide. Akinetes much larger than vegetative cells, cylindrical in shape, 16.9 μm long, and 12.9 μm wide. No heterocysts observed. Thin mucilage layer around filament, more obvious around akinete and adjacent cells. Filaments were long (>20 cells) in length, straight, solitary and free-floating. No geosmin production (Figure 7J).

Anabaena spiroides **Klebhan**

Vegetative cells spherical to barrel-shaped, dark green in colour, few gas vacuoles present, 7.2 µm long, and 4.7 µm wide. Heterocysts round and slightly wider and lighter than vegetative cells, 6.6 µm long, 5.4 µm wide. Thin mucilaginous layer surrounding filament. Filaments varied from short (<10 cells) to long (>20 cells) in length, solitary, somewhat bent and tangled. No geosmin production (Figure 7K).

Anabaena viguieri **Denis et Frémy**

Vegetative cells barrel-shaped, dark green in colour, wider than longer, 4.7 µm long, and 5.6 µm wide. Heterocyst spherical, larger than vegetative cells, yellowish in colour, 9.0 µm long, 8.4 µm wide No akinetes observed. Slight mucilaginous layer surrounding filament. Filaments long (>20 cells) in length, straight, solitary, free-floating. No geosmin production (Figure 7L).

Anabaena cylindrica **Lemmermann**

Vegetative cells barrel-shaped with more squared-off corners, end cells slightly tapered, dark green in colour, some gas vacuoles, 7.9 µm long, and 3.9 µm wide. Heterocyst lighter in colour, slightly larger than vegetative cells, 7.3 µm long, 4.7 µm wide. No akinetes observed. Slight mucilaginous layer surrounding filament. Filaments long (>20 cells) in length, straight, aggregated adjacent to other filaments to form a colony. No geosmin production (Figure 7M).

Trichormus variabilis **Kutzing**

Vegetative cells barrel-shaped, dark green in colour, not uniformly sized along filament, wider than long, 3.2 µm long, and 4.1 µm wide. No heterocysts, mucilage or akinetes

56

observed. Filaments varied from short (<10 cells) to long (>20 cells) in length, straight, solitary, free-floating. No geosmin production (Figure 7N).

Aphanizomenon flos-aquae **Linneaus Hamilton Harbour**

Vegetative cells barrel-shaped with somewhat squared-off corners, dark green in colour, end cells tapered, many gas vacuoles, 9.2 µm long, and 5.1 µm wide. Heterocyst lighter in colour and slightly wider than vegetative cells, 8.5 µm long, 5.4 µm wide. No akinetes observed. Thin mucilaginous sheath surrounding filament. Filaments varied from intermediate (10-20 cells) to long (>20 cells) in length, straight, aggregated or solitary. No geosmin production (Figure 7O).

Aphanizomenon gracile **Lemmermann**

Vegetative cells barrel-shaped with somewhat squared-off corners, dark green in colour, end cells tapered, many gas vacuoles, 9.2 µm long, and 5.1 µm wide. Heterocyst lighter in colour, slightly larger than vegetative cells, 9.8 µm long, 5.8 µm wide. No akinetes or mucilage observed. Filament long (>20 cells) in length, straight, aggregated adjacent to other filaments to form a colony. Geosmin producing (Figure 7P).

Calothrix parietina **(Thuret) Bornet et Flahault PCC 6303**

Vegetative cells barrel-shaped with somewhat squared corners, light to dark green in colour, 4.0 µm wide, shorter than wide towards the base of the trichome, and longer than wide in the middle of the trichome. Basal vegetative cells very dark (almost black) in colour, wider than rest of filament, 7.8 µm long. and 5.9 µm wide. Heterocysts were basal to the entire trichome, light green in colour, 7.0 µm long, and 4.8 µm wide. No intercalary heterocysts or akinetes observed. Filaments were wide at the base and tapered towards the

apex with sizes varying from intermediate (10-20 cells) to long (>20 cells) in length. Mucilaginous sheath surrounds entire filament. No geosmin production (Figure 7Q).

Calothrix sp. PCC 7507

Vegetative cells barrel-shaped with somewhat squared-off corners, green in colour, 4.1 μm wide, shorter than wide towards the base of the trichome, and longer than wide in the middle of the trichome. Basal heterocysts were light green in colour, 4.0 μm long, and 3.4 μm wide. Intercalate heterocysts were larger than basal, light green in colour, 7.2 μm long, and 5.2 μm wide. Mucilaginous sheath surrounded entire filament. No akinetes observed. Filaments varied from intermediate (10-20 cells) to long (>20 cells) in length. Geosmin producing (Figure 7R).

Microcystis aeruginosa Kutzing

Cells spherical, many gas vesicles, very dark green in colour. Cell diameter 3.9 μm with a thick and transparent mucilaginous layer (~1 μm). Cells were single and aggregated to form colonies. No geosmin production (Figure 7S).

Microcystis wesenbergii Komárek

Cells spherical, some gas vesicles, green in colour. Cell diameter 6.5 μm with a thick and transparent mucilaginous layer (~1 μm). Cells were single and aggregated to form colonies. No geosmin production (Figure 7T).

Planktothrix aghardii (Gomont) Anagnostidis et Komárek

Vegetative cells mostly shorter than wide, brownish in colour, crosswalls not narrowed, terminal cells rounded and smaller, gas vacuoles present, 4.3 μm long, and 6.4 μm wide. No mucilage observed. Filaments very long (>20 cells), straight, solitary, and free-floating. No geosmin production (Figure 7U).

Oscillatoria limosa (C. Aghard) Gomont

Vegetative cells mostly shorter than wide, brownish in colour, crosswalls not narrowed, end cell rounded and smaller, gas vacuoles present, 4.1 μm long, 10.2 μm wide. Thin mucilaginous layer surrounding entire filament. Filaments very long (>20 cells), straight, solitary, free-floating. Geosmin producing (Figure 7V).

Pseudanabaena sp. UTCC 593

Vegetative cells cylindrical to barrel-shaped, constricted at crosswalls, bright green in colour, apical cells truncate, 4.1 μm long, and 2.0 μm wide. No mucilage observed. Filaments varied from intermediate (10-20 cells) to long (>20 cells) in length, straight, solitary, free-floating. No geosmin production (Figure 7W).

Figure 7. Morphology of cyanobacteria isolates used in this study. *Anabaena flos-aquae* UTEX LB 2383 (A); *Anabaena flos-aquae* UTCC 64 (B); *Anabaena lemmermannii*: AL4, AL5, AL7, LONT2, LONT5, GIOL8 (C-H); *An*abaena sp. LOW (I); *Anabaena planktonica* (J); *Anabaena spiroides* (K); *Anabaena viguieri* (L). Heterocysts are indicated as Het. Akinetes are indicated as Ak.

A

Het

B

C

Het

D

Het

E

F

Het

G

Het

H

I

Het

J

Ak

K

Het

10 µm
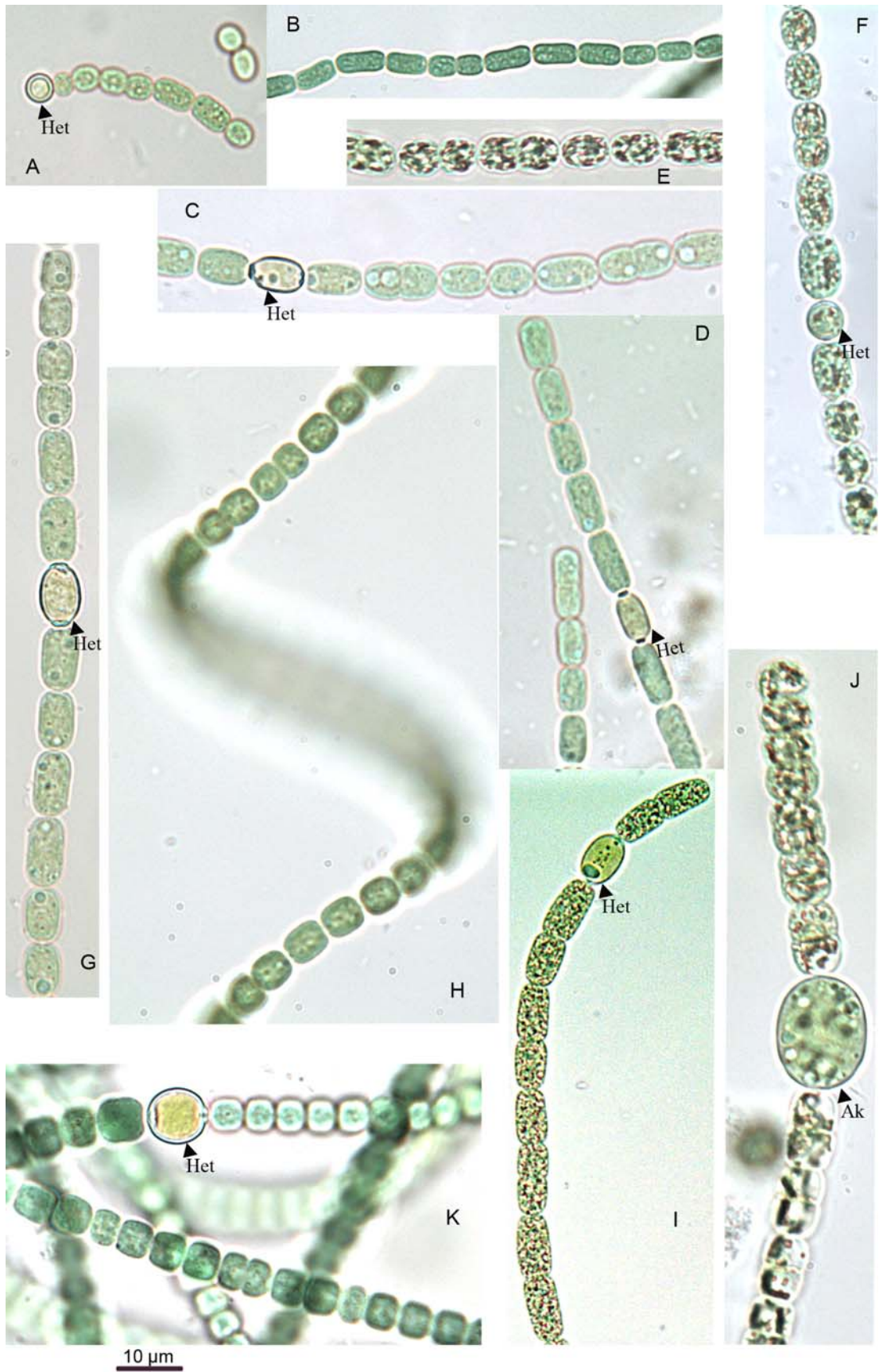
Figure 7 continued. Morphology of cyanobacteria isolates used in this study.

Anabaena cylindrica (M); *Trichormus variabilis* (N); *Aphanizomenon flos-aquae*

Hamilton Harbour (O); *Aphanizomenon gracile* (P); *Calothrix parietina* PCC 6303 (Q);

Calothrix sp. PCC 7507 (R); *Microcystis aeruginosa* (S); *Microcystis wesenbergii* (T);

Planktothrix aghardii (U); *Oscillatoria limosa* (V), *Pseudanabaena* sp. UTCC 593 (W).

Heterocysts are labelled as Het.
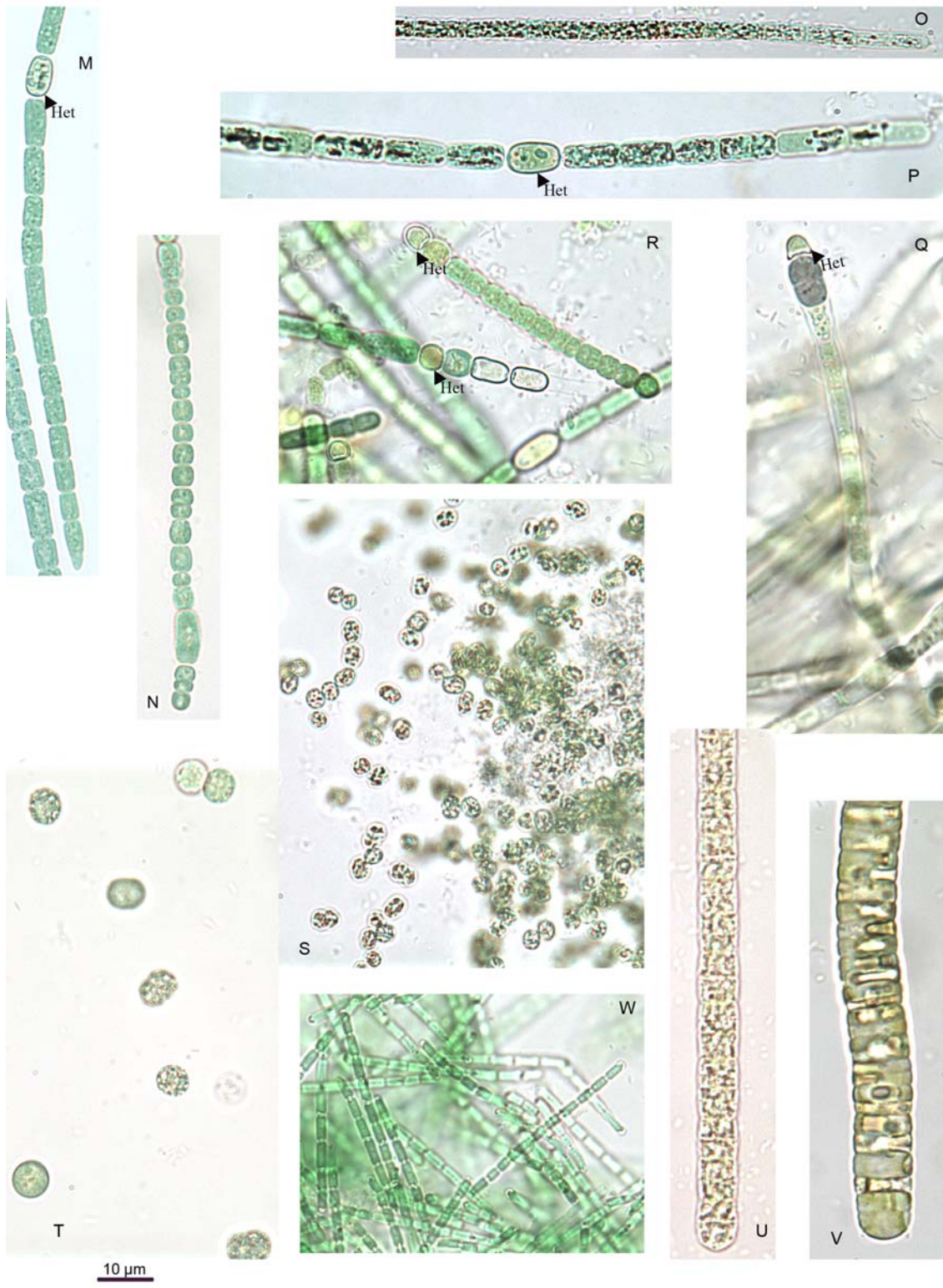
2.3.2 16S rRNA GENE BLAST RESULTS

The 16S rRNA gene BLAST results of the nearly complete 16S rRNA gene (~1203 bp) all corresponded to cyanobacteria within the morphologically identified genus and to the species level to most of the isolates with E-values approaching 0.0 for all queries (meaning that the sequences were nearly identical). Of the 23 isolates, 20 were previously morphologically identified to the species level while the three remaining were identified based on its genus or culture collection. A BLASTn query returned an exact match to the morphologically identified species of 18 isolates all with E-values approaching 0.0 for each. The two sequences not returning an exact species match were *Anabaena viguieri* and *Oscillatoria limosa*. Both matched to members of the same genus but not species even though their lineages did have 16S rRNA sequences in GenBank.

Previous morphological identification of *Pseudanabaena* sp. UTCC 593 and *Anabaena* sp. LOW was not accomplished with confidence to the species level due to a lack of characteristics, however their BLASTn results had strong E-values approaching 0.0 and matched to many members of their genus. *Pseudanabaena* sp. UTCC 593 returned five matches to *Pseudanabaena* sp. of the Pasteur Culture Collection (PCC) however none of these were identified to the species level. Many of the query results for *Anabaena* sp. LOW were matches to *Aphanizomenon flos-aquae*, *Anabaena lemmermannii*, and *Anabaena circinalis.*

Calothrix parietina PCC 6303 and *Calothrix* sp. PCC 7507 were obtained from the Pasteur Culture Collection and neither have 16S rRNA sequences in GenBank. BLAST results for *Calothrix parietina* PCC 6303 returned matches to four members of the *Calothrix* genus with two not identified to the species. One of the unidentified species of *Calothrix* was a

64

strain from the Pasteur Culture Collection; PCC 7714 while the other *Calothrix* strain was

of an unpublished source likely isolated from a benthic source in Belgium or Luxembourg

according to the title of the research (Willame *et al.*, 2005 submitted) in the GenBank

flatfile. One BLAST result returned a match to *Calothrix desertica* PCC 7102 while the

other match was to a *Calothrix* cf. *muscicolous* cyanobiont of lichen. All had E-values of

0.0. The *Calothrix* sp. PCC 7507 16S rRNA sequence was not available in GenBank and

did not have an exact match in BLAST. The majority of the BLAST results returned

matches to members of the Nostocales including *Trichormus variabilis*, and various *Nostoc*

sp. and *Anabaena* sp. all with E-values approaching 0.0. One match did correspond to

Calothrix brevissima with an E-value approaching 0.0. Sequence similarity of all returned

BLASTn taxa ranged from 94.0-100.0%.

2.3.3 SEPARATION OF CYANOBACTERIA STRAINS USING 16S RRNA DGGE

For the *An. lemmermannii* AL5 and *An. planktonica* samples electrophoresed in the parallel

DGGE, the best resolving range of denaturant was from 35-55%. This denaturant range

differentiated 23 cyanobacteria isolates encompassing 17 different species. The resulting

DGGE gel (Figure 9) demonstrated that the V3 region used was variable among the

isolates but it does not differ among all species. Many strains co-migrated with those of the

same species as well as different species and genera. Of the 17 isolates, 11 unique positions

in the gel were observed. Sequence analysis of co-migrated isolate DGGE bands resulted in

six groupings in which more than one isolate possessed an identical sequence. The majority

rule consensus (50%) unrooted Baysian likelihood phylogenetic tree (Figure 8) shows the

grouping of isolates with identical V3 sequences, those not grouped had unique sequences.

Previous gels (not shown) included *An. flos-aquae* UTEX LB 2383, *An. lemmermannii*

AL5 and AL7, as well as *An. viguieri* and showed co-migration of these strains to those

depicted in Figure 9. *An. flos-aquae* UTEX LB 2383 co-migrated with *An. flos-aquae*

UTCC 64 to form the group 1 band position. *An. lemmermannii* AL5, AL7 and *An. viguieri*

co-migrated with the other *An. lemmermannii* strains (AL4, LONT2, LONT5, GIOL8, *An.*

planktonica, Tr. variabilis) to form the group 2 band position. The completed DGGE

standard did not include every sample. Instead one isolate from each group was chosen as a

representative of that group in addition to *Osc. limosa* and *Calothrix* PCC 6303, which were

also chosen for the standard because their migration was distinctly separate from the other

6 bands. The final DGGE standard was composed of 7 DGGE fragments and included

Calothrix PCC 6303, *An. flos-aquae* UTCC 64, *An. lemmermannii* LONT5, *An. cylindrica*,

Anabaena sp. LOW, *Ap. flos-aquae* Hamilton Harbour, and *Osc. limosa* (Figure 10). The

third, fourth, fifth and seventh positions can be representative of potential geosmin-

producing species of cyanobacteria. The standard mixture was used in the identification of

potential geosmin-producing cyanobacteria in Lake Ontario water samples and is discussed

in chapter 3.

Figure 8. Majority rule consensus (50%) of the trees sampled in the Bayesian analysis based on 16S rRNA-V3 DNA sequences (128 bp). Estimates of clade probabilities are indicated above the branches. Each labelled group included isolates with identical 16S rRNA-V3 regions. Ungrouped strains had unique sequences for this region. One member of each group was chosen as a representative strain for its position in the gel. *Calothrix parietina* PCC 6303 and *Oscillatoria limosa* were included as unique band positions in the DGGE standard.
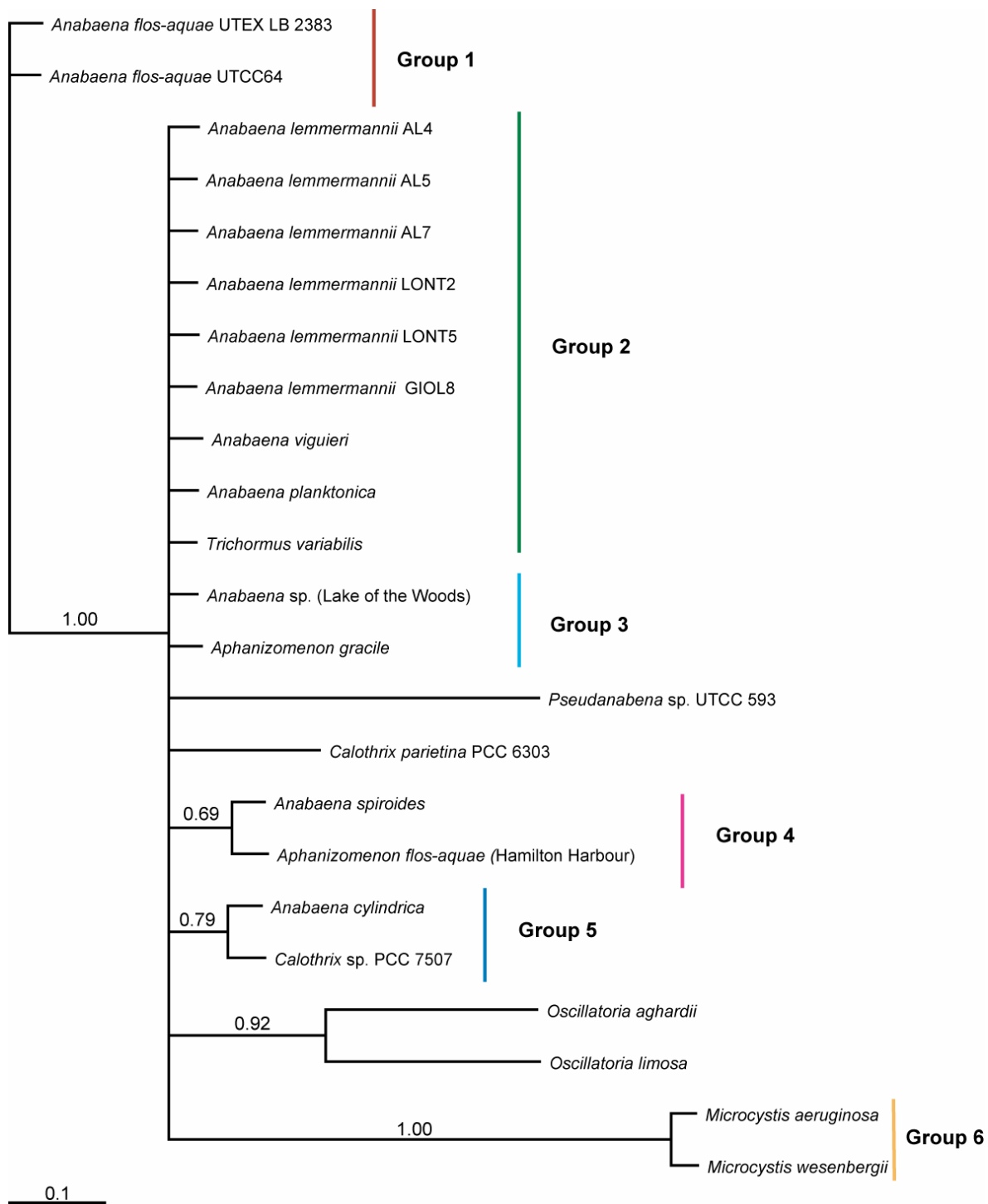
Figure 9. Separation of the 16S rRNA-V3 gene for 19 cyanobacteria isolates in a 35-55% denaturing gradient gel. The cyanobacterial strains in each lane include; *Anabaena* sp. LOW (lane 1), *An. lemmermannii* LONT2, LONT5, AL4, GIOL8 (lanes 2-5), *An. planktonica* (lane 6), *An. spiroides* (lane 7), *An. cylindrica* (lane 8), *An. flos-aquae* UTCC64 (lane 9), *An. variabilis* (lane 10), *Ap. flos-aquae* Hamilton Harbour (lane 11), *Ap. gracile* (lane 12), *Pseudanabaena* sp. UTCC 593 (lane 13), *M. aeruginosa* (lane 14), *M. wesenbergii* (lane 15), *C.*sp. PCC 7507 (lane 16), *C. parietina* PCC 6303 (lane 17), *Osc. aghardii* (lane 18), *Osc. limosa* (lane 19). Information about the strains is given in Table 1 (Cultures of cyanobacteria used in this study).
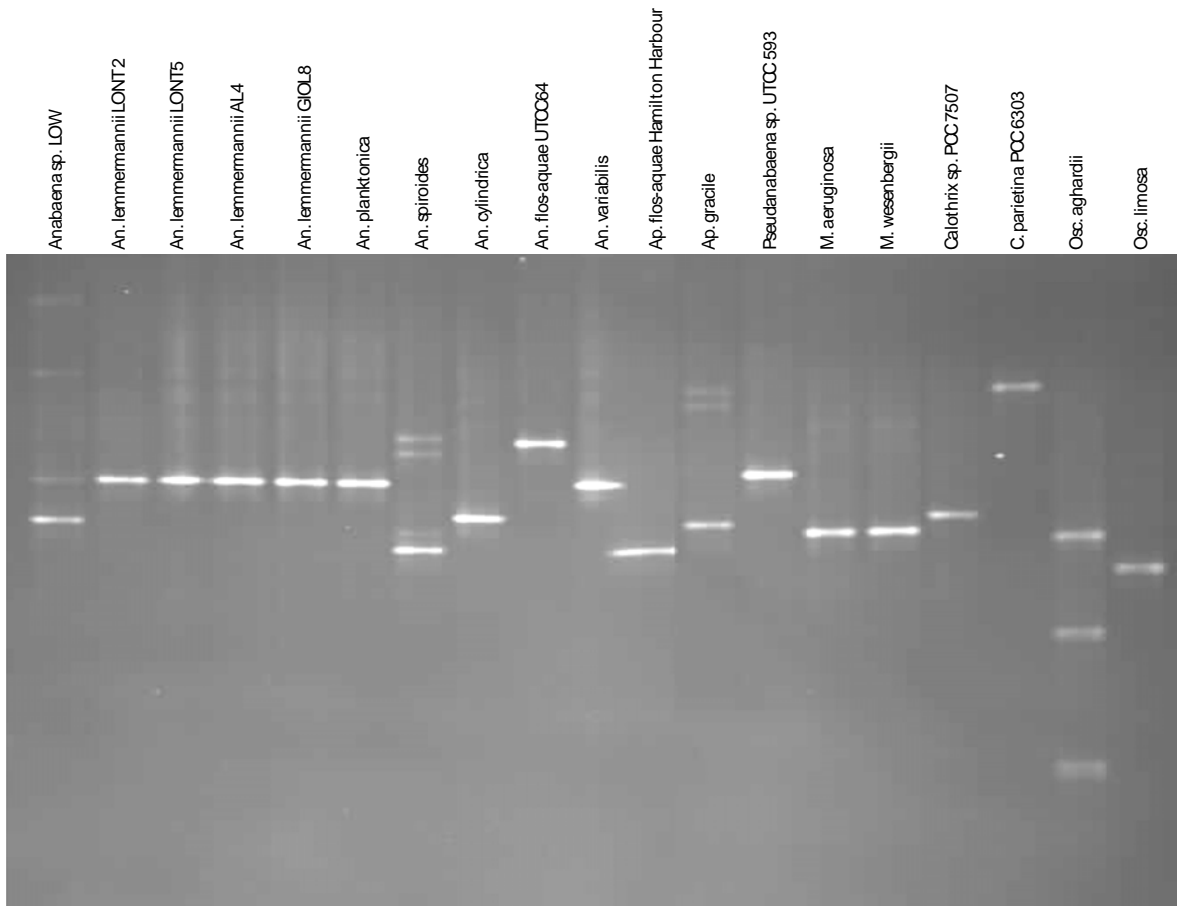
Figure 10. Completed 16S rRNA-V3 DGGE standard. Mixed standard to be used for identification of strains in Lake Ontario environmental samples (lane S), *C. parietina* PCC 6303 (lane 1), *An. flos-aquae* UTCC64 (lane 2), *An. lemmermannii* LONT5 (lane 3), *An. cylindrica* (lane 4), *An* sp. LOW (lane 5), *Ap. flos-aquae* Hamilton Harbour (lane 6), *Osc. limosa* (lane 7). The asterisk indicates a potential geosmin-producing species at that position in the gel.
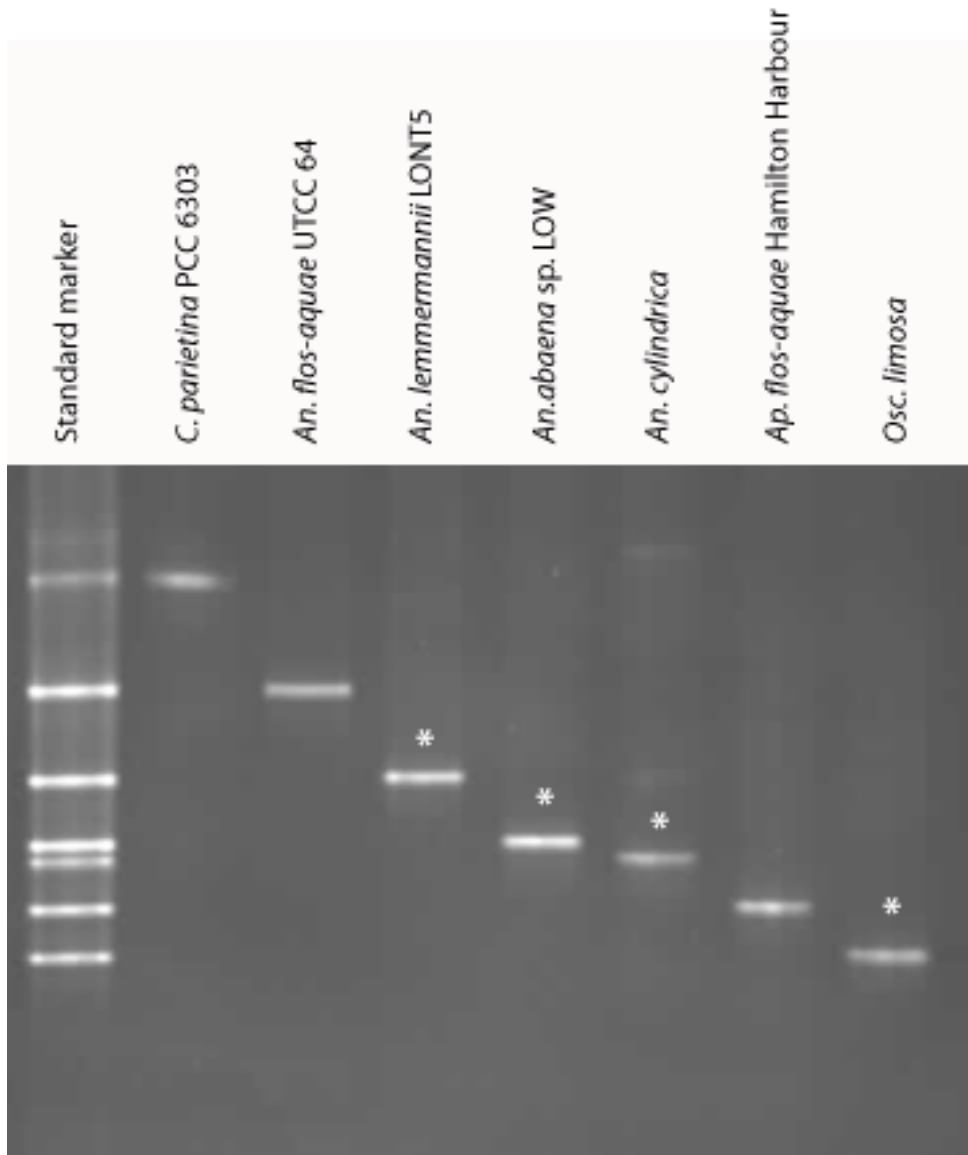
2.4   DISCUSSION

2.4.1 *CYANOBACTERIA ISOLATE IDENTITIES*

The morphological identifications of each isolate were compared to those described by
their authorities outlined in (Wehr and Sheath, 2003) and John et al. (2002).
Morphological discrepancies among isolate strains are likely due to plasticity that is
commonly exhibited by laboratory cultured isolates over prolonged periods. It is not
uncommon for cultured strains to alter their characteristics from that observed in nature
when grown under controlled environments (Komárek and Anagnostidis, 1989).
Additionally, discrepancies among GenBank nucleotide sequences and those sequenced
in this study do not necessarily suggest that an incorrect identification has been made for
the isolates in this study. Not all sequences in GenBank have been correctly identified
often due to a lack of morphological characteristics when being observed. Entries in the
GenBank database were not considered absolute and are regarded only as a guide for
morphological identification in this study.

Description of cyanobacteria isolates

***Anabaena flos-aquae* UTEX LB 2383 and UTCC 64**

Morphologically, these strains were very consistent with the type strain *Anabaena flos-
aquae* (Breb. Ex. Born. et Flah). Although no akinetes were observed in the cultures, the
vegetative cells were within the range of 2.5-8.3 μm long and 2.5-7 μm wide as
expected. The spherical heterocysts from the isolate were slightly smaller by ~0.5 μm
overall than the expected range of 5.0-8.5 μm long and wide, however this difference
was likely insignificant. Filaments were not surrounded in mucilage and were tangled as

expected. Both of these strains did not produce geosmin in culture and have not been known to do so (Watson unpublished). The majority of the filaments were short in the UTEX LB 2383 culture likely due to prolonged laboratory growth conditions at the time of examination. Throughout the course of culturing, these filaments have ranged from short to long in length. The BLASTn query of the UTEX LB 2383 16S rRNA sequence resulted in three matches to *Anabaena flos-aquae* strains with E-values approaching 0.0 and sequence identities ranging from 98.36-94.00% indicating that the sequence derived from this isolate was likely from the *Anabaena flos-aquae* genome. Of the three matches, one was the exact strain UTEX LB 2383 from which the isolate was obtained, providing even stronger support for the correct identification in this study. The BLASTn query of the UTCC 64 16S rRNA sequence resulted in an exact match to the UTCC 64 isolate sequence from GenBank with an E-value approaching 0.0 and sequence identity of 96.88%, also providing stronger support for the identification of this species. The consistency of morphological characteristics as well as identities to the exact 16S rRNA genes in GenBank is indicative that both isolates were correctly identified in this study.

Anabaena lemmermannii **AL4, AL5, AL7**

Morphologically, these strains were very consistent with the type strain *Anabaena lemmermannii* P. G. Richter. Although no akinetes were observed in these cultures, the vegetative cells were within the expected range of 2.6-12.6 µm long and 2.9-6.1 µm wide. The heterocysts from isolates AL4 and AL5 were longer by approximately 3µm than the expected range of 4.5-5.2 µm long and wide. Phenotypic plasticity may explain the slight discrepancy observed between these two isolates. No heterocysts were observed in the AL7 isolate and could not be assessed. None of the filaments were

surrounded in mucilage and all were straight, and solitary as expected. Many strains of *An. lemmermannii* have been known to produce geosmin and in Lake Ontario have been suspected as the primary contributors to taste and odour in drinking water (Watson, 2003). Although geosmin-producing strains exist, the unique characteristic of these three *An. lemmermannii* strains is that they have been isolated from Lake Ontario but have never produced geosmin in culture (Watson, unpublished).

All three 16S rRNA sequences obtained from these isolates were identical to one another. The BLASTn query of the sequences resulted in four matches to *An. lemmermannii* strains with E-values approaching 0.0 and sequence identities ranging from 99.46-98.57% indicating that the sequences derived from these isolates were likely from the *An. lemmermannii* genome. Consistencies of morphological characteristics as well as 16S rRNA sequence identities to *An. lemmermannii* in GenBank provide strong support for an accurate identification of these isolates.

Anabaena lemmermannii **LONT5**

This strain was isolated from Lake Ontario and was very similar morphologically to AL4, AL5, AL7 however there were differences that required a separate discussion of this particular isolate. The primary difference between this strain and the AL strains is that LONT5 produced geosmin in culture while none of the AL strains produced geosmin. When the culture was first obtained no geosmin was detected, however over the course of approximately six months, geosmin production was somehow activated. Growth conditions were always consistent for all isolates with no significant changes being noted from the time of receiving the culture to the start of geosmin production. No

factors were directly correlated to the induction of geosmin production for this isolate. This characteristic made it unique in comparison to the other *An. lemmermannii* isolates.

Vegetative cell sizes were consistent with the type strain *Anabaena lemmermannii*. Heterocysts were longer by approximately 4 μm and wider by 1 μm than the expected size range of 4.5-5.2 μm length and width. No akinetes were observed likely due to adequate growth conditions. The filaments were long, straight, solitary, free-floating and were not surrounded in mucilage, all consistent with the expected morphology.

The BLASTn query of the 16S rRNA gene sequence resulted in the same four matches to *An. lemmermannii* strains, as did the AL strains. All *An. lemmermannii* results returned E-values approaching 0.0 and sequence identities ranging from 99.38-98.48% indicating that the sequence derived from this isolate was likely from the *An. lemmermannii* genome. Consistencies of morphological characteristics as well as 16S rRNA sequence identities to *An. lemmermannii* in GenBank provided strong support for an accurate identification of this isolate.

Anabaena lemmermannii LONT2

This culture was isolated from Lake Ontario and the morphology was very similar to that of the AL and LONT5 isolates with the only difference being that it was a continuous geosmin producer. Vegetative cell sizes and shape were consistent with the type strain *Anabaena lemmermannii*. Heterocysts were slightly larger overall by approximately 0.5 μm, not a significant amount. No akinetes were observed likely due to adequate growth conditions. In addition, the filaments were straight, solitary, free-floating, and were not

surrounded by a mucilaginous layer, all of which were consistent with the expected morphology of *An. lemmermannii.*

The BLASTn query of the 16S rRNA gene sequence resulted in three matches to *An. lemmermannii* strains with two identical to those returned from the AL and LONT5 isolate 16S rRNA BLASTn. All *An. lemmermannii* results returned E-values approaching 0.0 and sequence identities ranging from 99.52-99.04% indicating that the sequence derived from this isolate is likely from the *An. lemmermannii* genome. Consistencies of morphological characteristics as well as 16S rRNA sequence identities to *An. lemmermannii* in GenBank provide strong support for an accurate identification of this isolate.

Anabaena lemmermannii GIOL8

Vegetative cell sizes were consistent with the type strain *Anabaena lemmermannii* however since no heterocysts or akinetes were observed, it was not possible to assess these characteristics. This strain was unique to the five other *An. lemmermannii* strains as it was not isolated from Lake Ontario but from a location in California, USA. It was a geosmin producer in culture, and its filament morphology was very different from the other *An. lemmermannii* isolates. It was expected that this strain would possess straight, solitary filaments however all filaments, whether short or long, in length were helical. This helical nature was similar to what would be expected of *An. spiroides* Klebahn however the BLASTn results of the 16S rRNA gene returned four *An. lemmermanni* strains with three of these matches being the same as for the AL isolates. Sequence identities for the four sequences ranged from 99.29-98.66% indicating that the sequence derived from this isolate was likely from the *Anabaena lemmermannii* genome.

The helical filament may be a product of maintenance under laboratory conditions and may not represent the natural morphology of the strain. Consistencies of morphological characteristics as well as 16S rRNA sequence identities to *An. lemmermannii* in GenBank provide strong support for an accurate identification of this isolate.

Anabaena sp. LOW

It was not possible to conclude with confidence the species to which this isolate belonged however it was tentatively identified to be *An. lemmermannii* (Watson and Kling, unpublished). In comparison to the six *An. lemmermanni* isolates used in this study, the vegetative cell sizes were much longer for *Anabaena* sp. LOW averaging 12.0 μm whereas the average length of the other isolates was approximately 5 μm shorter. The average width of the vegetative cells for this strain was in the same range as the other isolates and overall the morphology and vegetative cell sizes were within the expected range for the type strain *Anabaena lemmermannii*. The heterocysts were larger in comparison to the type *An. lemmermannii* although the sizes were within the same range as the heterocysts from the other *An. lemmermannii* isolates used in this study. Discrepancies in the slightly larger heterocyst size could be due to culturing under laboratory conditions. The average size of the akinetes observed in this culture fell within the expected range for *An. lemmermannii* with an average length of 19.1 μm and a 6.2 μm width. This isolate did not produce geosmin in culture, and the filaments were tangled and not surrounded by mucilage which differs from that observed in the *An. lemmermannii* isolates used in this study.

A BLASTn of the partial 16S rRNA gene resulted in *Ap. flos-aquae*, *An. lemmermannii*, *An. circinalis*, and other unidentified *An*abaena sp. matches with

sequence identities ranging from 99.36-99.20%. A greater number of *Anabaena* species were amongst the BLAST results than *Aphanizomenon* and with the morphological data this isolate is not likely of the *Aphanizomenon* genus. It is very probable that this strain is a species of *Anabaena,* however the speculation of this isolate belonging to *An. lemmermannii* may be incorrect. Future research could focus on sequence analysis of additional gene regions in order to obtain a well supported species inference.

Anabaena planktonica

Morphologically, this strain conformed to many of the characteristics of the type strain *An. planktonica* Brunneth. The vegetative cells were slightly smaller (approximately 3 µm long and wide) overall than the expected size of 10-12 µm long and 9-15 µm wide. No heterocysts were observed, however the average akinete cell size was within the expected range of 15-30 µm long and 10-20 µm wide. Filaments were straight, solitary, free-floating and were surrounded by a thin mucilaginous layer that was more obvious around the akinete, all consistent characteristics of the type strain (John *et al.*, 2002). *An. planktonica* is not known as a geosmin producer (Watson, 2003) and this isolate did not produce geosmin in culture.

The 16S rRNA gene BLASTn produced nine matches *An. planktonica* with sequence similarities of 99.91-99.82%. These BLASTn results strongly supported that the 16S rRNA gene was derived from the *An. planktonica* genome. Consistencies of morphological characteristics as well as 16S rRNA sequence identities to *An. planktonica* in GenBank provided strong support for an accurate identification of this isolate.

Anabaena spiroides

Morphologically, this strain was very consistent with the type strain *Anabaena spiroides* Klebahn. Although no akinetes were observed in the cultures, the vegetative cells were within the range of 6.0-8.0 µm long and but were slightly smaller in width by approximately 1.5 µm than the expected 6.0-8.0 µm. The spherical heterocysts from the isolate were within the expected range of 6.0-8.0 µm and were consistent with the type specimen. The major difference in this isolate was observed in the filament morphology. As expected from the name, the filament should be regularly coiled with the same distance between the coils, however that was not seen in this isolate. Instead, the filaments were tangled and did not have a particular pattern of morphology. A thick mucilaginous layer was anticipated from filaments in this culture, however the layer was present but thinner than expected. Morphological discrepancies within this isolate was could have been due to culturing under laboratory conditions and is probably not representative of the strain in its natural habitat. *An. spiroides* is not known to be a geosmin-producing species and as expected, this isolate was not a geosmin-producing strain.

The 16S rRNA gene BLAST results returned several similar sequences from *An. compacta* and *An. circinalis* with sequence identities of 99.82-98.12%, but there were also three *An. spiroides* matches with sequence identities of 97.76%. It was expected that the *An. spiroides* matches from GenBank would have greater sequence identity than *An. compacta* and *An. circinalis*, nonetheless there still remained support for the identification of this strain as *An. spiroides* with the combination of the morphological and 16S rRNA gene sequence data and it is probable that this strain has been identified

correctly for use in this study. Until more genes are sequenced and analyzed from this isolate there was not sufficient data to confidently rule-out a misidentification.

Trichormus variabilis

Anabaena variabilis (Kutzing) Bornet et Flahault has been revised and combined and named as *Trichormus variabilis* (Ralfs ex Bornet *et* Flahault) Komárek *et* Anagnostidis. The isolate used in this study conformed to *Trichormus variabilis* (Ralfs ex Bornet *et* Flahault) Komárek *et* Anagnostidis with vegetative cell sizes falling within the range of 2.5-6.0 µm long and 4.0-6.0 µm wide and very few, if any, gas vacuoles. No heterocysts or akinetes were observed in this culture and could not be assessed. The filament was long, straight, solitary, free-floating and without mucilage as was expected from the described type. No literature was found that describes *Tr. variabilis* as a geosmin-producing species and as expected this strain did not produce geosmin in culture. The 16S rRNA gene sequence results returned seven sequence matches to *Tr. variabilis* with a range of 100.00-98.84% sequence similarity. The *Tr. variabilis* strain from GenBank with 100% sequence identity was identified and sequenced by Rajaniemi, et al. (2005) and included the author Jiri Komárek who revised this genus. There was strong support from both morphological and molecular data suggested that the strain used in this study was not likely misidentified. This strain was originally identified as *An. variabilis*, however with the 16S rRNA results, the name was updated and changed to *Tr. variabilis*.

Anabaena viguieri

This strain conformed to the type specimen *Anabaena viguieri* (Denis et Frémy) with its vegetative cells ranging from 4.0-8.5 µm long and 6.0-7.0 µm wide. The filaments were

long, straight, solitary, free-floating, and had a very slight layer of mucilage surrounding the cells. No akinetes were observed however the heterocysts were larger overall by approximately 4 μm than the expected range of 4.0-5.5 μm long and 5.0-7.0 μm wide. *An. viguieri* has been known to producer geosmin (Watson, 2003) however this isolate did not produce geosmin over the course of this study.

The 16S rRNA BLASTn results did not show any matches to *An. viguieri* however there was only one GenBank sequence for this strain and if it was misdentified then it is possible that there were no accurate *An. viguieri* 16S rRNA sequences in GenBank. A pairwise sequence alignment performed using Bioedit© calculated a 94.9% sequence identity of the two *An. viguieri* strains, not as much similarity as was calculated for *Tr. variabilis* matches with 98.04-99.02% sequence identity. This was indicative that this isolate may be incorrectly identified however without an increased number of *An. vigiueri* 16S rRNA sequences in GenBank it is not possible to be certain whether this strain was incorrectly identified and will be considered *An. viguieri* for this study.

Anabaena cylindrica

This isolate conformed morphologically to the type strain *Anabaena cylindrica* Lemmermann with its vegetative cells within the expected range of 5.0-8.5 μm long and 3.0-4.5 μm wide. No akinetes were observed and heterocysts were slightly longer by approximately 2.5 μm than the expected spherical morphology and size of 4.5-5.0 μm diameter. Filaments were long, straight and aggregated adjacent to one another and were surrounded by a thin mucilaginous sheath as expected from the type specimen

description. This isolate did not produce geosmin in culture and as expected no literature has suggested that it is capable of producing this compound.

A BLASTn of the 16S rRNA gene sequence from this isolate resulted in three *An. cylindrica* sequence matches ranging from 99.38-96.97% sequence identity supporting that the sequence was likely derived from this species. Consistencies of morphological characteristics as well as 16S rRNA sequence identities to *An.cylindrica* in GenBank provided strong support for an accurate identification of this isolate.

Aphanizomenon gracile

This isolate was morphologically consistent with the type specimen *Aphanizomenon gracile* Lemmermann with many gas vacuoles, and tapered apical cells, however the size of the vegetative cells and heterocysts were larger than expected. The expected range for vegetative cells was 2.0-6.0 μm long and 2.0-3.0 μm wide however the cells in the isolate were approximately 3 μm larger and wider. In addition, the heterocyst cells were expected to range from 5.5-7.0 μm long and 2.5-3.5 μm wide yet those in culture were also longer and wider by approximately 3 μm. Filaments were long, straight, solitary, or aggregated, and were not surrounded by a mucilaginous sheath as expected from the described type specimen. *Ap. gracile* is a known geosmin-producing species (Watson, 2003) and as expected, the isolate used in this study produced large amounts of geosmin in culture. Support of the 16S rRNA gene indicated that this isolate was correctly identified with three *Ap. gracile* matches and a sequence identity ranging from 99.91-99.73%. Consistencies of morphological characteristics as well as 16S rRNA sequence identities to *Ap. gracile* in GenBank provided strong support for an accurate identification of this isolate.

***Aphanizomenon flos-aquae* Hamilton Harbour**

This isolate was morphologically consistent with the type specimen *Aphanizomenon flos-aquae* (Linneaus) Ralfs , Bornet et Flahault with many gas vacuoles present, and tapered, hyaline apical cells, however the width of the vegetative cells were slightly larger than expected. According to the type specimen, vegetative cells should range from 5.0-15.0 μm long and 5.0-6.0 μm wide yet the cells in this isolate from Hamilton Harbour, in the western basin of Lake Ontario, Canada were smaller in width by approximately 1.5 μm. Heterocyst sizes were consistent with those from the type specimen ranging from 7.0-20.0 μm long and 5.0-7.0 μm wide. In addition, the filament was straight, solitary, free-floating or in bundles, and surrounded by a sheath as expected. *Ap. flos-aquae* has not been known to produce geosmin (Watson, 2003) however the particular isolate used in this study was not a geosmin-producing strain.

The 16S rRNA gene sequence BLASTn results returned nine matches to *Ap. flos-aquae* with sequence identities ranging from 99.73-98.29%. Consistencies of morphological characteristics as well as 16S rRNA sequence identities to *Ap. flos-aquae* in GenBank provide strong support for an accurate identification of this isolate.

Planktothrix aghardii

This isolate conforms to the description given for the type specimen of *Planktothrix aghardii* (Gomont) Anagnostidis & Komárek. The species *Oscillatoria aghardii* was synonymized with *Pl. aghardii* by Anagnostidis et al. (1988). Initially this strain was named *Osc. aghardii* but was then changed to *Pl. aghardii* to conform to the most recent naming. The vegetative cell sizes were within the expected range of 2.5-4.0 μm long and 4.0-6.0 μm wide. The cells were much wider than they are long, did not constrict at the

cross walls, and did not produce heterocysts or akinetes. Filaments were single, slightly attenuated towards apex with the end cell rounded, solitary, free-floating, and without a mucilaginous sheath, all of which are consistent with the expected morphology. Certain strains of *Pl.* aghardii (as *Osc. aghardii*) have been known to produce geosmin (Watson, 2003) however the strain used in this study was not a producer at any point in time. The 16S rRNA gene sequence BLASTn results listed two matches to *Osc. aghardii* with sequence similarities of 99.66 and 99.03%, and nine matches to *Pl. aghardii* with sequence similarities of 99.78-99.03%. The similarities in sequences support the synonymy of *Osc. aghardii* and *Pl. aghardii*. Consistencies of morphological characteristics as well as 16S rRNA sequence identities to *Pl. aghardii* in GenBank provide strong support for an accurate identification of this isolate.

Pseudanabaena sp. UTCC 593

Although this isolate was not identified to the species level, it appeared to have many consistencies with the genus *Pseudanabaena* (Lauterborn) and specifically with the species *P. catenata*. Vegetative cell sizes were within the ranges for *P. catenata* having lengths between 2.0-5.0 µm and widths between 1.8-2.2 µm. Cells were cylindrical to barrel-shaped, constricted at cross walls, and with truncate apical cells as expected in *Pseudanbaena* sp. No heterocysts or akinetes were produced as expected, and the filaments were straight, solitary, free-floating and without a mucilaginous sheath, all of which were consistent with the type genus of *Pseudanabaena* members (John *et al.*, 2002). *P. catenata* has been observed to produce geosmin (Watson, 2003) however the uncharacterized strain used in this study did not produce geosmin at any point in time.

The 16S rRNA gene sequence BLASTn returned matches to five *Pseudanabaena* sp. not identified to the species level. Sequence similarities ranged from 99.37-98.56% with one match having the lowest identity of 90.98%, potentially a misidentified sample in GenBank. E-values were all apporaching 0.0 indicating that this isolate was likely of the *Pseudanabaena* genus. Consistencies of morphological characteristics as well as 16S rRNA sequence identities to *Pseudanabaena* sp. in GenBank provided strong support for an accurate identification of this isolate. No speculations have been made as to which species this isolate belongs; however the morphology was consistent with *P. catenata* and may be a possibility.

Microcystis aeruginosa

This isolate conformed morphologically to the type description of *Microcystis aeruginosa* (Kutzing) with its spherical cells falling within the range of 4.0-6.5 μm in diameter. Also as expected, many gas vacuoles were present making the cell appear darker and there was a thick mucilaginous layer surrounding individual cells as well as the colonies. These features indicated that morphologically this isolate appeared to be identified correctly. *M. aeruginosa* is not known to produce geosmin (Watson, 2003) and as expected no geosmin production was observed by the strain used in this study.

The 16S rRNA gene sequence BLASTn results returned five matches to *M. aeruginosa* with sequence similarities ranging from 99.64-99.11% indicating that the sequence was very likely obtained from the *M. aeruginosa* genome. Consistencies of morphological characteristics as well as 16S rRNA sequence identities to *M. aeruginosa* in GenBank provide strong support for an accurate identification of this isolate.

Microcystis wesenbergii

This isolate conformed to the type description of *Microcystis wesenbergii* (Komárek) with its spherical cells falling within the range of 4.0-8.5 μm in diameter. Also as expected, fewer gas vacuoles were present in comparison to *M. aeruginosa* that made the cell appear lighter in colour. In addition, a thinner mucilaginous layer than that observed in the *M. aeruginosa* isolate surrounded individual cells as well as the colonies and conforming to the type description of *M. wesenbergii*. These features indicated that this isolate was identified correctly. *M. wesenbergii* has not been known to produce geosmin (Watson, 2003) and as expected the isolate used in this study did not produce the odourous compound.

The 16S rRNA gene sequence BLASTn results returned two matches to *M. wesenbergii* with both sequence similarities calculated as 100%. Consistencies of morphological characteristics as well as 16S rRNA sequence identities to *M. wesenbergii* in GenBank provided strong support for an accurate identification of this isolate.

Calothrix parietina **PCC 6303**

The vegetative cell sizes of this isolate were within the expected range of 10.0-14.0 μm long and 5.0-10.0 μm wide and congruent with the type specimen *Calothrix parietina* (Thuret) Bornet et Flahault. A single heterocyst with a conical shape was observed at the basal end of the filament succeeded by one or more larger vegetative cells with a black colour, giving this culture a very dark appearance. In addition, there were no observed akinetes and the filaments were long, tapered at the apex and surrounded by a sheath, all of which are characteristic of *Calothrix parietina*. This particular strain has not been

known to produce geosmin in culture (Hockelmann *et al.*, 2004) and as expected the isolate in this study did not produce the odorous compound.

The 16S rRNA gene sequence BLASTn resulted in three matches to unidentified species of *Calothrix* with sequence identities ranging from 100.00-92.88% and two matches to *Calothrix desertica* with sequence identities of 93.51% and 92.47%. Only two *Calothrix parietina* 16S rRNA partial sequences are available in GenBank and they were not in the top 100 returned BLAST matches however in a pairwise sequence alignment using BioEdit© a sequence identity of 93.00% was calculated. Since the range of sequence similarity is quite large and only two clonal *Calothrix parietina* sequences exist in GenBank, there was not enough data to reject the identification of this isolate. Congruency with morphological characteristics of the type strain *Calothrix parietina* and a genus match of the 16S rRNA gene sequence, suggest that the isolate of *Calothrix parietina* used in this study was correctly identified.

Calothrix **sp. PCC 7507**

This isolate was not previously morphologically identified to the species level however it did have characteristics consistent with the genus *Calothrix* sp. Aghard ex Bornet et Flahault. The average vegetative cell size was much smaller in length than *Calothrix parietina* by approximately 5 μm and the cells were much lighter in colour. The filament did not have large, blackish cells succeeding the basal heterocyst as in *C. parietina* and the filament did not taper as much. Two types of heterocysts were observed; one conical shaped at the basal end, and an occasional intercalary heterocyst much longer and rectangular than the basal type. Both types are not uncommon in species of *Calothrix* and are characteristic of this isolate. The filaments were very long, somewhat tapered,

intermingled, and with a surrounding sheath, all of which are consistent with members of this genus. This particular strain is a known and potent geosmin producer (Hockelmann *et al.*, 2004) and as expected in this study it produced a large amount of the compound.

There were approximately 20 16S rRNA gene sequences in GenBank, however only one belonged to *Calothrix brevissima.* The GenBank sequence from this species matched with the isolate sequenced in this study with a similarity of 95.94%. The remaining sequence matches belonged to several *Cylindrospermum, Trichormus*, and *Nostoc* species but none had a sequence identity with more than 97.22% indicating that a sequence from this unidentified species of *Calothrix* was simply not in the database and cannot be accurately identified based on its sequence data. Morphological characteristics of this isolate were consistent with *Calothrix brevissima* from the returned BLAST results and with this data it can be concluded that the isolate used in this study was of the genus *Calothrix*, however no speculations were made as to the species it belongs.

Oscillatoria limosa

This isolate conformed to the description given for the type specimen of *Oscillatoria limosa* (C. Agardh) Gomont with its vegetative cell sizes within the expected range of 2.0-5.0 μm long and 11.0-22.0 μm wide, much larger than *Pl. aghardii*. The cells were much wider than they are long, did not constrict at the cross walls, and did not produce heterocysts or akinetes. Filaments were very long, single, slightly attenuated towards apex with the end cell rounded and with a slightly thicker membrane. Filaments were also solitary, free-floating, and with a thin mucilaginous sheath, all of which were consistent with the expected morphology. Several strains of *Osc. limosa* are geosmin-producing and are often implicated in drinking water taste and odour episodes (Guasch

et al., 2005) and the strain of *Osc. limosa* used in this study continually produced a large amount of geosmin, not a surprising characteristic of this species.

The 16S rRNA gene sequence BLASTn results listed two matches to unidentified species of *Oscillatoria* with sequence identities of 98.94 and 98.18%. Only one *Osc. limosa* 16S rRNA gene sequence was available in GenBank and it did not appear in the BLASTn results. A pairwise sequence analysis using BioEdit© calculated a sequence identity of 89.70% with the GenBank *Osc. limosa* 16S rRNA sequence suggesting that either this isolate is not *Osc. limosa* or that the sequence in GenBank was misidentified. Since a strong sequence identity exists between the isolate sequence and other *Oscillatoria* sequences and only one *Osc. limosa* sequence is available in GenBank, there was not sufficient data to conclude that the 16S rRNA gene obtained from this isolate was not *Osc. limosa*. In addition, consistencies of morphological characteristics provide even stronger support that this isolate has been correctly identified for use in this study.

2.4.2   *PCR amplification in the negative control was primer stacking*

During the second round of PCR amplification, it was observed that the negative control had some unexpected amplification when the blank from the first round was used as template for the second round of amplification. A new negative control using PCR grade sterile water as template for the second round of amplification was attempted to determine whether amplification in the previous negative control was due to primer stacking or actual contamination from a reagent. When this was done, no amplification was observed in the sterile water reaction, indicating it was likely not reagent contamination but primer stacking due to multiple primers used in the amplification. To

confirm with certainty that this was not true contamination, DNA from the negative control was sequenced and a BLASTn of the sequence against those present in GenBank did not result in identical sequence matches to any organisms. None of the BLAST results included the full gene fragment sequence and its alignment was fragmented across the regions of similarity. When the negative control sequence was aligned with the full 16S rRNA gene of all the isolates it was also very staggered across the region indicating that it was likely primer stacking occurring in the sample and not true contamination. When electrophoresed in DGGE this band migrated to a very high denaturing concentration of approximately >50%, away from the range of the DGGE standard bands, and were not considered true bands in future analyses.

2.4.3 SEPARATION OF GEOSMIN-PRODUCING AND NON-PRODUCING ISOLATES USING 16S RRNA-V3 DGGE

The 16S rRNA-V3 gene region using DGGE was expected to separate the 23 cyanobacteria isolates based on slight differences in nucleotide sequences. Morphological and genetic characterization identified 17 different species of the 23 isolates examined. Seventeen unique bands were expected subsequent to separation of the 23 isolates using DGGE, each representative of a single species however only ten unique positions were observed suggesting that a single band was representative of more than one species. Six groups of isolates migrated to the same gel position (Figure 9 and 10) with two species representing a single band for groups 1, 3, 4, 5, and 6 and four species representing the group 2 band position. Since DGGE can separate DNA fragments differing in as little as one nucleotide (Wang *et al.*, 2005), it was expected that the co-migrating fragments also had exact sequences. Sequencing of each of the ten

fragments verified that each fragment at a position in the gel was identical in DNA sequence. Sequencing results of the fragments were as expected with ten different positions representing ten different V3 sequences. Although each sequence within the groups was identical to one another, two of those groups included isolates of different species and the other three groups included those of different genera. Group 1 included the two *An. flos-aquae* strains with identical V3 sequences as expected of the same species. Although there is no cutoff value of 16S rRNA sequence similarity for the species definition, it is apparent from the vast amount of studies that the majority of recognized species examined to date differ in their 16S rRNA sequence from related species of the same genus by at least 1% of the sites, and typically more (Song *et al.*, 2003). Sequencing of the final (approximately 1000) nucleotides from *An. flos-aquae* UTCC 64 was not possible even after several attempts with numerous alterations in the PCR protocol and was likely due to the development of secondary structure in the molecule. Pairwise sequence alignment of the first 399 nucleotides in the 16S rRNA gene was possible for both isolates in group 1 and a 98.70% sequence similarity was calculated. This is very close to a 1% difference, validating that the isolates are of the same species and as expected migrated to the same position in the gel.

Group 2 included the most isolates for a single grouping with seven members of identical sequences in the V3 region. All of the *An. lemmermannii* isolates (geosmin-producing and non-producing) were a part of this group indicating that the V3 region was identical for this species and it does not distinguish the geosmin-producers from the non-producers. The fact that geosmin-producing strains could not be separated in the 16S rRNA-V3 region suggests that geosmin production may not be nuclear encoded and

the genes for this trait may be located on a plasmid. Strains that possess the plasmid would thus not separate from those that do not produce the compound because the 16S rRNA gene used for discrimination is derived from the nuclear genome. Since *An. planktonica* and *An. viguieri* were included at this gel position it was thought that they might have been of the same species however sequence similarities of 1203 nucleotides of the 16S rRNA gene for both isolates ranged from 97.10-94.70% to others in this group, suggesting that they are two very distinct species. The band for this group was included in the DGGE standard and was representative of three *Anabaena* species as well as geosmin-producing and non-producing *An. lemmermannii* species. The application of this particular band in the identification of potential geosmin-producing *An. lemmermannii* in Lake Ontario is discussed in chapter 3.

Group 3 isolates included *An. cylindrica* (geosmin non-producing) and *Calothrix* sp. PCC 7507 (geosmin-producing), both of the Nostocales order yet belong to the Nostocaceae and Rivulariaceae families respectively. Since these isolates are very different morphologically and genetically it was not expected that they would possess identical V3 sequences. A pairwise alignment of 1203 nucleotides of the 16S rRNA gene calculated a 95.00% sequence similarity, clearly indicative of two different lineages. The band for this group was included in the DGGE standard and was representative of a potential geosmin-producing and a geosmin non-producing species within the Nostocales.

Groups 4 and 5 included planktonic isolates that were not as distantly related as those of group 3. *Ap. gracile* (geosmin-producing) and *Anabaena* sp. LOW (geosmin non-producing) of group 4 and *Ap. flos-aquae* Hamilton Harbour and *An. spiroides* (both

geosmin non-producers) of group 5 were all of the Nostocaceae family but differed in genera. Sequencing of the mid ~200 nucleotides of the 16S rRNA gene was not successful for *Anabaena* sp. LOW even after several attempts with different primers, amplification conditions, and extractions, therefore the pairwise alignment of the isolates in group 4 involved 855 nucleotides. Percent similarities of *Ap. gracile* and *Anabaena* sp. LOW were very high at 99.4% and not expected for sequences of different genera. Gugger et al. (2002) examined sequence data of the 16S rRNA, ITS1, and rbcL regions and revealed that planktonic species of *Anabaena* and *Aphanizomenon* form a single clade suggesting they should be considered members of the same genus regardless of their morphologies (Gugger *et al.*, 2002). It is therefore probable that *Ap. gracile* and *An*abaena sp. LOW are more closely related than previously thought however morphologically they are very distinct and likely not the same species. Analysis of the pairwise alignment of 1203 nucleotides in the 16S rRNA genes of group 5 isolates resulted in a sequence similarity of 97.5% and suggested the two isolates were not the same species. Both groups 4 and 5 were included in the DGGE standard with the group 4 band position representing geosmin-producing and non-producing genera and the group 5 band position representing non-geosmin producers all within the Nostocaceae.

     M. aeruginosa and *M. wesenbergii* of the Chroococcales order and Subsection I, comprised the group 6 isolates with identical V3 regions. The 180 nucleotides in the center of the 16S rRNA gene of *M. wesenbergii* could not be sequenced even after attempts with various primers, PCR programs, and extractions therefore 921 nucleotides were aligned pairwise and were calculated to have 98.9% similarity. This number is slightly higher than expected for members of different species, however if the missing

nucelotides were included in the analysis there may have been an increase in sequence differences. With the support of morphological variation among the two species as well as some sequence differences, *M. aeruginosa* and *M. wesenbergii* should not be considered the same species. This group was included in the DGGE standard because its band positioned in close proximity to the group 3 band and may have caused difficulties in band matching. Additionally, these were not geosmin-producing strains and were not of primary interest in this study.

As expected some isolates migrated to unique positions on the gel including *C. parietina* PCC 6303, *Pseudanabaena* sp. UTCC 593, *Pl. aghardii*, and *Osc.limosa.* Of these isolates, *C. parietina* PCC 6303 and *Osc. limosa* were included in the DGGE standard because they did not display multiple amplicons and their positions were distinctly separate from other isolate fragments. *Pseudanabaena* sp. UTCC 593 was not included in the DGGE standard marker as it did not produce geosmin. In addition, its band was located very close to those of the group 3 band position which may have caused difficulties in band matching to Lake Ontario water samples (chapter 3). It is not uncommon for a strain of cyanobacteria to possess more than one 16S rRNA gene, a condition that has led to multiple amplicons in some isolates subsequent to PCR amplification (Crosbie *et al.*, 2003). *Anabaena* sp. LOW, *An. spiroides*, *Ap. gracile*, and *Pl. aghardii* exhibited four amplicons that were either heteroduplex artifacts or simply multiple 16S rRNA genes in the genome. For sequence analysis, the most intense bands of these samples were used since they were consistently the brightest, suggesting that the band at that gel position was the most common and representative of the strain. Since *Anabaena* sp. LOW and *Ap. gracile* share a common V3 sequence and band, *An*abaena

sp. was chosen to be included in the standard because more than one amplicon was rarely observed. Amplification products of the three remaining multiband isolates (*Calothrix* sp. PCC 7507, *Ap. gracile*, and *Pl. aghardii*) were not used in the DGGE standard because more than one band would be representative of the three isolates. Instead, one isolate from each group with a single amplicon was used to represent the isolates at that particular position in the gel.

2.5   CONCLUSIONS

It was apparent from the 16S rRNA-V3 PCR-DGGE standard that the V3 region could not resolve to the species level among the 17 different species examined in this study and did not distinguish geosmin-producing and non-producing strains of *An. lemmermanni.* It thus should not be used alone to infer with certainty that a matching band in an environmental sample is or is not a geosmin producer. With the support of other data such as geosmin concentration in the environmental sample, correlations may exist that support the presence of a DGGE band from a potential geosmin-producing species, thus strengthening support for its presence in the environment at the time of collection. Although the standard was not capable of identifying a single species for each band, it is likely that the use of DGGE in conjunction with other data will enable a more accurate identification of potential geosmin-producing cyanobacterial species in water samples. This approach was assessed in chapter 3 using Lake Ontario water samples and geosmin concentrations. The seven other hypervariable regions of the 16S rRNA gene are less variable and would not be as useful individually compared to the V3 region for identification of species. A combination of the V3 and other hypervariable regions such as the V5 however may provide increased variation among the cyanobacteria isolates and may be a more resolved genetic marker (Yu and Morrison, 2004).

The ITS-1 region is a highly evolving untranslated region that was briefly examined in this study to assess whether it may be species-specific and resolve to the species level more effectively than the 16S rRNA-V3 region. Sixteen of the 23 isolates were sequenced and demonstrated too much variation even within the same species. The ITS-1 region in cyanobacteria has been observed to possess $tRNA_{ile}$ and/or $tRNA_{ala}$ each

approximately 100 nucleotides in length. Results of this study showed multiple

amplicons in the majority of the isolates. This region would not be useful for identifying

individual species from environmental samples due to the high variability and multiple

amplicons among strains. In addition, amplicons with identical sequences were observed

in geosmin-producing and non-producing *An. lemmermannii* suggesting that this region

would also not be capable of distinguishing these strains from one another.

Other regions such as the RNA polymerase C1 (*rpo*C1), useful for supporting

16S rRNA phylogenies (Seo and Yokota, 2003), the subunit B protein of DNA gyrase

(*gyr*B), a faster evolving gene than 16S but more conserved than ITS (Seo and Yokota,

2003), and Rubisco (*rbc*LX) that has been successful in identifying relationships among

toxin-producing cyanobacteria (Gugger *et al.*, 2002) should be investigated for use in

identifying geosmin-producing strains in DGGE. It is also possible that genes involved

in the MEP or the hypothetical phosphate pathways observed in *Synechocystis* sp. may

also be present in other species of cyanobacteria. Analysis of one or more of these gene

sequences may be capable of distinguishing geosmin-producing from non geosmin-

producing strains from one another. Investigations of new genes may separate individual

species to unique positions on the gel or, ideally, distinguish geosmin-producing and

non-producing strains of cyanobacteria from one another and allow for accurate

molecular identifications in DGGE without requiring additional data. Furthermore, the

incomplete separation of geosmin-producing species from non-producing species in this

study suggests the possibility that the genes involved in geosmin production could be

encoded on a plasmid rather than in the nuclear genome. Complementation studies with

geosmin-producing and non-producing strains should be investigated to determine if this

trait is acquired through plasmid transfer. With further research into the biosynthetic pathway for geosmin, it is hopeful that a single gene responsible for the final product will be identified and allow for a straightforward identification of geosmin-producing strains.

2.6   REFERENCES

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). Basic local alignment search tool. Journal of Molecular Biology 215: 403-410.

Blaha, L., S. Sabater, P. Babica, E. Vilalta and B. Marsalek (2004). Geosmin occurrence in riverine cyanobacterial mats: is it causing a significant health hazard? Water Science and Technology 49(9): 307-312.

Boutte, C., S. Grubisic, P. Balthasart and A. Wilmotte (2006). Testing of Primers for the study of Cyanobacterial molecular diversity. Journal of Microbiological Methods 65: 542-550.

Boyer, S. L., V. R. Flechtner and J. R. Johansen (2001). Is the 16S-23S rRNA internal transcribed spacer region a good tool for use in molecular systematics and population genetics? A case study in Cyanobacteria. Molecular Biology and Evolution 18(6): 1057-1069.

Casamatta, D. A., J. R. Johansen, M. L. Vis and S. T. Broadwater (2005). Molecular and morphological characterization of ten polar and near-polar strains within the Oscillatoriales (Cyanobacteria). Journal of Phycology 41: 421-438.

Casamatta, D. A., M. L. Vis and R. G. Sheath (2003). Cryptic species in cyanobacterial systematics: a case study of *Phormidium retzii* (Oscillatoriales) using RAPD molecular markers and 16S rDNA sequence data. Aquatic Botany 77: 295-309.

Castenholz, R. W. (1992). Species Usage, Concept, and Evolution in the Cyanobacteria (Blue-Green Algae). Journal of Phycology 28: 737-745.

Crosbie, N. D., M. Pockl and T. Weisse (2003). Rapid establishment of clonal isolates of freshwater autotrophic picoplankton by single-cell and single-colony sorting. Journal of Microbiological Methods 55: 361-370.

Cybernetics, M. (1999). Image-Pro Express. Silver Spring, MD, Media Cybernetics.

Guasch, H., E. Vilalta, I. Munoz, A. Romani and S. Sabater (2005). Causes of geosmin production in lotic systems. The case of the Llobregat river. ASLO Santiago 2005 Summer Meeting, Santiago de Compostela, Spain.

Gugger, M., C. Lyra, P. Henriksen, A. Coute, J.-F. Humbert and K. Sivonen (2002). Phylogenetic comparison of the cyanobacterial genera *Anabaena* and *Aphanizomenon*. International Journal of Systematic and Evolutionary Microbiology 52: 1867-1880.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95-98.

Hockelmann, C., T. Moens and F. Juttner (2004). Odor compounds from cyanobacterial biofilms acting as attractants and repellents for free-living nematodes. Limnology and Oceanography 49(5): 1809-1819.

Hu, Q., M. Sommerfield, L. Baker and P. Westerhoff (2003). Canal wall brushing - a control measure for taste and odour problems in drinking water supplies and in environments. Journal of Water Supply Research and Technology-Aqua 52(8): 545-554.

Huelsenbeck, J. P., F. Ronquist and P. v. d. Mark (2005). Mr. Bayes 3.1. Computer program distributed by the authors.

Janse, I., W. Edwin, A. Kardinaal, M. Meima, J. Fastner, P. M. Visser and G. Zwart (2004). Toxic and non-toxic Microcystis colonies in natural populations can be differentiated on the basis of rRNA gene internal transcribed spacer diversity. Applied and Environmental Microbiology 70(7): 3979-3987.

Janse, I., M. Meima, W. E. A. Kardinaal and G. Zwart (2003). High-Resolution Differentiation of Cyanobacteria by Using rRNA-Internal Transcribed Spacer Denaturing Gradient Gel Electrophoresis. Applied and Environmental Microbiology 69(11): 6634-6643.

John, D. M., B. A. Whitton and A. J. Brook, Eds. (2002). The Freshwater Algal Flora of the British Isles. Cambridge, The press syndicate of the University of Cambridge.

Jüttner, F., J. Leonhardt and S. Moehren (1983). Environmental Factors Affecting the Formation of Mesityl Oxide Di Methyl Allylic Alcohol and Other Volatile Compounds Excreted By Anabaena-Cylindrica. Journal of General Microbiology 129: 407-412.

Kan, J. J., K. Wang and F. Chen (2006). Temporal variation and detection limit of an estuarine bacterioplankton community analyzed by denaturing gradient gel electrophoresis (DGGE). Aquatic Microbial Ecology 42: 7-18.

Kolmonen, E., K. Sivonen, J. Rapala and K. Haukka (2004). Diversity of Cyanobacteria and heterotrophic bacteria in cyanobacterial blooms in Lake Joutikas, Finland. Aquatic Microbial Ecology 36: 201-211.

Komárek, J. and K. Anagnostidis (1989). Modern Approach to the Classification System of Cyanophytes 4. Nostocales. Archiv fur Hydrobiologie Supplement 82: 247-345.

Lewis, P. O. (2001). Phylogenetic systematics turns over a new leaf. Trends in Ecology and Evolution 16: 30-37.

Litvaitis, M. K. (2002). A molecular test of cyanobacterial phylogeny: inferences from constraint analyses. Hydrobiologa 468: 135-145.

Muyzer, G. (1999). DGGE/TGGE a method for identifying genes from natural ecosystems. Current Opinion in Microbiology 2: 317-322.

Muyzer, G., E. C. DeWaal and A. Utterlinden (1993). Profiling complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Applied and Environmental Microbiology 59(3): 695-700.

Nakamura, S., N. Nakamura and S. Ito (2001). Determination of 2-methylisoborneol and geosmin in water by gas chromatography-mass spectrometry using stir bar sorptive extraction. Journal of Separation Science 28(8): 674-677.

Nubel, U., F. Garcia-Pichel and G. Muyzer (1997). PCR primers to amplify 16S rDNA from Cyanobacteria. Applied and Environmental Microbiology 63(8): 3327-3332.

Oren, A. (2004). A proposal for further integration of the Cyanobacteria under the Bacteriological Code. International Journal of Systematic and Evolutionary Microbiology 54: 1895-1902.

Palinska, K. A., W. Liesack, E. Rhiel and W. E. Krumbien (1996). Phenotype variability of identical genotypes: the need for a combined approach in cyanobacterial taxonomy demonstrated on *Merismopedia*-like isolates. Archives of Microbiology 166: 224-233.

Posada, D. and K. A. Crandall (1998). Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817-818.

Rajaniemi, P., P. Hrouzek, K. Kastovska, R. Williame, A. Rantala, L. Hoffman, J. Komarek and K. Sivonen (2005). Phylogenetic and morphological evaluation of the genera *Anabaena*, *Aphanizomenon*, *Trichormus* and *Nostoc* (Nostocales, Cyanobacteria). International Journal of Systematic and Evolutionary Microbiology 55: 11-26.

Rashash, D., A. Deitrich, R. Hoehn and B. Parker (1995). The influence of growth conditions on odour-compound production by two chrysophytes and two cyanobacteria. Water Science and Technology 31: 165-172.

Seo, P.-S. and A. Yokota (2003). The phylogenetic relationships of cyanobacteria inferred from 16S rRNA, *gyr*B, *rpo*C1, and *rpo*D1. Journal of General Applied Microbiology 49: 191-203.

Song, Y., C. Liu, M. McTeague and S. M. Finegold (2003). 16S Ribosomal DNA Sequence-Based Analysis of Clinically Significant Gram-Positive Anaerobic Cocci. Journal of Clinical Microbiology 41(4): 1363-1369.

Stein, J. E. (1973). Handbook of Phycological Methods: Culture Methods and Growth Measurements, Cambridge University Press.

Svenning, M. M., T. Eriksson and U. Rasmussen (2005). Phylogeny of symbiotic cyanobacteria within the genus *Nostoc* based on 16S rDNA sequence analyses. Archives of Microbiology 183: 19-26.

Swofford, D. L. (2001). PAUP* Phylogenetic analysis using parsimony and other methods, Sinauer.

Thompson, J. R., L. A. Marcelino and M. F. Polz (2002). Heteroduplexes in mixed-template amplifications: formation, consequences, and elimination by reconditioning PCR. Nucleic Acids Research 30(9): 2083-2088.

Wang, Q., J. R. Deeds, A. R. Place and R. Belas (2005). Dinoflagellate community analysis of a fish kill using denaturing gradient gel electrophoresis. Harmful Algae 4: 151-162.

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. (2004). Aquatic taste and odour: a primary signal of drinking water integrity. Journal of Toxicological and Environmental Health, Part A 67: 1779-1795.

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

Wehr, J. D. and R. G. Sheath, Eds. (2003). Freshwater Algae of North America: ecology and classification. San Diego, Academic Press.

Willame, R., C. Boutte, S. Grubisic, J. Komarek, L. Hoffmann and A. Wilmotte (2005 submitted). Morphological and molecular characterization of planktonic cyanobacteria from Belgium and Luxembourg. Environment and Biotechnologies Research Submitted paper, unpublished.

Yu, Z. and M. Morrison (2004). Comparisons of different hypervariable regions of rrs genes for use in PCR-denaturing gradient gel electrophoresis. Applied and Environmental Microbiology 70(8): 4800-4806.

Zwart, G., M. P. K.-v. Agterveld, I. v. Werff-Staverman, F. Hagen, H. L. Hoogveld and H. J. Gons (2005). Molecular characterization of cyanobacterial diversity in a shallow eutrophic lake. Environmental Microbiology 7(3): 365-377.

CHAPTER 3

IDENTIFICATION OF POTENTIAL GEOSMIN-PRODUCING *ANABAENA LEMMERMANNII* AND OTHER CYANOBACTERIA IN LAKE ONTARIO USING DGGE

3.1 INTRODUCTION

One of the most frequently reported odours in western Lake Ontario is an earthy, muddy, musty smell commonly caused by geosmin (*trans*-1, 10-dimethyl-*trans*-9-decalol), a terpenoid metabolite released by certain members of the Cyanobacteria, bacteria and some fungi (Skafel and Yerubandi, 2003). Because taste and odour characteristics of drinking water are paramount in the public perception of its quality, it is essential that these compounds be reduced to undetectable levels before reaching the tap. The public can become accustomed to low-level background taste and odour from minerals, chlorine residue, or seasonal patterns, however abrupt changes in these characteristics often elicits consumer concern. Water quality is largely assessed using taste and odour as a primary indicator and as a result this can have many negative effects on consumption (Watson, 2004). Although it has no effect on health, consumers often associate the scent of geosmin with poor water quality and will avoid its consumption.

The suspected primary contributor to Lake Ontario taste and odour is the cyanobacterium *Anabaena lemmermannii* (Watson, 2004). Peak geosmin concentrations occur annually in late August and early September when the lake is stratified and there is a distinct thermocline in the upper 30 m. With a warm upper layer, increased abundances of *An. lemmermannii* are common and when they senesce, their cells breakdown and intracellular geosmin is released into the water, thereby increasing the concentration of

the compound significantly (Watson, 2004). Studies suggest that downwelling induced by strong easterly winds drives water towards the northwest shoreline and deeper into the mixed layer. As a result, offshore geosmin is transported nearshore at lower depths and into water treatment facility intakes (Skafel and Yerubandi, 2003). When sufficient geosmin concentrations reach intakes, it often evades conventional drinking water treatment resulting in undesirable taste and odour in tap water. Conventional treatment such as sand sedimentation and sand filtration are largely ineffective at removing most dissolved algal volatile organic compounds (AVOCs), additionally, carbon beds and advanced filtration technology performance has been substantially reduced in the presence of high concentrations of dissolved organic material (Watson, 2004). Past taste and odour episodes in Lake Ontario have been severe, reaching levels more than two to three times higher than the odour threshold concentration ($\sim$4-10 ng $\cdot$ l^{-1}. During these times (summers of 1987, 1994, 1998, 1999) water treatment facilities were unable to prepare for high geosmin concentrations due to the unpredictable nature of the events.

Several studies suggest that geosmin production by cyanobacteria is not directly correlated with specific environmental factors. Iron (Nakashima and Yagi, 1992), temperature, nitrogen, phosphorus, light (Saadoun *et al.*, 2001) and copper (Dionigi, 1995) have shown varying relationships with geosmin production in some species under laboratory conditions. It appears that for cyanobacteria, increased light intensity and non-optimal temperatures favours geosmin production by *Anabaena* sp. (Jüttner, 1995; Saadoun *et al.*, 2001) Limited iron conditions tend to increase geosmin release by some cyanobacteria into growth media (Jüttner, 1995) and nitrogen in the form of dinitrogen (N_2) rather than nitrate (NO_3) tends to increase geosmin production. Increased copper

has resulted in reduced geosmin levels in several genera of Cyanobacteria and bacteria due to its inhibitory effect on overall growth (Dionigi, 1995; Saadoun *et al.*, 2001). These experiments that have tested various environmental conditions against geosmin production have shown that geosmin has neither a constant ratio to protein nor to chlorophyll-a content suggesting that it does not directly correlate to cell biomass. The range of variation of this metabolite was limited and has not exceeded a factor of two to three, indicating that there may be some limitations to its production by the cell (Jüttner, 1995).

Geosmin is a sesquiterpene metabolite derived from one or more isoprenoid biosynthetic pathways: the mevalonate (MVA), the methylerythritol pathway (MEP) (Lange *et al.*, 2000), and possibly the hypothetical pentose phosphate pathway (Ershov *et al.*, 2002). The exact route to geosmin has not been completely elucidated subsequent to the production of isopentenyl diphosphate (IPP), the central intermediate of isoprenoid biosynthesis (Lange *et al.*, 2000). Because the conditions that may promote translation of enzymes involved in the biosynthesis of geosmin are unknown and geosmin production does not consistently correlate to a specific environmental variable, it has been very difficult to predict when or why taste and odour episodes occur.

Water quality managers have generally been unable to prevent or control taste and odour episodes of geosmin-producing cyanobacteria via proactive action prior to episodes due to a lack of knowledge of the cyanobacterial communities in the waters as well as the inability to predict when geosmin will reach intake pipes as a result of physical processes. A proactive approach to controlling taste and odour in drinking water sources by monitoring cyanobacterial species composition in water samples for potential

taste and odour producers is an option that generally has not been employed by water managers. Characterization of species composition in a water body could provide an advanced warning of the presence of potential taste and odour-causing species in source waters. This would enable water quality managers to prepare for taste and odour events and water treatment could in turn be optimized. In the long term, identification of cyanobacterial taxa that occur in source water could contribute towards a more robust management of drinking water through corrective or preventative measures (Watson, 2004).

Identification of potential geosmin-producing species of cyanobacteria in water samples can be performed using traditional microscopy, biochemical means such as marker pigments, or via molecular biology techniques. The first method is very useful, however extremely time-consuming, costly and requires expert opinions for correct identification to species levels. It also cannot positively identify odour producers, since different strains of the same (morphologically defined) species can vary considerably in their ability to produce geosmin (as with toxin production). The second approach only provides identification of the presence of cyanobacteria, with no resolution among taxa. The last approach can use a variety of techniques to examine cyanobacterial communities present in an environmental sample based on DNA or protein data.

The 16S rRNA gene is the basis for systematic assignment in the latest edition of Bergey's Manual of Systematic Bacteriology and has been useful in distinguishing broad taxonomic groups as well as individual species and is frequently used in the classification and identification of cyanobacteria (Casamatta *et al.*, 2003; Litvaitis, 2002; Rajaniemi *et al.*, 2005; Svenning *et al.*, 2005). Direct sequence analysis of this region

has been fundamental in phylogenetic studies of the Cyanobacteria and has resulted in a

large expansion of the molecular data that is available in public databases such as

GenBank (Svenning *et al.*, 2005). Although the analysis of nucleotide sequences

provides the most characteristics for distinguishing taxonomic groups, it nevertheless is

a lengthy process. To characterize species of cyanobacteria in environmental samples

using sequence analysis of the full 16S rRNA gene, time-consuming procedures such as

cloning or isolation and culturing of each strain would be necessary. Although a

substantial amount of data can be gained through sequence analysis of genes such as the

16S rRNA, this technique is neither rapid nor simple and would not be ideal for routine

use in the identification of geosmin-producing cyanobacteria to facilitate the prediction

or monitoring of taste and odour episodes.

Denaturing gradient gel electrophoresis (DGGE) is a molecular technique that

has been used to assess the genotypic diversity in environmental samples as well as to

evaluate the purity of cultures without the need for strain isolation or lengthy DNA

sequence analysis. This technique works through sequence-dependant separation of PCR

products and is based on the decreased electrophoretic mobility of partially melted

double stranded DNA molecules in a polyacrylamide gel containing a linear gradient of

denaturants in the form of formaldehyde and urea (Muyzer, 1999). When a double

stranded DNA fragment migrates by electrophoresis through a denaturing gradient, it

partially melts and changes form. Nucleotide "domains" that suddenly dissociate at the

same time, dramatically slow down migration, resulting in denaturation of the DNA

fragment (Wang *et al.*, 2005). Slight changes in base pair (bp) composition, often as

little as 1 bp, will shift domain boundaries, thereby altering the conditions needed for

domain dissociation. Accordingly, the denaturing concentrations required for domain dissociation of all of the DNA fragments in an environmental sample will vary. Each species within an environmental sample will correlate to a DNA fragment and migration of the fragments to different positions in the gel, will result in a DGGE profile (Kolmonen *et al.*, 2004). DGGE profiles from environmental samples can give insight into the diversity of certain groups of organisms and has been used in a number of cyanobacterial studies to assess community structure (Zwart *et al.*, 2005), monitor diversity over time (Kolmonen *et al.*, 2004), and identify characteristic strains of cyanobacteria such as toxin-producing *Microcystis* (Janse *et al.*, 2003). This technique has proven to be efficient, cost-effective, and relatively simple in studies of species composition in environmental samples and is applied for the first time in this study for the identification of potential geosmin-producing cyanobacteria in Lake Ontario.

The DGGE standard marker developed using the characterized isolates described in chapter 2 has been used in this study to identify potential geosmin-producing species of cyanobacteria present in Lake Ontario water samples. Since *An. lemmermannii* is the suspected primary contributor to western Lake Ontario taste and odour, it was hypothesized that this species would be present in samples from this region that contained geosmin concentrations greater than the human threshold odour concentration. Sampling sites with increased geosmin and no *An. lemmermannii* were also examined for other potential geosmin-producing species of cyanobacteria using the developed marker. The aim of this study was to address the following questions:

Is DGGE an effective in the identification of potential geosmin-producing cyanobacteria in Lake Ontario?

Does the presence of *An. lemmermannii* and increased geosmin concentrations correlate to one another and support geosmin and cyanobacteria community trends throughout the lake?

Would DGGE be a useful technique for use in monitoring potential geosmin-producing cyanobacteria in drinking water management?

3.2 MATERIALS AND METHODS

3.2.1 SAMPLING AND GEOSMIN QUANTIFICATION

Water samples from 32 inshore and offshore locations surrounding Lake Ontario were collected from less than 0.5 m depths during August 29 – September 2, 2005 on the CCGS Limnos (Fig. 17). Sample volumes of 1 l of sample were filtered through a 47 mm diameter, 1.5 μm nominal pore sized glass fibre filter (GF/C) (Whatman, USA) and stored in small petri dishes at -20ºC. Within 24 h of collection the samples were stored at –80ºC until DNA was isolated.

For geosmin analysis using GC-MS, 1 l of water was drawn off into and stored in 0.5-1.0 l narrow-necked glass amber bottles to minimize photosynthetic activity or photo-oxidation. A volume of 50 mL of this was replaced with 3-(3,4-dichlorophenyl)-1,1-dimethylurea DCMU (to inhibit photosynthesis) and recovery standard was added by syringe (250 μl of 200 ng · ml^{-1} napthalene-d$_8$ + biphenyl-d$_{10}$ in methanol, giving a final concentration of 25 ng · l^{-1}. Quantitative analysis of geosmin was performed using a HP Agilent gas chromatograph and mass spectrometer.

3.2.2 DNA ISOLATION AND PCR AMPLIFICATION OF THE 16S RRNA-V3
REGION

DNA was extracted using the chelex-100 protocol (Walsh *et al.*, 1991). To evaluate the reproducibility of the sampling and DNA isolation method, seven triplicate DNA isolations were performed for randomly chosen filter samples that included stations 738, 86, 55, 71, 5, 82, and 1001. Each filter was cut into 1/16th using sterile scissors and placed in a 1.5 ml microfuge tube to which 400 μl 5% Chelex-100 (BioRad, Canada)

solution was added. The tubes were incubated at 100ºC for 30 min with shaking approximately every 5 min. After incubation each tube was vortexed on high for 10 s and subsequently centrifuged at 13 000 rpm for 2 min. The supernatant was transferred to a newly labeled tube and used as template for PCR.

The 16S rRNA-V3 region (~161 base pairs) of each sample was amplified in a nested PCR protocol using cyanobacterial specific primers (Zwart *et al.*, 2005) as in chapter 3 using 8 µl of template DNA in the first round and 6 µl in the second round. The 25 µl reaction product was used in DGGE for separation and analysis of all amplicons.

3.2.3 DGGE

Optimal acrylamide percentage and denaturant concentrations for parallel DGGE were determined as in chapter 2. After optimization experiments and development of a DGGE marker for use in the identification of potential geosmin-producing species in the Lake Ontario samples (chapter 2), perpendicular DGGE was performed using the D-Code Universal Mutation Detection System (BioRad, Canada) 25 µl of PCR product from the V3 region from each of the 32 samples was electrophoresed in an 8% polyacrylamide gel containing a gradient of denaturant ranging from 35-55% (100% denaturant is 7 M urea and 40% deionized formamide). Electrophoresis was performed using the D-Code Universal Mutation Detection System (BioRad, Canada) in a buffer containing 0.04 M Tris-acetate and 0.001 M EDTA (pH 7.6) (0.5X TAE buffer) for 16 h at 60 V and a constant temperature of 60˚C. The gel was stained in 0.5 µg ethidium bromide ml^{-1} for 1 h and subsequently visualized using the Syngene Bioimaging System (Synoptics Ltd., United Kingdom). Band matching to the DGGE marker and richness calculations (total

number of bands in a given profile) were performed using the GeneTools software program for Windows (Synoptics Ltd., United Kingdom).

3.2.4 SEQUENCING OF DNA FROM DGGE PCR PRODUCTS AND BANDS

Bands observed in the DGGE gel that matched with those in the marker were purified using the QIAquick PCR Purification kit© (Qiagen, Canada). Each band was cut out using a sterile blade and incubated in 50 μl Biotech grade water (Fisher, Canada) for 24 h at 4˚C. After incubation, the tubes were centrifuged for 30 s at 13 000 rpm and 0.5 μl of the supernatant was re-amplified using the second round of amplification protocol. PCR products were re-run alongside their matching marker band in another 8% polyacrylamide gel containing a linear gradient of 35-55% denaturant in the form of urea and formamide and was electrophoresed under the same conditions as previous DGGE gels. This was done to verify that the products being sent in for sequencing were single and did not contain more than one amplicon. For samples that exhibited more than one amplicon, the band matching to the marker was cut out with a sterile blade and re-amplified again using the second round of amplification protocol. PCR products were purified using the QIAquick PCR Purification kit© (Qiagen, Canada) and sequenced using the Applied Biosystems 3130XL Genetic Analyzer with either the F357GC, F357 without a GC clamp, or R518 primers (Zwart *et al.*, 2005). Sequence reaction products were visualized and compared to matching DGGE marker band sequences using Bioedit sequence alignment editor and analysis program (Hall, 1999).

3.2.5 *ANALYSIS OF ANABAENA LEMMERMANNII DGGE BANDS AND GEOSMIN CONCENTRATIONS*

The relationships between the presence of an *An. lemmermanni* band in DGGE gels, richness of DGGE bands, and GC-MS determined concentrations of geosmin were analyzed using Microsoft Excel and Systat© ver. 10. Average total geosmin concentration for sites with *An. lemmermannii* bands was compared to average total geosmin concentration for sites without *An. lemmermannii* bands. Results were graphed and analyzed.

3.2.6 *REPRODUCIBILITY OF SAMPLING AND CHELEX-100 DNA ISOLATION METHODS*

Triplicate DNA isolations using the chelex-100 protocol (Walsh *et al.*, 1991) were performed on seven randomly selected samples to examine the efficiency and reproducibility of sample collection and DNA isolation of Lake Ontario water samples. Subsequent to PCR amplification and DGGE analysis of these samples, profiles among triplicates were scored based on presence/absence of bands and richness similarities were measured using the Jaccard similarity index: $J = 100 ([c / a + b - c])$ where a is the number of bands of sample A, b the number of bands in sample B, c the number of bands that are in common to samples A and B (Lyautey *et al.*, 2005).

RESULTS

3.3.1 IDENTIFICATION OF POTENTIAL GEOSMIN-PRODUCING CYANOBACTERIA IN LAKE ONTARIO USING DGGE

Profiles of the 31 Lake Ontario samples were electrophoresed along four DGGE gels and are depicted in Figures 11-14. The first set of profiles showed band matches to *An. lemmermannii* in six samples belonging to all triplicate stations in 738 as well as 01, 12, and 739. No other bands matched to the marker bands. A negative control is not seen in this image as it was electrophoresed in a subsequent DGGE gel under the same conditions, no bands were observed in this sample as expected (Figure 11). The second set of profiles did not show any band matches to *An. lemmermannii* in any of the samples nor did any bands match to those in the marker (Figure 12). The third set of samples electrophoresed in DGGE showed band matches to *An. lemmermannii* in four samples including two of the triplicate samples for station 5 as well as stations 1193 and 498. No other bands in the samples matched to those in the marker (Figure 13). The final set of profiles showed five samples with a band match to *An. lemmermannii* with two belonging to the triplicates from station 1001 as well as 741, 750, and 744. In addition, four samples showed bands from *Ap. flos-aquae* with all triplicates from station 1001 possessing the band along with 741 and 750. No other bands matched to marker bands in the remaining samples (Figure 14). In total 11 samples showed the presence of *An. lemmermannii* in the DGGE profiles.

An. lemmermannii band sequences from stations 1001, 498, 5, 741, 744, 750, 1193, and 12 were 100% identical to those in the marker. Stations 738 and 739 band sequences were not identical. Station 738 differed by 4 nucleotides (4.2% dissimilarity)

116

of 97 with good quality peaks. Station 739 differed by 3 nucleotides (5.5% dissimilarity) of the 68 with good quality peaks. *Ap. flos-aquae* band sequences from stations 741, and 750 were identical to that of the standard marker while station 1001 differed by 7 nucleotides (6.8% dissimilarity) of the 103 that were sequenced with good quality.

3.3.2 *TRIPLICATE DNA ISOLATION, PCR AND DGGE*

Triplicate DNA isolation using the chelex-100 method was performed on seven randomly selected stations — 738, 86, 55, 71, 5, 82, and 1001 — and similarity indices were calculated and averaged for each set based on the number of detected bands (or richness) (Table 2). Scoring was a measure of presence or absence of a DGGE band for a given profile. Richness similarities ranged from 46.4% (lowest) to 93.3% (highest) and were never 100% identical for a single triplicate set. The average overall similarity for all triplicates was 71.1%.

3.3.3 *GEOSMIN CONCENTRATIONS VS. DGGE PROFILES*

Geosmin concentrations varied across Lake Ontario sites in 2005, ranging from 2.84 – 62.56 ng · l^{-1}. Since the perceived threshold odour concentration for geosmin is approximately 10 ng · l^{-1}, geosmin levels were divided into low (< 10 ng · l^{-1}), moderate (10 to 20 ng · l^{-1}), and high (>20 ng · l^{-1}) concentrations for each location (Table 3). The general trend observed was that higher geosmin levels tended to reside in samples from the western basin while lower geosmin levels were observed in the eastern basin. The central region had both high and moderate geosmin concentrations.

Richness for sites varied from 0-17 bands in a profile with the higher numbers generally in the western regions and the lower in the east (Table 3). Average richness values for

117

eastern, central, and western regions were 5.95, 5.50, and 8.70 respectively with an overall lake average of 7.29 (Figure 15). Abundance (more than 10 bands) was observed in 10 samples, six of which were located in the western basin, two in the Bay of Quinte, and two in the eastern basin. To determine the relationship between geosmin concentration and richness for each station a plot of the values (Figure 16) revealed an $R^2 = 0.0016$, indicating there was a poor linear correlation between the two variables. Geosmin concentrations do not increase with increasing richness.

The only geosmin-producing species identified in the DGGE profiles were *An. lemmermannii*, detected at 11 stations, two of which occurred in the Bay of Quinte one in the central region, and seven in the western region. Seven *An. lemmermannii* bands were detected in the high geosmin group while three were detected in the low geosmin group. One *An. lemmermannii* band was detected in the moderate geosmin level group. In addition, *An. lemmermannii* was not identified in the DGGE profiles of stations 740, 33 and 3, yet geosmin levels were in the high range for these sites. The average geosmin concentration from stations with a presence of the *An. lemmermannii* band was 26.70 ng · l^{-1}, while the average geosmin concentration for stations with no detected *An. lemmermannii* was 13.30 ng · l^{-1}. The relationship of geosmin concentrations with the presence of *An. lemmermannii* bands and absence of *An. lemmermannii* bands for samples with low geosmin concentrations is shown in Figure 17. The t-test indicated that the null hypothesis of geosmin concentration and presence of An. lemmermannii bands do not correlate could not be rejected. The p value of 0.087 was too high to reject the null hypothesis. It is therefore not possible to state that increased geosmin concentrations correlated to the presence of *An. lemmermannii* DGGE bands. A map of Lake Ontario

with sampling sites and geosmin concentrations in comparison to presence/absence of

An. lemmermannii bands in DGGE is depicted in Figure 18; discrepancies have been

noted.

Figure 11. DGGE profiles of 16S rRNA-V3 amplified samples from Lake Ontario. Lanes 1-3 show triplicate samples isolated and amplified under identical conditions. Lanes 1-4, 6, and 7 had band matches to *An. lemmermannii*. Lanes include 16S rRNA-V3 standard (lanes S), Stn 738-1 (lane 1), Stn 738-2 (lane 3), Stn 738-3 (lane 3), Stn 01 (lane 4), Stn 22 (lane 5), Stn 12 (lane 6), Stn 739 (lane 7), Stn 89 (lane 8), Stn 79 (lane 9), Stn 75 (lane 10), Stn 1196 (lane 11), Stn 90 (lane 12). Asterisks represent *An. lemmermannii* bands.
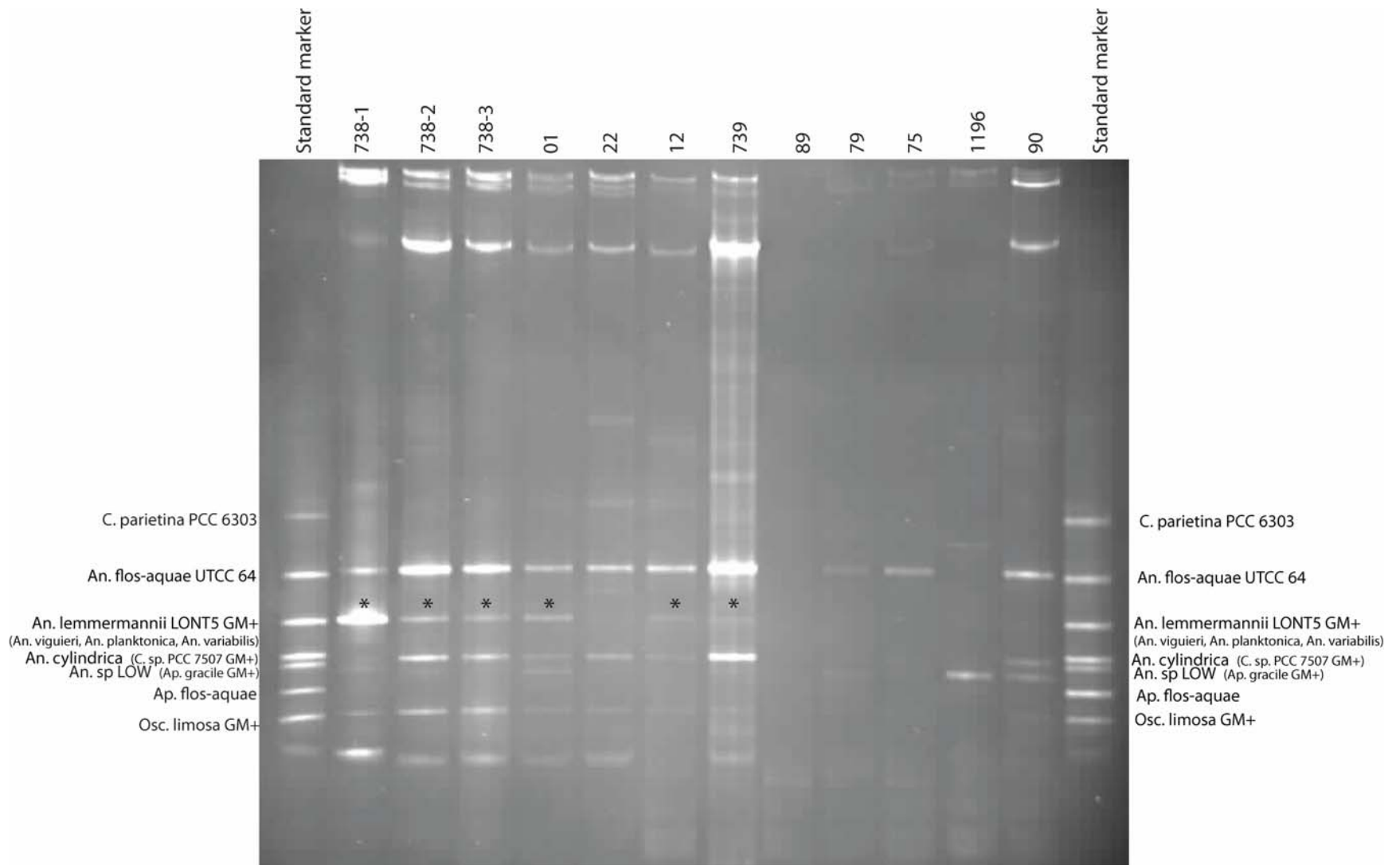
Figure 12. DGGE profiles of 16S rRNA-V3 amplified samples from Lake Ontario.

Lanes 1-3 show triplicate samples isolated and amplified under identical conditions. No

lane had band matches to *An. lemmermannii* or any other isolates. Lanes include 16S

rRNA-V3 standard marker (lanes S), Stn 86-1 (lane 1), Stn 86-2 (lane 3), Stn 86-3 (lane

3), Stn 98 (lane 4), Stn 81 (lane 5), Stn 66 (lane 6), Stn 103 (lane 7), Stn 740 (lane 8),

Stn 78 (lane 9), Stn 100 (lane 10), negative control (lane 11).
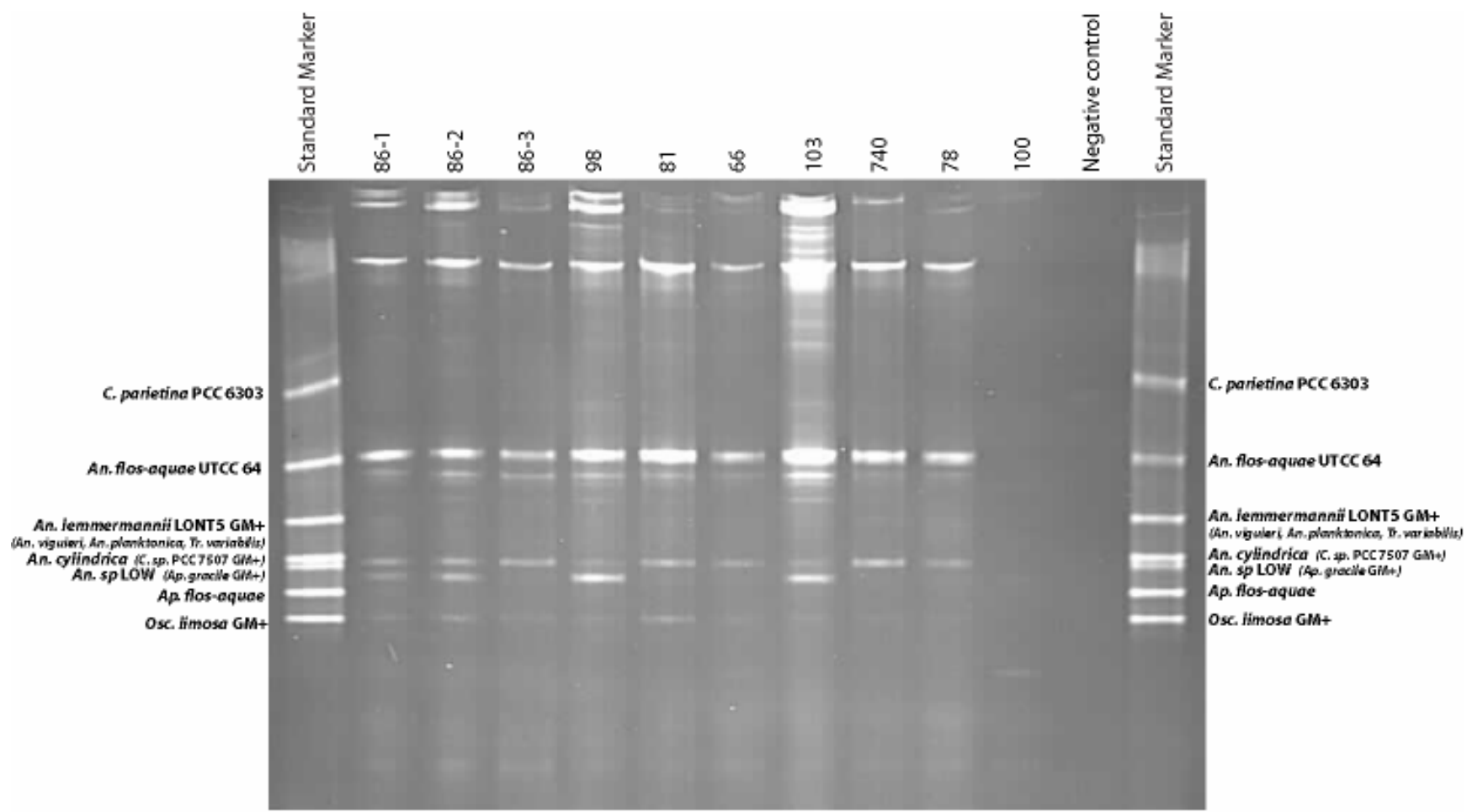
Figure 13. DGGE profiles of 16S rRNA-V3 amplified samples from Lake Ontario.

Lanes 1-6 show triplicate samples isolated and amplified under identical conditions.

Lanes 4, 5, 7, 10 had band matches to *An. lemmermannii*. Lanes include 16S rRNA-V3

standard (lanes S), Stn 55-1 (lane 1), Stn 55-2 (lane 3), Stn 55-3 (lane 3), Stn 5-1 (lane

4), Stn 5-2 (lane 5), Stn 5-3 (lane 6), Stn 1193 (lane 7), Stn 13 (lane 8), Stn 33 (lane 9),

Stn 498 (lane 10), Stn 742 (lane 11), Stn 754 (lane 12), negative control (lane 13).

Asterisks indicate *An. lemmermannii* band match.
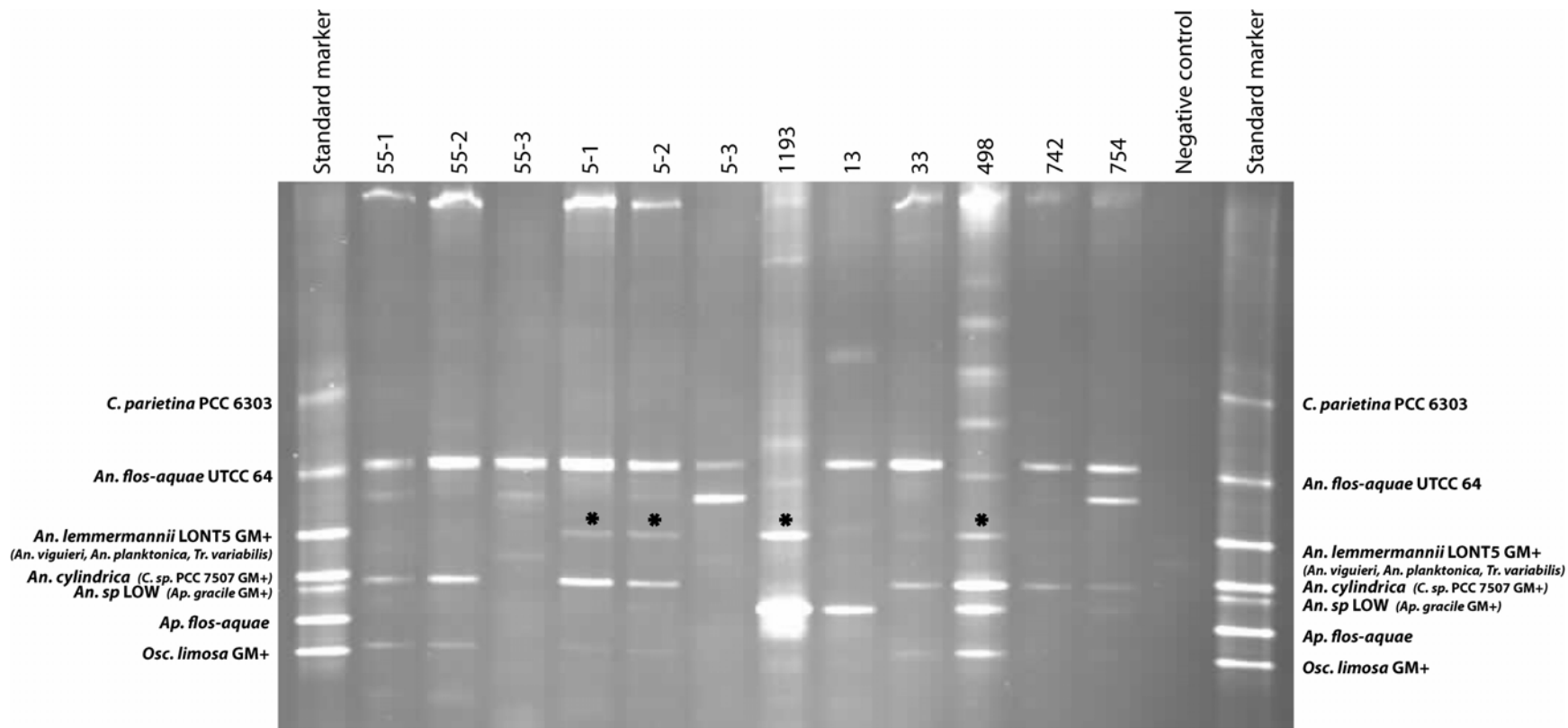
Figure 14. DGGE profiles of 16S rRNA-V3 amplified samples from Lake Ontario.
Lanes 1-9 show triplicate samples isolated and amplified under identical conditions.
Lanes 8-12 had band matches to *An. lemmermannii* (asterisks just above bands). Lanes
7-11 had band matches to *Ap. flos-aquae* (asterisks just above bands) . Lanes include
16S rRNA-V3 standard (lanes S), Stn 71-1 (lane 1), Stn 71-2 (lane 3), Stn 71-3 (lane 3),
Stn 82-1 (lane 4), Stn 82-2 (lane 5), Stn 82-3 (lane 6), Stn 1001-1 (lane 7), Stn 1001-2
(lane 8), Stn 1001-3 (lane 9), Stn 741 (lane 10), Stn 750 (lane 11), Stn 744 (lane 12),
negative control (lane 13). Asterisks indicate band matches to *An. lemmermannii* (upper)
and *Ap. flos-aquae* (lower)
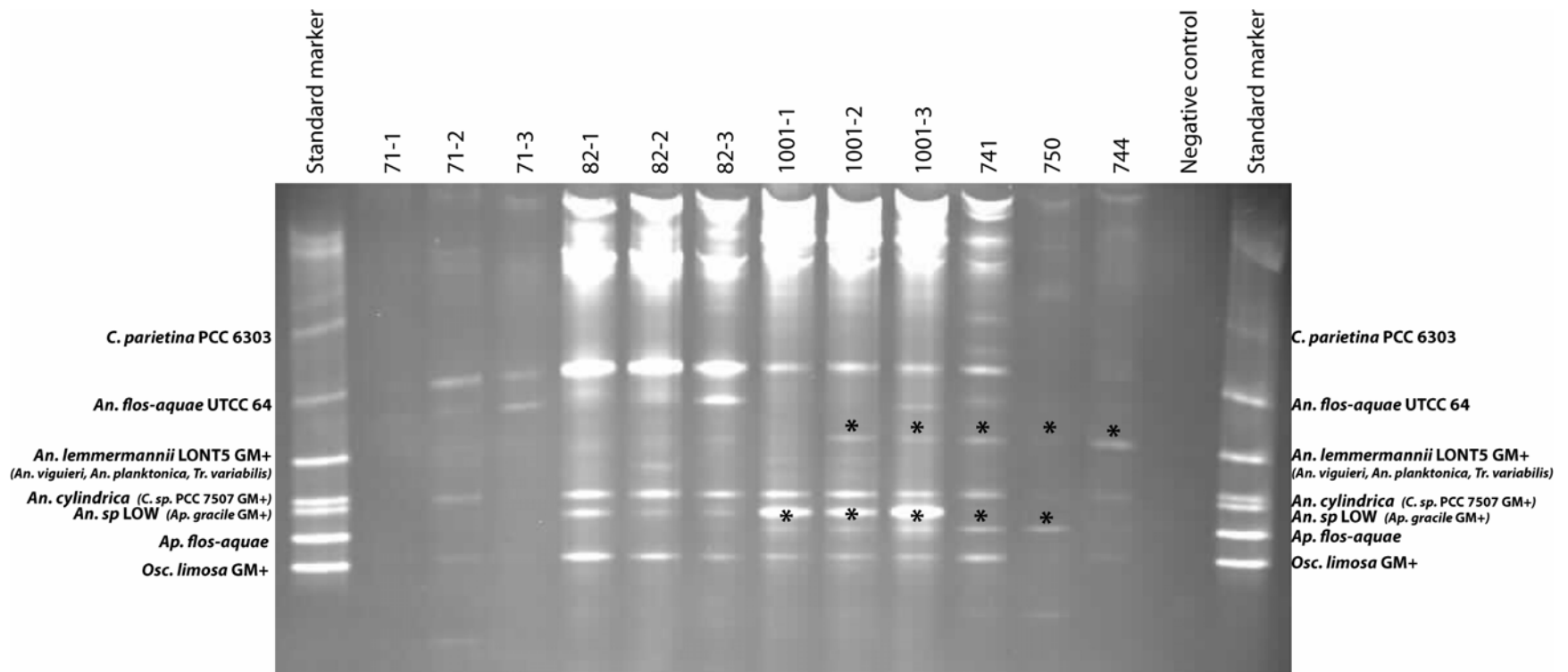
Table 2. Average Jaccard Similarity Indices for richness of triplicate DGGE profiles. Richness was measured as number of bands in a profile. Banding was scored as presence or absence in a profile.

Station	Jaccard Similarity Index Average (%)
738	72.7
86	90.5
55	46.4
71	52.4
5	68.2
82	93.3
1001	74.4
Overall	71.1

Table 3. Geosmin Concentrations and Richness for sampling regions surrounding Lake Ontario. BQ = Bay of Quinte; E = Eastern; NE = Northeast; SE = Southeast; C = Central; W = West; NW = Northwest; SW = Southwest

Low (<10 ng · l^{-1})				Moderate (10 to 20 ng · l^{-1})				High (> 20 ng · l^{-1})			
Region	Station	Geosmin (ng · l^{-1})	Richness	Region	Station	Geosmin (ng · l^{-1})	Richness	Region	Station	Geosmin (ng · l^{-1})	Richness
BQ	1193	2.84	10	E	100-1	16.33	1	BQ	498	62.56	11
BQ	1196	4.41	4	NE	103	13.35	14	C	744	46.10	5
BQ	90	5.83	7	NE	79	11.65	3	C	33	20.80	6
E	75	6.81	7	NE	754	15.26	6	NW	12	43.20	7
E	89	7.22	0	SE	66	14.01	7	NW	738-1	29.20	11
E	98	6.25	10	SE	71-1	11.18	3	W	5-1	32.60	7
NE	81	5.00	5	C	55-1	14.31	7	W	739	22.40	17
NE	82-1	6.71	9	C	742	15.42	4	W	750	25.32	4
W	1001-1	4.50	10	SW	22	11.10	10	W	3	29.30	5
SW	741	6.3	12	SW	86-1	14.40	8	SW	740-2	37.50	5
				W	01	19.20	10				

Figure 15. Distribution of DGGE richness among regions for Lake Ontario

collection sites. Average richness for east, central and west was 5.95, 5.50, and 8.70

respectively. Asterisk denotes richness average for triplicate stations.

Richness (total number of DGGE bands)
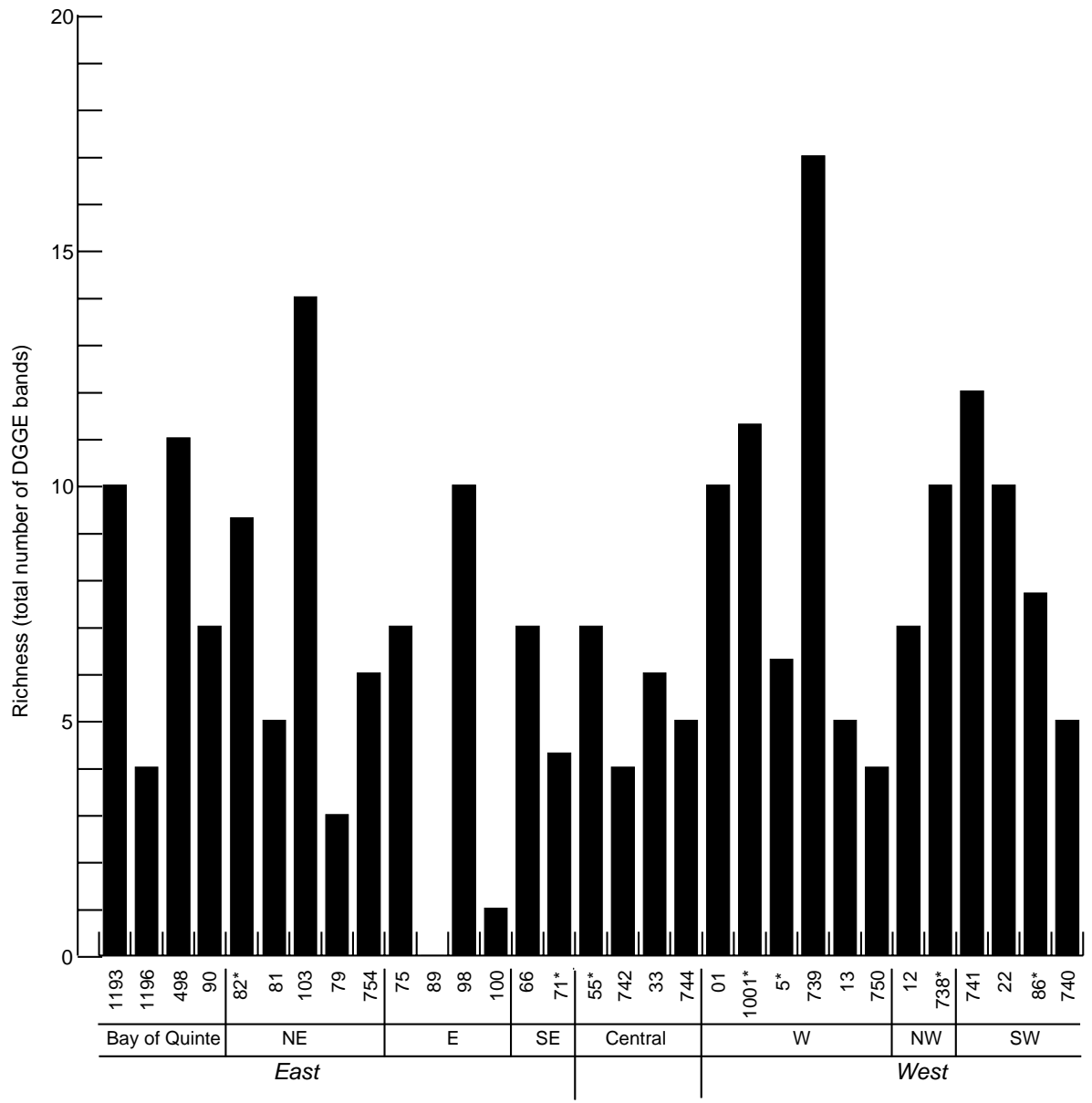
Lake Ontario station number and region

Station	Region	Group
1193, 1196, 498, 90	Bay of Quinte	East
82*, 81, 103, 79, 754	NE	East
75, 89, 98, 100	E	East
66, 71*	SE	East
55*, 742, 33, 744	Central	
01, 1001*, 5*, 739, 13, 750	W	West
12, 738*	NW	West
741, 22, 86*, 740	SW	West

Figure 16. The relationship of geosmin and DGGE band richness. No correlation was observed among the data. High geosmin concentrations do not indicate an increase in species richness. $R^2=0.0016$.

$y = 0.1528x + 16.834$
$R^2 = 0.0016$
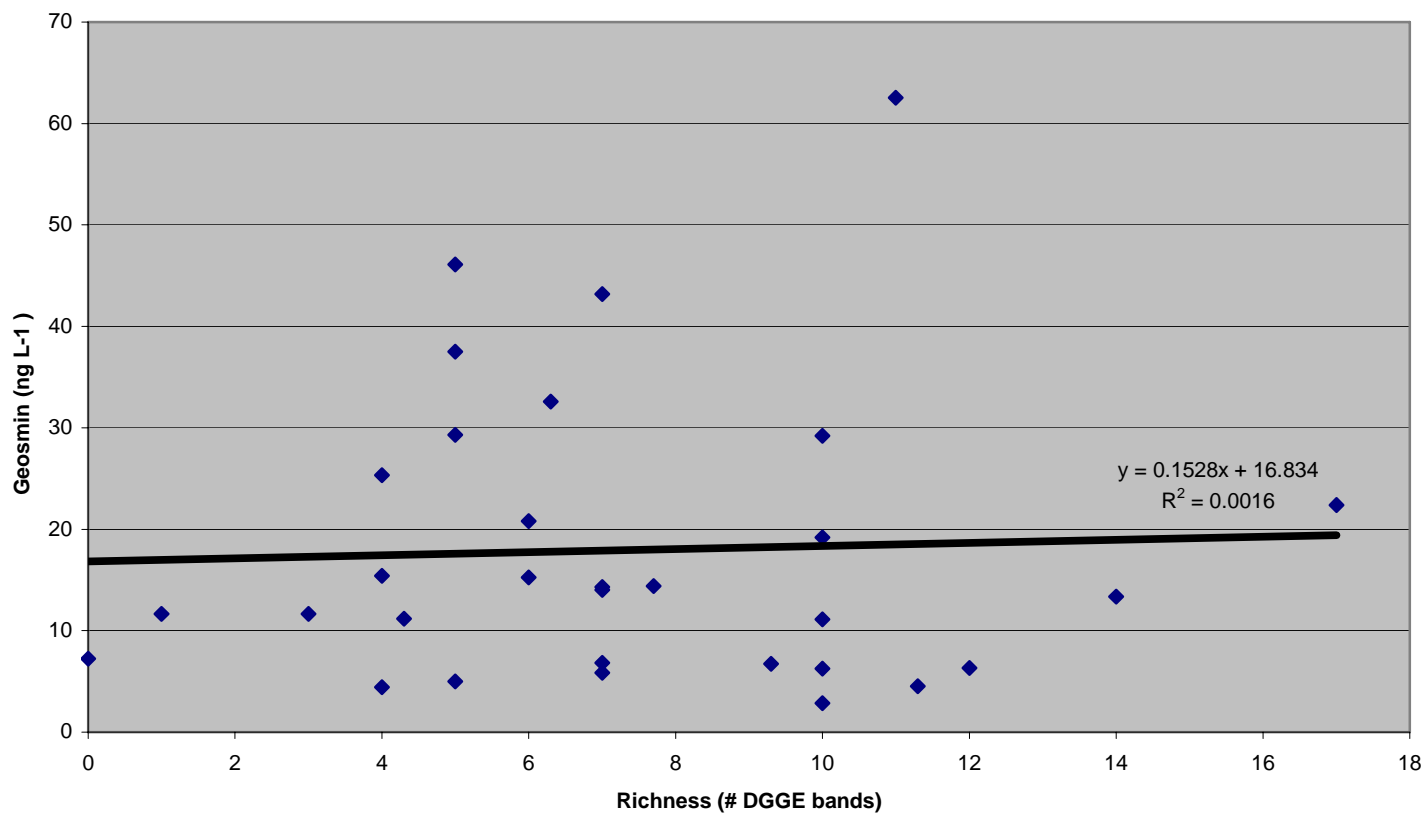
Figure 17. Geosmin concentrations (ng · l$^{-1}_{\text{log10}}$) for Lake Ontario sampling sites in comparison to presence and absence of DGGE *Anabaena lemmermannii* band from DGGE profiles. Sites with increased geosmin (avg. 26.70 ng · l^{-1}) showed a relationship to the presence of *An. lemmermanni* in the DGGE profiles. Sites with decreased geosmin concentrations (avg. 13.30 ng · l^{-1}) showed a relationship with an absence of *An. lemmermannii* in the DGGE profiles. The t-test showed no significant correlation (p=0.087) of mean geosmin concentrations (ng · l$^{-1}_{\text{log10}}$) and presence and absence of *An. lemmermannii* in DGGE profiles.
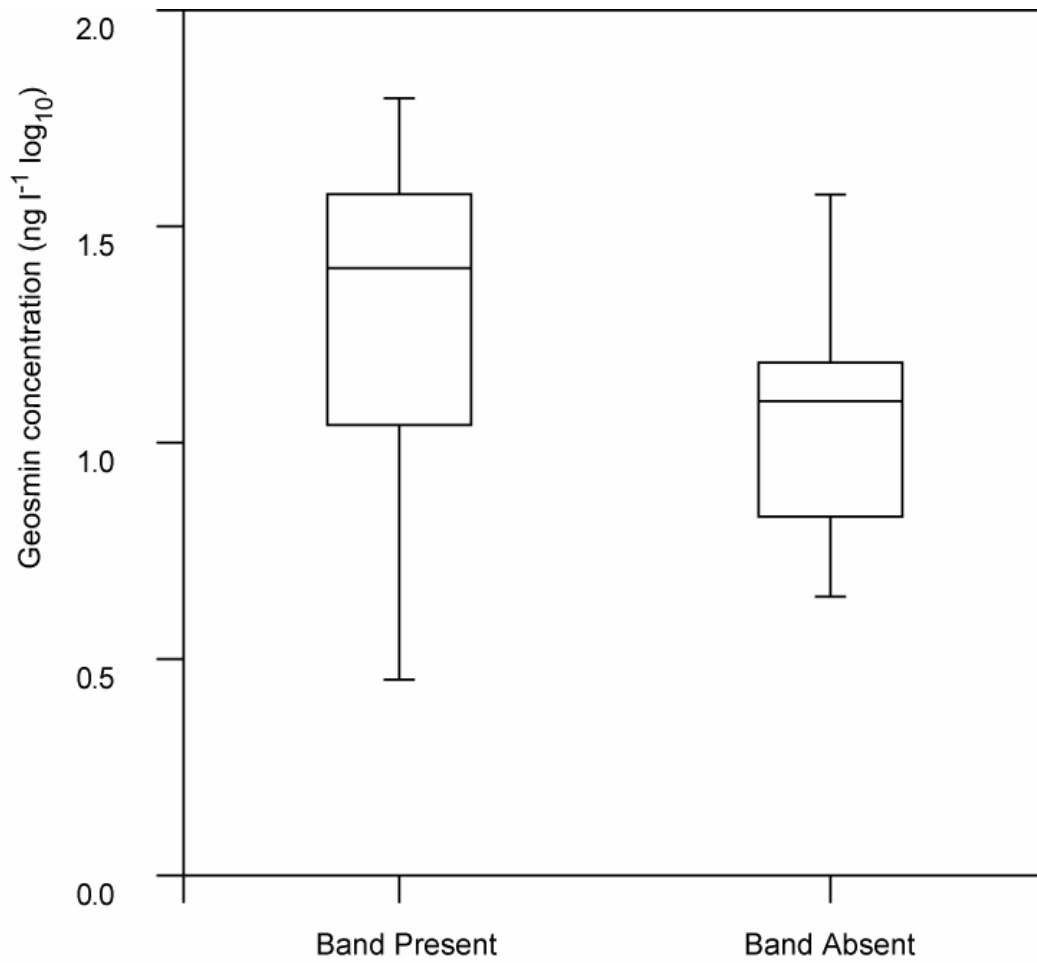
**Figure 18. Range of geosmin concentrations and presence of *An. lemmermannii*
DGGE bands for Lake Ontario sampling locations.** High geosmin was considered >20
ng · l^{-1}, Low geosmin is <20 ng · l^{-1}

Lake Ontario Environmental Sample Analysis: DGGE and GC-MS Data

Taste and Odour in Drinking Water

2005-00-003

CCGS Limnos

August 29-September 2, 2005

Legend

- A. lemmermannii DGGE band & high [geosmin]
- A. lemmermannii DGGE band & low [geosmin]
- No An. lemmermannii DGGE band & low [geosmin]
- No A. lemmermannii DGGE band & high [geosmin]
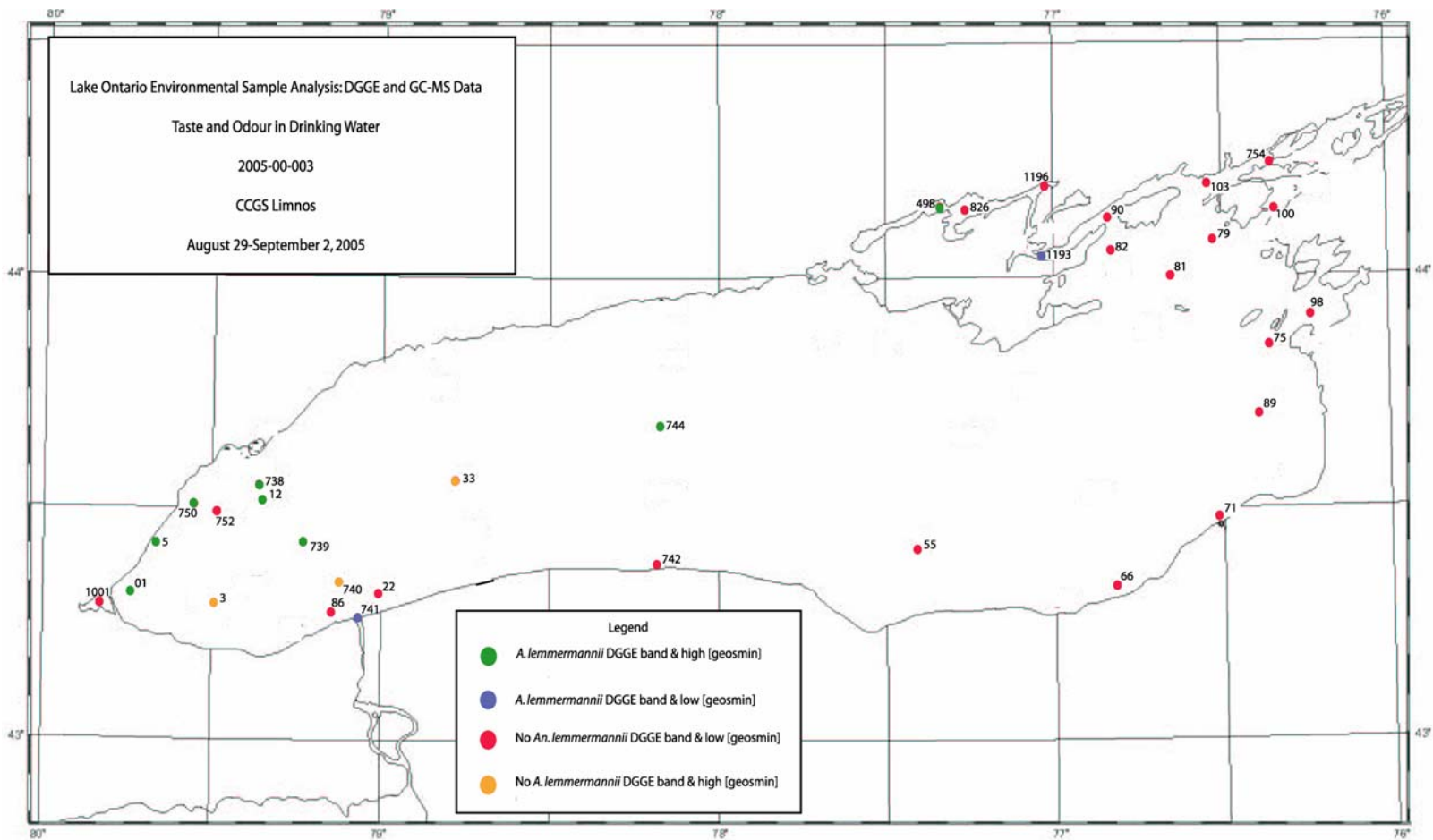
3.4 DISCUSSION

3.4.1 *DGGE PROFILES OF LAKE ONTARIO SITES — WHAT DO THEY INDICATE?*

The DGGE profiles from Lake Ontario have provided insight into community richness as well as composition when compared to the developed 16S rRNA-V3 marker. Species richness assessed with DGGE is a measure of total bands in a given profile and is indicative of the relative number of species present in a sample (Yu and Morrison, 2004). Since the introduction of filter-feeding dreissenid (zebra or quagga mussel) populations over the past decade, a dramatic change in substrate, nutrient cycling, and food webs in the inshore areas has been linked to an overall increase in lake transparency and phyoplankton richness (Mills *et al.*, 1999). Richness in DGGE bands for Lake Ontario were greater in the western basin in comparison to the central and eastern regions, indicating that there was more diversity in the west. Richness values determined using DGGE in this study likely represent values resembling actual counts however it is by no means an exact measure. Previous studies have determined that actual cell counts have corresponded to DGGE band numbers and support the reliability of the technique for identifying species of organisms that are not at trace levels (Kolmonen *et al.*, 2004). In general, DGGE cannot detect populations whose abundance is less than 1% of the total cell count (Muyzer *et al.*, 1993) therefore this technique would be useful only for analyzing cyanobacterial communities that are not at trace levels (under 1%). A comparison of actual cell counts to richness and community composition would be a useful approach to determining the accuracy of DGGE in predicting species richness and should be pursued in future studies.

In the event that actual cell counts and identification of the species from fixed stations were possible, it was expected that the populations comprising greater than 1% of

the sample would correspond to a band in the DGGE profile. If counts could have been performed and did not correspond to DGGE results it is possible that the DNA isolation and/or sample collection was the root of the problem and may need revising in future studies. The chelex-100 DNA isolation method was chosen based on the simplicity and rapidity of the procedure in comparison to conventional phenol-chloroform protocols. If the DGGE technique were to be used in the analysis of high volumes of samples in management, a conventional method such as phenol-chloroform would slow down the overall process by more than 24 hours whereas the chelex-100 protocol can be as rapid as one to two hours (Walsh *et al.*, 1991). This DNA isolation technique has been used in forensics and requires fewer transfers than conventional phenol-chloroform extraction leaving less opportunity for cross-contamination (Walsh *et al.*, 1991) and making it a good candidate for use in high-throughput of samples for characterization using DGGE. In addition, this extraction method has been proven to be as effective as phenol-chloroform (Walsh *et al.*, 1991) therefore it is unlikely that the DNA isolation in this study has led to false results. The discrepancies observed in the reproducibility of the DGGE profiles observed in the triplicates suggests that the sample collection in conjunction with DNA isolation should be re-examined.

For sampling, the cyanobacteria cells were collected on a 47 mm diameter filter of which only a fraction (1/16th) was used in DNA isolation. This small portion may not be representative of the total cyanobacterial communities collected and may in the future lead to non-corresponding results to actual cell counts and identification. Additionally, this method assumes a uniform distribution of the cells on the filter, which does not always occur during filtration. Use of the entire filter for DNA extraction may result in DGGE

profiles more representative of populations in the sample. In the future, DNA isolation and sampling protocols should be tested in conjunction with other techniques to determine the most accurate representation of Cyanobacteria diversity.

Richness values from DGGE profiles did not correlate to geosmin concentration. It was not hypothesized that a relationship would exist between these variables since geosmin is strain-specific. A high concentration of geosmin could have been present without a diversity of cyanobacteria. For sites where richness was high, it was not necessarily due to the presence of increased cyanobacteria. Furthermore, for sites with low richness and increased geosmin concentrations, it is possible that dissolved geosmin was present in the absence of a diversity of cyanobacteria.

3.4.2 IS *ANABAENA LEMMERMANNII* THE SOURCE OF GEOSMIN IN LAKE ONTARIO?

Analysis of the 31 stations for presence of *An. lemmermannii* using DGGE revealed many expected results with a few that are questionable. The Saint Lawrence River (SLR) and the eastern basin of Lake Ontario have suffered severe and prolonged taste and odour caused primarily due to attached cyanobacteria, growing on substrates and macrophytes in patchy inshore zones rather than free-floating plankton in surface waters (Watson, 2004). Additionally, dreissenid-associated or other actinomycetes may contribute considerably to taste and odour in this region (Watson, 2004). Being a planktonic species, *An. lemmermannii* was not hypothesized to be prevalent in the eastern basin. None of the DGGE profiles included *An. lemmermannii* and geosmin concentrations were low to moderate (between 5 and 15.26 ng l^{-1}) with none in the extreme range. These results suggest that planktonic free-floating species of cyanobacteria, including *An. lemmermannii,*

were not primary contributors to taste and odour in this region. Results from Watson and Ridal (2004) also determined that the associated biofilms on macrophytes contained significant amounts of other geosmin-producing cyanobacteria, notably Oscillatoriales, major contributors to taste and odour in this region therefore it was not surprising that fewer planktonic cyanobacteria and no *An. lemmermannii* was detected in the eastern regions.

The Bay of Quinte is a region that experiences yearly eutrophication and summer blooms of cyanobacteria despite substantial nutrient loading reductions since the 1970s. *Anabaena* and *Aphanizomenon* largely dominated these blooms, however over the past 5 years there has been an 80% increase in *Microcystis* (Nicholls *et al.*, 2002). Although *M. aeruginosa* and *M. wesenbergii* isolates were used in the development of the DGGE marker, they were not focal species in this study because they do not produce geosmin. They were to be included in the marker if their bands were clearly separate from others but they positioned close to more valuable geosmin producing isolate bands and were thus not included. *Microcystis* populations were thus not analyzed. Of the four profiles taken from this region of the lake, two included *An. lemmermannii,* station 498, the most inward region sampled in the Bay, and 1193. Station 498 had the highest geosmin levels of all those sampled, more than six times the odour threshold concentration and it is likely that this region may have suffered a taste and odour episode during this time (Watson, unpublished). Observation of a band corresponding to *An. lemmermannii* in the DGGE profile was therefore not surprising for this region since species of *Anabaena* and other cyanobacteria are not uncommon to these waters during the late summer (Watson, 2004). This species is likely the primary contributor to geosmin in this region. The *An.*

lemmermannii band from station 1193 did not correspond to high geosmin levels and was one of the discrepancies observed in this study. Since the 16S rRNA-V3 region was concluded not to be species discriminative (refer to chapter 2), it is likely that for this result the band is indicative of another geosmin non-producing species. This locus was representative of *An. lemmermannii*, *An. viguieri*, *An. planktonica*, and *Tr. variabilis* in this study and suggests that that the band detected in this sample likely corresponds to one or more of these three or possibly another species from the Nostocales.

In addition to the presence of *An. lemmermannii*, *An. flos-aquae* was detected at station 498, a result that was not unexpected. This species is common in the Bay of Quinte and is a geosmin non-producer but can produce toxic metabolites such as anatoxin-a that have been implicated in water impairment for recreational use and drinking (Gupta *et al.*, 2002). No other stations from the Bay of Quinte had bands that matched to the DGGE standard marker. It is probable that *Microcystis* populations were present in the Bay of Quinte but no detection was possible without a marker band. Future studies could investigate the potential of DGGE for detection of toxin-producing *Microcystis* populations in this region.

Planktonic *An. lemmermannii* is common in central offshore Lake Ontario and thrive in the warmer offshore waters. These populations are brought to water treatment plant intakes during downwelling events in the offshore region (Skafel and Yerubandi, 2003). The most northern station of the central region was the only one that possessed an *An. lemmermannii* band in its DGGE profile and as expected corresponded to high geosmin ($46.10 \text{ ng} \cdot \text{l}^{-1}$). It is possible that at the time of collection, water at this site was warmer and growth conditions were optimal for *An. lemmermannii* to thrive. Another possibility is that

An. lemmermannii populations were located in this region due to circulation of the waters. At station 33 high geosmin levels were detected but no *An. lemmermannii* band or other potential geosmin-producing cyanobacteria were detected. This discrepancy is indicative that the collection and/or DNA isolation methods may need revising for reasons previously stated. It is also possible that dissolved geosmin levels were present in this region while *An. lemmermannii* was absent. The western basin of Lake Ontario experiences geosmin peaks annually and of the 11 profiles with *An. lemmermannii* bands, seven occurred in the western basin, an expected result considering *An. lemmermannii* is generally common in this region during the late summer. Richness was also higher for sites in this area. Six of these corresponded to geosmin levels greater than 19 ng $\cdot$ l^{-1} located to the northwest, while one location in the southwest (station 741) was likely a discrepancy due to low geosmin concentrations at this site. The high number of profiles with *An. lemmermannii* and increased geosmin in the northwest basin indicates that this species is likely the primary contributor to geosmin in this region. Since station 741 experienced low to moderate geosmin levels (6.3 ng $\cdot$ l^{-1}), it is possible that the detected band at this locus corresponds to *An. lemmermannii*, however since most of the *An. lemmermannii* bands detected across the lake were above 19 ng $\cdot$ l^{-1} it is most probable that the band is representative of another geosmin non-producing species. Another discrepancy was observed from station 740 which was expected to possess a band representative of *An. lemmermannii* as it was experiencing geosmin concentrations greater than 20 ng $\cdot$ l^{-1}. It is likely that the species was present and the collection and DNA extraction did not use a piece of filter with representative cells, again supporting that these methods may need revision.

The northwest basin of Lake Ontario not only suffers from taste and odour episodes but occasionally experiences increased levels of toxins corresponding to increased levels of *Microcystis* sp. (Murphy *et al.*, 2003). *Microcystis* was not analyzed in this study however *Ap. flos-aquae* was detected at stations 741 (southwest), 750 (northwest) and 1001 (west, Hamilton Harbour) for the western region. It was not unexpected that *Ap. flos-aquae* was observed at station 1001 because the isolate used in the standard marker was derived from this site. In addition, significant blooms dominated by *Aphanizomenon* and/or *Microcystis* have been common in this region since 1999. Low geosmin was detected at stations 1001 and 741 which suggest that the strain of *Ap. flos-aquae* at these sites was not contributing to geosmin. It was not surprising to see a band match to *Ap. flos-aquae* in the station 750 DGGE profile since this location is not very far from Hamilton Harbour and may have been experiencing similar lake dynamics at the time of collection. Station 750 had 25.32 ng l^{-1} of geosmin that was likely caused by *An. lemmermannii* (detected by DGGE). According to the t-test results, there was no significant correlation of the presence of *An. lemmermannii* DGGE bands and increased geosmin concentrations. The lack of correlation between these two variables was likely due to the fact that the 16S rRNA-V3 region did not discriminate to the species level. Samples with *An. lemmermannii* bands and low geosmin were likely indicative of a different species of geosmin non-producing *Anabaena.* This would account for the discrepancies observed in the data and the high standard deviation (0.442 ng · l^{-1}log $_{10}$) from the mean (1.274 ng · l^{-1}log $_{10}$). Samples with an absence of *An. lemmermannii* detected using DGGE showed less standard deviation (0.250 ng · l^{-1}log $_{10}$) from the mean (1.056 ng · l^{-1}log $_{10}$) indicating that the technique is more reliable for undetected *An. lemmermannii* and low geosmin concentrations. Although statistically there

was no correlation between the presence of *An. lemmermannii* and increased geosmin concentrations, on average the geosmin concentrations were higher and corresponded to regions where *An. lemmermannii* is common. This suggests that with the use of a more species-specific gene region, DGGE may be useful in monitoring potential taste and odour-causing cyanobacterial populations.

Band sequencing for *An. lemmermannii* and *Ap. flos-aquae* was not of highest quality until the protocol was revised. Simple excision and re-amplification was not sufficient for obtaining sequences of good quality. It was thought that the acrylamide was interfering with sequencing, thus concentrations used as template in re-amplification were reduced. This alone was not effective therefore to identify if multiple amplicons from the excised region were interfering with sequencing, the re-amplified products were re-run in DGGE. In the majority of samples more than one amplicon was observed, indicating that the excision was likely not precise and caused poor quality sequences. Once the bands were excised and re-amplified a second time, sequencing was successful. Stations 738 and 739 did not match in sequence to the *An. lemmermannii* marker and 1001 did not match to *Ap. flos-aquae* in the marker while the remaining sequences were successfully matched to their corresponding markers. Presence of *An. lemmermannii* and *Ap. flos-aquae* in profiles were coresponded to the expected regions in which they were observed suggesting that the DGGE profiles are likely representative of the communities at that time and discrepancies are probably due to protocol imperfections that should be revised and improved in future studies.

3.5   CONCLUSIONS

Although the statistical analyses showed that the was no correltation of the relationship of geosmin to the presence of an *An. lemmermannii* DGGE band, DGGE community profiles corresponded to geosmin trends in Lake Ontario. With geosmin levels for stations with *An. lemmermannii* bands averaging 26.70 ng · l^{-1} and those with no *An. lemmermannii* bands averaging 13.30 ng · l^{-1}, the data supports that *An. lemmermannii* is likely the primary contributor to Lake Ontario geosmin. The fact that no other potential geosmin-producing species of cyanobacteria were detected in the waters adds support to this conclusion and suggests that *An. lemmermannii* is the primary and possibly sole contributing cyanobacteria to western Lake Ontario and the Bay of Quinte taste and odour caused by geosmin. Additionally, results for the eastern region of the lake are coincident with previous studies and support that geosmin occurrences are not due to planktonic, free-floating cyanobacteria (including *An. lemmermannii*) but likely originate from cyanobacteria-associated biofilms on macrophytes. Profiles were also indicative that *Ap. flos-aquae* was still prevalent in the Hamilton Harbour and the southwest and that these locations could experience increased toxicity during late summer as well if a toxic strain is present. The DGGE profiles were not uncharacteristic of the expected results however with the 16S rRNA-V3 region, a precise account of the cyanobacteria community to the species level was not possible. Although this region gave primarily expected results, occasional discrepancies were observed, indicating that a different or longer nucleotide region may be more informative using this technique. Development of this method with examination of various genetic markers and standardization of the technique may lead to a promising future for use of DGGE in the prediction and management of taste and odour events caused by cyanobacteria.

REFERENCES

Casamatta, D. A., M. L. Vis and R. G. Sheath (2003). Cryptic species in cyanobacterial systematics: a case study of *Phormidium retzii* (Oscillatoriales) using RAPD molecular markers and 16S rDNA sequence data. Aquatic Botany 77: 295-309.

Dionigi, C. P. (1995). The effects of copper sulfate on geosmin biosynthesis by *Streptomyces tendae, Streptomyces albidoflavus*, and *Penicillum expansum*. Water Science and Technology 31(11): 135-138.

Ershov, Y. V., R. R. Gantt, J. Francis X. Cunningham and E. Gantt (2002). Isoprenoid Biosynthesis in *Synechocystis* sp. Strain PCC6803 Is Stimulated by Compounds of the Pentose Phosphate Cycle but Not by Pyruvate or Deoxyxylulose-5-Phosphate. Journal of Bacteriology 184(18): 5045-5051.

Gupta, N., A. S. B. Bhaskar and P. V. L. Rao (2002). Growth characteristics and toxin production in batch cultures of *Anabaena flos-aquae*: effects of culture media and duration. World Journal of Microbiology 18: 29-35.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95-98.

Janse, I., M. Meima, W. E. A. Kardinaal and G. Zwart (2003). High-Resolution Differentiation of Cyanobacteria by Using rRNA-Internal Transcribed Spacer Denaturing Gradient Gel Electrophoresis. Applied and Environmental Microbiology 69(11): 6634-6643.

Jüttner, F. (1995). Physiology and biochemistry of odorous compounds from freshwater Cyanobacteria and algae. Water Science and Technology 31(11): 69-78.

Kolmonen, E., K. Sivonen, J. Rapala and K. Haukka (2004). Diversity of Cyanobacteria and heterotrophic bacteria in cyanobacterial blooms in Lake Joutikas, Finland. Aquatic Microbial Ecology 36: 201-211.

Lange, B. M., T. Rujan, W. Martin and R. Croteau (2000). Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. Proceedings of the National Academy of Sciences 97(24): 13172-13177.

Litvaitis, M. K. (2002). A molecular test of cyanobacterial phylogeny: inferences from constraint analyses. Hydrobiologa 468: 135-145.

Lyautey, E., B. Lacoste, L. Ten-Hage, J.-L. Rols and F. Garabetien (2005). Analysis of bacterial diversity in river biofilms using 16S rDNA PCR-DGGE: methodological settings and fingerprints interpretation. Water Research 39: 380-388.

Mills, E., J. Chrisman, B. Baldwin, R. O. Owens, T. Howell, E. Roseman and M. Raths (1999). Changes in the Dreissenid community in the lower Great Lakes with emphasis on southern Lake Ontario. Journal of Great Lakes Research 25: 187-197.

Murphy, T. P., K. Irvine, J. Guo, J. Davies, H. Murkin, M. Charlton and S. B. Watson (2003). New microcystin concerns in the lower Great Lakes. Water Quality Research Journal of Canada 38(1): 127-140.

Muyzer, G. (1999). DGGE/TGGE a method for identifying genes from natural ecosystems. Current Opinion in Microbiology 2: 317-322.

Muyzer, G., E. C. DeWaal and A. Utterlinden (1993). Profiling complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Applied and Environmental Microbiology 59(3): 695-700.

Nakashima, S. and M. Yagi (1992). Iron forms that influence the growth and musty odor production of selected Cyanobacteria. Water Science and Technology 25: 207-216.

Nicholls, K. H., L. Heintsch and E. Carney (2002). Univariate step-trend and multivariate assessments of the apparent effects of P loading reductions and zebra mussels on the phytoplankton of the Bay of Quinte. Journal of Great Lakes Research 28: 15-31.

Rajaniemi, P., P. Hrouzek, K. Kastovska, R. Williame, A. Rantala, L. Hoffman, J. Komarek and K. Sivonen (2005). Phylogenetic and morphological evaluation of the genera *Anabaena*, *Aphanizomenon*, *Trichormus* and *Nostoc* (Nostocales, Cyanobacteria). International Journal of Systematic and Evolutionary Microbiology 55: 11-26.

Saadoun, I., K. K. Schrader and W. T. Blevins (2001). Environmental and nutritional factors affecting geosmin synthesis by *Anabaena* sp. Water Research 35(5): 1209-1218.

Skafel, M. and R. Yerubandi (2003). Physical Processes in Western Lake Ontario during a taste and odour episode. Canadian Coastal Conference 2003, NWRI, Environment Canada.

Svenning, M. M., T. Eriksson and U. Rasmussen (2005). Phylogeny of symbiotic cyanobacteria within the genus *Nostoc* based on 16S rDNA sequence analyses. Archives of Microbiology 183: 19-26.

Walsh, S. P., D. A. Metzger and R. Higuchi (1991). Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10(4): 506-513.

Wang, Q., J. R. Deeds, A. R. Place and R. Belas (2005). Dinoflagellate community analysis of a fish kill using denaturing gradient gel electrophoresis. Harmful Algae 4: 151-162.

Watson, S. B. (2004). Aquatic taste and odour: a primary signal of drinking water integrity. Journal of Toxicological and Environmental Health, Part A 67: 1779-1795.

Yu, Z. and M. Morrison (2004). Comparisons of different hypervariable regions of rrs genes for use in PCR-denaturing gradient gel electrophoresis. Applied and Environmental Microbiology 70(8): 4800-4806.

Zwart, G., M. P. K.-v. Agterveld, I. v. Werff-Staverman, F. Hagen, H. L. Hoogveld and H. J. Gons (2005). Molecular characterization of cyanobacterial diversity in a shallow eutrophic lake. Environmental Microbiology 7(3): 365-377.

CHAPTER 4

CYANOBACTERIA SYSTEMATICS

INTRODUCTION

4.1.1 THE CLASSIFICATION OF CYANOBACTERIA

The Cyanobacteria are commonly known as the Cyanophyta or the blue-green algae (Graham and Wilcox, 2000) and are prokaryotic, possessing many characteristics common to the Eubacteria. Over the past few decades there has been considerable controversy over the taxonomic classification of the Cyanobacteria. Historically, this group was classified using morphological data and followed the International Code of Botanical Nomenclature ('Botanical Code') (John *et al.*, 2002). However, because cyanobacteria behave and possess characteristics similar to bacteria, a bacteriological approach to their classification was proposed byRoger Stanier and his colleagues (1978) at the Pasteur Institute in Paris (John *et al.*, 2002). Subsequent to this, Rippka et al. (1979) classified the Cyanobacteria under this code and divided them into five sections based on phenotypic characteristics and GC content. These sections do not, for the most part, correspond to major taxa such as order and family that are recognized by phycologists and have been referred to as sections:

Section I: include the simplest unicellular Cyanobacteria that reproduce by binary fission or budding and their cells are spherical, cylindrical, or oval.

Section II: reproduce by multiple fission and their vegetative cells are always enclosed by an additional fibrous layer on the outer membrane.

Section III: have cells that form a filament or trichome, that are often enclosed by a tubular sheath, the trichome is composed solely of vegetative cells and often displays false branching when the reproducing trichome breaks.

Sections IV: have cells that form a filament or trichome, that are often enclosed by a tubular sheath, can develop heterocysts (nitrogen fixing cells) and akinetes (reproductive spores), trichomes are uniseriate and unbranched with cells that divide at right angles to the long axis of the filament.

Section V: have cells that form a filament, or trichome, that are often enclosed by a tubular sheath, can develop heterocysts (nitrogen fixing cells) and akinetes (reproductive spores), filaments are partly multiseriate with and develop uniseriate lateral branches that break off and become hormogonia (reproductive cells) (Rippka *et al.*, 1979).

Komárek and Anagnostidis (1986, 1988, 1989, 1990) modernized the classification of the Cyanobacteria by including new data such as life cycles, and biochemical data to establish new hypotheses from the previous system of botanical taxonomy. They suggested that discussions of an exclusively bacteriological or botanical approach to this group of organisms was insolvable and proposed applying names to the sections of the Bacteriological Code with some variations. This revised system is comprised of four (orders rather than five sections): the Chroococcales (seven families), Oscillatoriales (six families), Nostocales (four families), and Stigonematales (eight families). These authors do not view this system as definitive and encourage updates as more knowledge becomes available from a variety of morphological, biochemical, physiological, and molecular studies (Anagnostidis and Komárek, 1988, 1990; Komárek and Anagnostidis, 1986, 1989).

One very unique group of prokaryotic oxygenic photosynthesizers are the Prochlorophytes. These organisms carry out photosynthesis using chlorophylls-a and b, but lack phycobiliproteins as light-harvesting pigments. These characteristics distinguish them from Cyanobacteria, which contain phycobiliproteins and lack chlorophyll b (Palenik and Haselkorn, 1992). Although the Prochlorophytes differ in this vital trait, they also share many characteristics with the Cyanobacteria which have made their classification very difficult. Taxonomy for this group was estabilshed by Lewin (1977) and Buger-Wiersma et al. (1989) and included them in the Cyanobacteria class with one order, the Prochlorales, containing three genera: *Prochloron*, *Prochlorothrix*, and *Prochlorococcus* (Burgerwiersma *et al.*, 1989; Lewin, 1977). Because the Prochlorophytes contain chlorophylls-a and b in addition to having stacked thylakoids (not seen in the Cyanobacteria) they were hypothesized to share a common ancestor with the endosymbiont ancestor of the euglenoid, green algal and streptophyte chloroplast (Lewin and Cheng, 1989). Molecular phylogenetic studies however have shown that the Prochlorales likely do not share this common ancestor due to the multiple evolutionary origins displayed in this group (Palenik and Haselkorn, 1992; Palenik and Swift, 1996; Urbach *et al.*, 1992). There has been notable controversy surrounding the order Prochlorales and as a result the order was rejected in 1997 under the Bacteriological Code and subsequently the genera *Prochloron* and *Prochlorococcus* were included in Section II and the genus *Prochlorothrix* was included in Section III of the Cyanobacteria in Bergey's Manual of Systematic Bacteriology (Garrity *et al.*, 2004). Komárek & Anagnostidis (1986, 1988, 1989, 1990) did not classify the Prochlorales into the revised orders and under their classification they have been considered a separate order.

Molecular studies apply phylogenetics as an approach to classification by examining the evolutionary relationships among groups of organisms. A closely related group of organisms is identified by monophyly, a group that contains all taxa descended from a common ancestor. Under the phylogenetic species concept, a species is identified by estimating the phylogeny of closely related populations and finding the smallest monophyletic group (Freeman and Herron, 2001). Cladistic methods use techniques that identify monophyletic groups on the basis of shared derived characters. Characters derived from molecular data such as DNA and protein sequences are commonly used to determine the evolutionary relationships of one organism to another and to construct phylogenetic trees that illustrate these relationships (Freeman and Herron, 2001). Phylogenetic trees can be constructed using various methods, two of which were used in this study; distance and maximum likelihood. Distance methods use the degree of sequence dissimilarity as a measure of evolutionary distance among taxa. This distance can be calculated in several ways depending on the model of character evolution (distance model) used in the analysis (Freeman and Herron, 2001). The first distance model was developed by Jukes and Cantor (JC) (1969) who considered the rate of nucleotide substitution the same for all pairs of the four nucleotides (A, T, G, and C) (Jukes and Cantor, 1969). Since then, many investigators have developed more complicated models such as Kimura 2-Parameter (K2P) and Tamura-Nei (TN) that account for differences in mutations such as transitions and transversions (Kimura, 1980; Nei and Kumar, 2000). Another commonly used but more complex distance model is the general time reversible (GTR) model of nucleotide substitution that uses base frequencies to determine the evolutionary distance among taxa (Rodriguez *et al.*, 1990). Phylogenetic trees can be calculated using a variety of algorithms, the most

155

common being neighbour-joining developed by Saitou and Nei (1987) (Felsenstein, 2004). The approximation given by this algorithm is relatively accurate and it can analyze hundreds of sequences very rapidly in comparison to maximum likelihood methods (Felsenstein, 2004).

The maximum likelihood approach for molecular phylogenies considers all possible trees and calculates the probability of producing the observed data from each tree. This probability is called the likelihood. The tree with the highest likelihood is chosen as the most probable (Felsenstein, 2004). This method also requires the input of a nucleotide substitution model and is very computationally slow and cannot analyze very large datasets (Freeman and Herron, 2001).

Rationale for utilizing molecular data in conjunction with the phylogenetic species concept include: i) DNA and protein sequences are strictly heritable entities whereas morphological and biochemical characters can be affected by environmental conditions, ii) the description of molecular characters is unambiguous while morphological characters are not always clear-cut, iii) molecular data are more amenable to quantitative treatments, and iv) some molecular data allow for very distantly related organisms to be assessed (ie. among different kingdoms and phyla) whereas morphological characters are not comparable at these distances (Graur and Li, 2000). Due to the lack of a consensus phylogeny and an unambiguous system of classification, unresolved evolutionary relationships within the Cyanobacteria continue to persist. A combination of morphological and molecular data of the studied organisms will strengthen the currently used taxonomy and perhaps result in a system that is mutually agreed upon by scientists.

Cyanobacteria possess an rRNA cistron comprised of three genes; the 16S small subunit (SSU), 23S large subunit (LSU) and the 5S subunit each separated by an internal transcribed spacer region (ITS) that occasionally contains a tRNA$_{ala}$ and/or tRNA$_{ile}$ region (refer to Chapter 1). The 16S rRNA gene is often used in phylogenetic studies and within the Cyanobacteria, sequence information from this gene is widely regarded as one of the most valid criteria for determining relationships between closely related groups, such as species or genera. It is the basis for systematic assignment in the latest edition of Bergey's Manual of Systematic Bacteriology and has been useful in distinguishing broad taxonomic groups as well as individual species (Casamatta *et al.*, 2005; Litvaitis, 2002; Rajaniemi *et al.*, 2005; Svenning *et al.*, 2005). The 16S rRNA gene has regions that are variable and other regions that are highly conserved that allow for straightforward nucleotide alignments (Litvaitis, 2002). The conservative nature of the gene, its universal distribution and the availability of information in public databases (GenBank,) make the 16S rRNA gene very useful for taxonomic studies (Svenning *et al.*, 2005).

Establishment of relationships among species of cyanobacteria may provide fundamental information to understanding their role in nature. Phylogenetic relationships determined using genetic characteristics may give insight into the relatedness of organism levels in which no gene can be directly examined, such as a secondary metabolite. Geosmin, an odorous compound implicated in drinking water quality, is a strain-specific secondary metabolite that some cyanobacteria are capable of producing. Although many strains have been known to produce this odorous compound, the factors controlling it are unknown. When conditions are favourable and geosmin-producing cyanobacteria increase in numbers, the earthy, musty muddy odour can often be released into the waters after cell

death and lysis. In large quantities, geosmin evades conventional drinking water treatment, such as activated carbon filtration, and reaches the taps of consumers. Although there are no health effects of geosmin ingestion, the public often associates poor taste and odour in water with compromised safety (Watson, 2004). Since geosmin is a secondary metabolite of isoprenoid biosynthesis, a largely uncharacterized pathway in the Cyanobacteria, there is no known gene that directly translates into the compound. Isopentenyl diphosphate (IPP) is the central intermediate of isoprenoid biosynthesis and, through many conversions, leads to geosmin. IPP can be produced via three possible routes: the mevalonate (MVA), the methylerythritol (MEP) (also called the non-mevalonate) pathway (Lange *et al.*, 2000), and the hypothetical pentose phosphate pathway (Ershov *et al.*, 2002). Many organisms utilize one or more of the pathways however the exact route to geosmin subsequent to IPP production has not been completely elucidated. An overwhelming majority of eubacteria including the cyanobacterium, *Synechocystis* sp. exclusively use the MEP pathway for isoprenoid production. However, *Synechocystis* sp. is the only cyanobacterium that has been investigated for the occurrence of putative genes involved in the production of IPP. A review by Eisenrich (2004) summarized that *Synechocystis* sp. possesses genes only of the MEP pathway, and hypothesized that that the production of geosmin occurs via the MEP pathway. Radiolabelling experiments carried out by Ershov et al., 2002 however showed that although *Synechocystis* sp PCC 6803 possesses genes of the MEP pathway, it is not likely that this is the route to IPP. Their data supported alternative routes of entry of pentose phosphate cycle substrates derived from photosynthesis rather than entry from glycolysis (Chapter 1, Figure 3c) (Ershov *et al.*, 2002). More studies are needed to elucidate the exact route to IPP formation in cyanobacteria before it is possible to infer the

pathway in which geosmin is produced. It seems probable that homologous genes from the MEP pathway will be discovered in more species of cyanobacteria as research continues however it does not necessarily indicate that this pathway is utilized. Only a few organisms have been examined for the route of IPP to geosmin and the metabolic pathway is largely uncharacterized. As a result, the relationship of geosmin production within and among strains of the Cyanobacteria remains unclear.

Phylogenetic studies of the Cyanobacteria have demonstrated that genetic relationships sometimes conflict with the morphological classification (Gugger and Hoffmann, 2004; Iteman *et al.*, 2002; Lyra *et al.*, 2001). With a lack of isolates of many species and inadequate morphological data of sequenced strains, the classification of the Cyanobacteria is by far incomplete. Komárek and Anagnostidis (1989) estimated that more than 50% of the strains in culture collections are misidentified, therefore it is important that new isolates are studied with morphological and genetic approaches (Komárek and Anagnostidis, 1989). The objectives of this study were to examine the phylogenetic relationships of 25 morphologically identified cyanobacteria isolates (Chapter 2) and two additional isolates not studied in chapter 2 using the 16S rRNA gene with particular focus on the relationships of geosmin-producing strains in comparison to non-producing strains. Twenty-three isolates, six of which were geosmin-producing in culture, were sequenced for the 16S rRNA gene and phylogenetic analyses of these in addition to 38 other species within GenBank was performed using distance and maximum likelihood methods. The aim of this study was two-fold: 1) to compare the resulting phylogenies with current taxonomies established by both the bacteriological classification system as well as the system described by Komárek and Anagnostidis and 2) to determine if the ability to

produce geosmin by certain cyanobacteria has been lost or acquired many times or once over the course of time.

4.2 MATERIALS AND METHODS

4.2.1 DNA ISOLATION, PCR AMPLIFICATION, AND SEQUENCE ANALYSIS OF THE 16S rRNA GENE FROM CYANOBACTERIA ISOLATES

Twenty-five isolates of cyanobacteria from the Nostocales, Oscillatoriales, and Chroococcales (Chapter 2, Table 1) were identified based on morphological characteristics. Two geosmin non-producing isolates, *An. oscillaroides* and *An. compacta*, provided by Dr. Freiderich Jüttner of the University of Zurich, Switzerland, were added to those described in Chapter 2, Table 1. DNA was isolated and the 16S rRNA gene (approximately 1400 bp) was PCR amplified and sequenced from all isolates for phylogenetic analysis as in Chapter 2. The 25 isolate sequences were aligned based on primary DNA structure with 46 different cyanobacterial 16S rRNA partial gene sequences and two bacterial outgroup sequences of *Escherichia coli* and *Bacillus subtilis* obtained from Genbank (www.ncbi.nih.gov) using the program Muscle version 3.6 (Edgar, 2004). The alignment was manually edited using the program BioEdit version 7.0.4 (www.mbio.ncsu.edu) by removing the first ~200 nucleotides and last ~200 nucleotides to make all sequences the same length. Identical sequences were treated as a single entity in all phylogenetic analyses. The neighbour-joining (NJ) phylogenetic analysis was carried out using the program PAUP* version 4.0 beta 10 (Swofford, 2001). Maximum likelihood (ML) analysis was carried out using the program PHYML ver. 2.4.4 (Guindon and Gascuel, 2003). The Modeltest program version 3.06 (Posada and Crandall, 1998) was used to examine 56 possible models of DNA substitution and to identify the model that best fit the data set. The Modeltest program selected the GTR+I+G model (Rodriguez *et al.*, 1990) with a proportion of invariable sites (pinvar=0.4420) and the gamma distribution shape

parameter ($a = 0.4915$). The tree was inferred in NJ using this model and 1000 bootstrap

replicates were also performed using the same model. The ML tree was inferred using

PHYML ver. 2.4.4 (Guindon and Gascuel, 2003) using the GTR+I+G model selected by

Modeltest. Bootstrap replicates were performed (1000) for the ML tree. PHYML ver. 2.4.4

(Guindon and Gascuel, 2003) is a simple, fast and accurate algorithm used to estimate large

phylogenies and was used in place of PAUP for the ML analysis due to the large number of

sequences (73) and immense computational time required with PAUP. For both trees,

Escherichia coli K12 and *Bacillus subtilis* were used as outgroup sequences and the trees

were rooted on the branch length leading to these taxa. All bootstrap replicates are

displayed as a percentage and only values greater than 60% are shown. The trees were

viewed using TreeView version 1.6.6 and edited using Adobe Illustrator version 9.0. The

NJ tree is depicted in Figure 19 and the ML tree in Figure 20.

RESULTS

4.3.1 PHYLOGENETIC RELATIONSHIPS

The Modeltest program selected the GTR+I+G model (Rodriguez *et al.*, 1990) with a proportion of invariable sites (pinvar=0.4420) and the gamma distribution shape parameter (*a* = 0.4915). Maximum likelihood determined nucleotide frequencies were (A = 0.25156, C =0.22117, G = 0.31801, T = 0.20927), a six parameter rate matrix: R(a) [A-C] = 0.87061, R(b) [A-G] = 2.12488, R(c) [A-T] = 1.40090, R(d) [C-G] = 0.49075, R(e) [C-T] = 3.53041, R(f) [G-T] = 1.0 (fixed). These parameters were used in the neighbour-joining and maximum likelihood methods for inferring phylogenetic trees (Figures 19 and 20). The 16S rRNA gene was amplified and sequenced for 25 isolates of cyanobacteria. Identical sequences were considered as 0.999-0.998% similar and were observed in three *An. lemmermannii* isolates, LONT 5, AL4 and AL7. Hence, the former two were removed from the analyses. In both trees, the Nostocales and Stigonematales orders, corresponding to Sections IV and V respectively, separated from the Chroococcales, Oscillatoriales, and Prochlorales into two principal well supported clades [NJ, 93% bootstrap support; ML, 92% bootstrap support]. Within the Nostocales, the Rivulariaceae and Nostocaceae families were intermixed. One clade was formed solely of species of *Anabaena* and *Aphanizomenon* and had high bootstrap support [NJ, 94%; ML 93%]. Within this clade, all of the *An.. lemmermannii* isolates sequenced in this study (both geosmin-producing and non-producing) grouped into a subclade for which there was a poor NJ bootstrap support of 65% and no support in the ML tree. Divergence was low and ranged from 0.27-2.60% for this subclade. In addition, *An. planktonica*, *Ap. flos-aquae* (Hamilton Harbour), *An. spiroides*, *An. compacta*, *An* sp. LOW, and *Ap. gracile* (geosmin-producer)*,* all of which

163

were sequenced in this study, grouped within the *Anabaena/Aphanizomenon* subclade in both trees. Divergence within the *Anabaena* and *Apanizomenon* clade was low, ranging between 0.1-6.7%. In both trees, three small subclades with little divergence formed and were comprised exclusively of isolates sequenced in this study. The first included *An. spiroides* and *An. compacta* with high bootstrap support [NJ, 100%; ML, 100%]; the second included both *An. flos-aquae* strains UTEX LB 2383 and UTCC 64 as well as *An. oscillaroides* and was well supported [NJ, 94%; ML, 90% bootstrap support]; and the third comprised *An. cylindrica*, *An. variabilis*, and *An. viguieri* for which there was poor support in the NJ tree (67%) and none in the ML tree. *Calothrix* sp. PCC 7507 (geosmin-producer) was basal to the subclade containing both *An. flos-aquae* strains, *An. oscillaroides*, along with *Anabaenopsis circularis* and *Cylindrospermum* sp. PCC 7417 and did not group with other *Calothrix* species (no bootstrap support). Although the Rivulariaceae were polyphyletic, a subclade of three species of *Calothrix* formed in the NJ tree that included only strains of *Calothrix*; *Calothrix* sp. PCC 7714, *Calothrix desertica*, and *Calothrix parietina* PCC 6303 (sequenced in this study) and had a strong bootstrap support of 81%. A similar clade formed in the ML tree but also included *Calothrix* sp. CCMEE5093 and had strong bootstrap support of 81% as well. Members of the Rivulariaceae spanned a single genus, *Calothrix*, and were polyphyletic among four subclades. Divergence was very high for this genus and ranged from 4.7-15.6%. A subclade of all the Stigonematales and members of Section V formed within the Nostocales and had poor bootstrap support in the NJ tree (60%) but strong support in the ML tree (83%). The families within the Stigonematales were polyphyletic with *Hapalosiphon welwitschii* and *Chlorogleopsis* sp.

164

Figure 19. Neighbour-joining tree of 71 cyanobacterial species and strains, constructed using 16S rRNA gene sequences (1203 bp). Numbers at nodes represent bootstrap percentages of 1000 replicates; only values above 60% are shown. Taxa in bold were isolates used in this study. The asterisk denotes geosmin-producing strains in culture.

Figure 20. PHYML Maximum-likelihood phylogenetic tree of 71 cyanobacterial species and strains, constructed using 16S rRNA gene seuqences (1203 bp). Numbers at nodes represent percentages of 1000 bootstrap replicates; only values above 60% are shown. Taxa in bold were isolates used in this study. The asterisk denotes geosmin-producing strains in culture.

64 gi51105635 *Anabaena viguieri* strain TAC433
Anabaena planktonica
gi16945135 *Anabaena spiroides* strain PMC9702
gi55162837 *Anabaena smithii* 1tu39s8
84 gi51105617 *Anabaena affinis* strain NIES40
gi55162836 *Anabaena sigmoidea* 0tu38s4
81 Aphanizomenon flos-aquae (Hamilton Harbour)
Aphanizomenon gracile*
Anabaena sp (Lake of the Woods)
100 gi38045370 *Aphanizomenon gracile* strain LMECYA40
Anabaena compacta
Anabaena spiroides
82 79 **Anabaena lemmermannii** LONT5*
Anabaena lemmermannii AL5
Anabaena lemmermannii LONT2*
93 **Anabaena lemmermannii** GIOL8*
gi55162840 *Anabaena sp* 0tu39s7
gi16945132 *Anabaena lemmermannii* strain 256.
96 gi55162847 *Aphanizomenon issatschenkoi* 0tu37s7
Anabaena vigueri
Trichormus variabilis
83 **Anabaena cylindrica**
gi8571953 *Anabaena bergii*
gi10505043 *Nodularia harveyana*
99 **Anabaena flos-aquae UTEX LB2383**
84 **Anabaena flos-aquae UTCC64**
90 **Anabaena oscillaroides**
gi8896061 *Anabaenopsis circularis* NIES21
gi6522669 *Cylindrospermum sp* PCC7417
Calothrix sp PCC7507*
100 gi51105634 *Anabaena kisseleviana* strain TAC34
gi3659905 *Cylindrospermopsis raciborskii*
gi55162829 *Anabaena oscillarioides* BO HINDAK 1984 43
gi23978189 *Calothrix brevissima*
65 gi14625359 *Fischerella muscicola*
gi14517316 *Hapalosiphon welwitschii*
gi49169 *Chlorogloeopsis sp*
92 83 100 gi89242004 *Calothrix desertica*
gi6522670 *Calothrix sp* PCC7714
81 **Calothrix parietina PCC6303**
gi24850289 *Calothrix sp* CCMEE5093
100 gi6522866 *Oscillatoria sp* CYA128R
Planktothrix aghardii
95 gi3242118 *Halothece sp*
gi3242140 *Dactylococcopsis sp*
78 gi2760616 *Spirulina subsalsa*
gi3242112 *Cyanothece sp*
gi1109755 *Gloeothece membranacea*
gi1684704 *Merismopedia glauca*
gi2760614 *Oscillatoria rosea*
100 **Microcystis aeruginosa**
Microcystis wesenbergii
gi45800 *Prochloron sp*
gi47716884 *Lyngbya majuscula* HECT
100 gi47933466 *Oscillatoria spongeliae* 32P1
gi1289269 *Oscillatoria cf corallinae*
100 gi3559794 *Trichodesmium tenue*
gi14625371 *Oscillatoria sancta*
gi3242226 *Lyngbya sp*
Oscillatoria limosa*
gi14625370 *Oscillatoria acuminata*
94 gi2760615 *Phormidium mucicola*
90 gi14625372 *Pseudanabaena* PCC6802
100 **Pseudanabaena sp**
gi3336898 *Oscillatoria limnetica*
gi2760617 *Phormidium ambiguum*
gi22759843 *Oscillatoria princeps* NIVACYA
gi12004674 *Pseudanabaena tremula* UTCC471
gi1289224 *Leptolyngbya boryanum*
gi3336915 *Prochlorothrix hollandica*
gi45598 *Prochlorococcus marinus*
100 gi100222932 *Bacillus subtilis*
gi49175990 *Escherichia coli* K12

Nostocales
(Nostocaceae)

Nostocales (Riv.)
Nostocales
(Nostocaceae)
Nostocales (Riv.)
Stigonematales
Nostocales (Riv.)

Oscillatoriales
Chroococcales
Oscillatoriales
Chroococcales
Oscillatoriales
Chroococcales
Prochlorales (Prochloraceae)

Oscillatoriales

Prochlorales (Prochlorococcaceae)
Bacteria
Outgroup

IV

III
N/C
I
III
I
N/C
III
I

III

N/C
III
N/C
III
I

0.1

of the Mastigocladaceae grouping with *Fischerella mucicola* of the Fischerellaceae and divergence among them was moderate 8.6-16.0%. When examining the tree based on the bacteriological Code, only one family exists in Section V (analogous to the Stigonematales) and includes *Chlorogloeopsis* sp. and *Fischerella mucicola*. *Hapalosiphon welwitschii,* a species that has not officially been validated in this subsection but is considered a member of the Stigonematales, was included in the subclade as well.

The Oscillatoriales, Chroococcales, and Prochlorales, analogous to Sections I and III (Bacteriological Code) formed a separate but polyphyletic clade from the Nostocales and Stigonematales (Sections IV and V) with strong bootstrap support [NJ, 93%; ML, 92%]. *Spirulina subsalsa* of the Oscillatoriales grouped with an almost exclusive subclade of the Chroococcales (and Section I) members in both trees for which there was no bootstrap support. Included in this clade of the NJ tree were *Halothece* sp., *Dactylococcopsis* sp., *Cyanothece* sp., *Gloeothece membranaceae* (Chroococcales/Section I), and *Spirulina subsalsa* (Oscillatoriales/Section III) as well as *Prochloron* sp., a member of Section I under the Bacteriological Code and of the Prochlorales (Botanical Classification) grouping basally. *Prochloron* sp. grouped differently in the ML tree and was basal to another subclade containing mostly Chroococcales (and Section I) members. Two smaller subclades comprised exclusively of the Oscillatoriales/Section III taxa formed groups in both trees however no bootstrap support was available for these particular groups. A small subclade of the Prochlorales that included *Prochlorococcus marinus* and *Prochlorothrix hollandica* formed in the NJ tree for which there was a poor bootstrap support of 66%. In the ML tree these species did not group together and were basal to the Oscillatoriales/Chroococcales/Prochlorales (Sections I and III) however *Prochlororhrix*

hollandica was more closely related to Section III members for which it has been classified. Divergence among the Prochlorales was very high, ranging between 16.9-28.0%. Of the isolates sequenced in this study, those of the Oscillatoriales, *Osc. limosa* (a geosmin-producer) and *Planktothrix aghardii*, grouped within a small subclade of this order and Section III in the NJ tree. *Osc. aghardii* was on the same branch as the GenBank sequence of *Oscillatoria* sp. CYA128R with a high bootstrap support of 100%. *Osc. limosa* grouped with *Osc. princeps* and *Phormidium ambiguum* within this small subclade for which there was no bootstrap support. In the ML tree, *Pl. aghardii* formed a very small group with *Oscillatoria* sp. CYA128R with very high bootstrap support (100%) that was distant from *Osc. limosa*, who grouped with *Trichodesmium tenue*, *Osc. sancta*, and *Lyngbya* sp. in a small clade for which there was no bootstrap support. *Osc. limosa* grouped in a small subclade for both trees but with different members of the Oscillatoriales (Section III). In the NJ tree, *Microcystis aeruginosa* sequenced in this study grouped in a small subclade with *Merismopedia glauca*, another Chroococcales and member of Section I. *Microcystis wesenbergii* however grouped basal to the entire Oscillatoriales/Chroococcales/Prochlorales (Section I and III) clade for which there was also no bootstrap support. In the ML tree however, *M. aeruginosa* and *M. wesenbergii* of the Chroococcales (Section I) grouped together with high bootstrap support (100%). Together they formed a small clade with *Merismopedia glauca* (Chroococcales, not classified under the Bacteriological Code) and *Oscillatoria rosea* (Oscillatoriales/Section III). Divergence within the Chroococcales was broadly ranged from 0.7-23.0% and similarly within the Oscillatoriales, 0.5-29.3%, 5.0-9.7% more divergence than that

exhibited by the Nostocales (Section IV) and 7.0-12.7% more than the Stigonematales

(Section V).

DISCUSSION

4.4.1 HOW WAS THE GEOSMIN TRAIT ACQUIRED IN CYANOBACTERIA?

According to the results of both phylogenetic trees, the 16S rRNA gene does not distinguish geosmin-producing strains of cyanobacteria from one another. The Lake Ontario *An. lemmermannii* strains were intermixed in a single subclade with a small divergence range of 0.27-2.6%, indicating that geosmin-producing strains identified as the same species and collected from the same source were no more closely related than those of different sources. This supports previous studies that have shown geosmin production is strain specific rather than species specific (Watson, 2004) and provides evidence for the possibility that the genes involved in geosmin biosynthesis could be found on a plasmid. The geosmin-producing isolates of the Nostocales exhibited a small divergence ranging from 0.4-7% while *Osc. limosa* within the geosmin-producers was most divergent (between 12.8-16.9%). The divergence and grouping of geosmin-producing cyanobacteria with geosmin non-producing strains suggests that geosmin production did not evolve from a single ancestor within the Cyanobacteria. The genes responsible for the conversion of intermediates in the terpenoid pathways may have evolved in the Cyanobacteria through lateral gene transfer from an ancestral non-photosynthetic bacterium such as those of the geosmin-producing actinomycetes or via convergent evolution of genes that were acquired separately in the Oscillatoriales and Nostocales. The most parsimonious explanation would be that the gene(s) for geosmin biosynthesis were acquired once in the Cyanobacteria and subsequently lost in various species. Those species that do not exhibit geosmin production could possibly possess the gene(s) necessary for its production but do not express them. Other markers may be more informative when determining the origins and relationships of

geosmin-producing cyanobacteria rather than the 16S rRNA gene. The internal transcribed spacer (ITS) (Gugger *et al.*, 2002; Rocap *et al.*, 2002), RNA polymerase C1 (rpoC1)(Seo and Yokota, 2003), or Rubisco (rbcLX) (Gugger *et al.*, 2002) have proven to be of effective in studies of toxin-producing cyanobacteria such as those that produce anatoxin-a or hepatotoxins (Gugger *et al.*, 2002) and may also be useful in determining relationships among geosmin-producing strains of cyanobacteria. Further research into the terpenoid metabolic pathways leading to geosmin will hopefully uncover a gene responsible for its production, one that may convert the final intermediate into the odourous metabolite. If such a gene is uncovered, it may be applied in assays to directly identify geosmin-producing strains by the absence or presence of the gene in a similar manner towards the identification of microcystin-producing cyanobacteria (Fergusson and Saint, 2000; Hisbergues *et al.*, 2003). Furthermore, the *isp* genes involved in isoprenoid production in *Synechocystis* sp. may also be present in other species of cyanobacteria and could show more resolved relationships among geosmin-producing strains.

4.4.2 A COMPARISON OF CYANOBACTERIAL TAXONOMY

The phylogenies of the 71 species of cyanobacteria used in this study has shown and supported many of the ambiguities seen in the current classification. The subclades within the Nostocales (Section IV) were for the most part very well supported however they were not monophyletic for their genera. Many studies have shown that the Nostocales (Section IV) largely group together however members of the Stigonematales (Section V) have also been included in a subclade of this order (Gugger and Hoffmann, 2004) as was seen in the NJ and ML trees in this study. One commonality between these two groups is that they are the only heterocyst-forming cyanobacteria and grouping was monophyletic based on this

trait. The primary morphological difference between the Stigonematales (Section V) and the Nostocales (Section IV) is that the former exhibits true branching while the latter does not (Anagnostidis and Komárek, 1990). This characteristic does not correspond to genetic differences among the heterocyst-forming cyanobacteria and in many other studies these groups have been intemixed; however the Stigonematales (Section V) tend to form their own subclades and the respective species of the groups do not appear to intermix with the Nostocales (Gugger and Hoffmann, 2004; Rajaniemi *et al.*, 2005). Divergence within the heterocyst-forming clade ranged from 0.1% to 23%, with the most divergence seen between members of the Nostocales (Section IV) and Stigonematales (Section V), particularly between *Calothrix desertica* and *Dactylococcopsis* sp. Studies have shown that genera within the Nostocales such as *Anabaena*, *Aphanizomenon*, *Anabaenopsis*, *Cylindrospermopsis*, and *Nodularia* are polyphyletic based on the 16S rRNA gene (Iteman *et al.*, 2002; Rajaniemi *et al.*, 2005; Svenning *et al.*, 2005), and as seen in the NJ and ML trees in this study they did not group monophyletically. The topologies seen in both trees of this study are congruent with those of other studies and suggest that the taxonomy of the heterocystous cyanobacteria is not completely resolved and that higher taxonomic levels (orders and sections) appear to be more resolved than those below (family, genus, species).

The non-heterocystous cyanobacteria comprise the Chroococcales (Section I), Section II (formerly named the Pleurocapsales order), and the Oscillatoriales (Section III). In this study the relationships of the Chroococcales (Section I) to the Oscillatoriales (Section III) were shown to be polyphyletic based on the 16S rRNA gene. Other studies have shown that the classification of these two groups is not resolved based on phylogenies of the 16S rRNA gene (Casamatta *et al.*, 2005; Ishida *et al.*, 1997; Litvaitis, 2002) and as

seen in this study none of the groups exhibited monophyly with bootstrap support. Divergence among the Chroococcales (Section I) ranged from 0.7%-28.0% and for the Oscillatoriales 0.5-29.3% indicating that there is a greater evolutionary distance among the members of these groups in comparison to those of the Nostocales (0.1-19.0% divergence) and the Stigonematales (0.9-16.0% divergence). The greater divergence and lack of bootstrap support for the Chroococcales (Section I) and Oscillatoriales (Section III) in both trees suggests that taxonomy for these groups needs revision. *Prochloron* sp. and *Prochlorococcus marinus* were former members of the Prochlorales and in the past decade were re-classified into the Section I, Family 1.1 under the Bacteriological Code (Pinevich *et al.*, 1997). These two genera exhibited the greatest divergence (28%) within the Chroococcales in this study and interestingly *Prochlorothrix hollandica* (Section III) exhibited less divergence to *Prochlorococcus marinus*. The grouping of *Prochloron* sp. with several members of the Chroococcales in both trees suggests that the reclassification of this genus was improved, however the grouping of *Prochlorothrix hollandica* and *Prochlorococcus marinus* of differing subsections in both trees suggests that they are more closely related to one another and that the taxonomy may still require revision.

4.5 CONCLUSIONS

The biochemistry and molecular biology of geosmin production remains unclear. Production of this metabolite in drinking water sources such as Lake Ontario has caused concern for the quality of tap water in Ontario even though geosmin is not a toxic compound. The phylogenetic analyses carried out in this study on six geosmin-producing isolates of cyanobacteria have given some insight into the evolutionary origins of this trait. Since the geosmin-producing strains did not form a monophyletic group, it was not possible to state a single evolutionary event has led to the acquisition of the gene(s) involved in its metabolism. It is therefore possible that these genes were acquired from an ancestral bacterial species (perhaps of the actinomycetes) and have been lost throughout the course of time. Because this trait is not species specific, it is possible that members of the same species possess the genes for geosmin biosynthesis but they are repressed throughout the life cycle of certain strains. Reasons for a suppressed gene expression are unknown as are the origins of the genes involved. Future phylogenetic studies using other genes such as ITS, rpoC1, phycocyanin, and the *isp* genes may be more informative in determining the origins and relationships among geosmin-producing cyanobacteria. Additionally, it is possible that the genes for geosmin-production are located outside of the nuclear genome, perhaps on a plasmid, and become transferred to certain strains. Future studies should also investigate the possibility that geosmin is not nuclear encoded to determine the genes that directly result in the odorous compound for phylogenetic and other molecular studies.

Since the first extensive accounts by Bornet & Flahault [1886-88] and Gomont [1829-93], cyanobacterial taxonomy has undergone extensive revisions, with the most widely applied classification based on Castenholz & Waterbury (1989) under the

Bacteriological Code and of Komárek & Anagnostidis (1985, 1986, 1988, 1989, 1990), a modern approach (John *et al.*, 2002). The results of this study indicate none of the orders (Chroococcales, Oscillatoriales, Nostocales and Stigonematales) and subsections (I, III, IV, and V) are monophyletic. The Nostocales (Section IV) appears to be the most resolved of the groups with low divergence and a near-monophyletic grouping. Within this order/subsection however the genera are not resolved, as is the case with all of the studied taxa. The heterocystous cyanobacteria were distinct from the non-heterocystous and showed less divergence and polyphyly in comparison to the non-heterocystous cyanobacteria. It is apparent that the non-heterocystous cyanobacteria require a greater amount of revision to the classification owing to the intermixing of the orders/subsections and greater divergence among taxa. Although both systems of classification used in this study are similar, nomenclature and groupings are occasionally different among the groups. As a result, it demonstrates the incongruity between bacteriologists and botanists and emphasizes the need for a consensus system of classification for the Cyanobacteria.

4.6 REFERENCES

Anagnostidis, K. and J. Komárek (1988). Modern approach to the classification system of Cyanophytes. 3. Oscillatoriales. Archiv fur Hydrobiologie Supplement 80: 327-472.

Anagnostidis, K. and J. Komárek (1990). Modern Approach to the Classification of the Cyanophytes 5 - Stigonematales. Algological Studies 59: 1-73.

Burgerwiersma, T., L. J. Stal and L. R. Mur (1989). *Prochlorothrix hollandica* gen-nov, sp-nov, a filamentous oxygenic photoautotrophic Prokaryote containing chlorophyll-a and chlorophyll-b - Assignment to Prochlorotrichaceae fam-nov and order Prochlorales Florenzano, Balloni, and Materassi 1986, with emendation of the ordinal description. International Journal of Systematic Bacteriology 39(3): 250-257.

Casamatta, D. A., J. R. Johansen, M. L. Vis and S. T. Broadwater (2005). Molecular and morphological characterization of ten polar and near-polar strains within the Oscillatoriales (Cyanobacteria). Journal of Phycology 41: 421-438.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32(5): 1792-1797.

Ershov, Y. V., R. R. Gantt, J. Francis X. Cunningham and E. Gantt (2002). Isoprenoid Biosynthesis in *Synechocystis* sp. Strain PCC6803 Is Stimulated by Compounds of the Pentose Phosphate Cycle but Not by Pyruvate or Deoxyxylulose-5-Phosphate. Journal of Bacteriology 184(18): 5045-5051.

Felsenstein, J. (2004). Inferring Phylogenies. Sunderland, Massachusetts, Sinauer Associates, Inc.

Fergusson, K. M. and C. P. Saint (2000). Molecular Phylogeny of *Anabaena circinalis* and Its Identification in Environmental Samples by PCR. Applied and Environmental Microbiology 66(9): 4145-4148.

Freeman, S. and J. C. Herron (2001). Evolutionary Analysis. Upper Saddle River, Prentice Hall.

Garrity, G. M., J. A. Bell and T. G. Lilburn (2004). Taxonomic Outline of the Prokaryotes. New York, Springer-Verlag.

Graham, L. E. and L. W. Wilcox (2000). Algae. Upper Saddle River, Prentice-Hall.

Graur, D. and W. H. Li (2000). Fundamentals of Molecular Evolution. Sunderland, MA, Sinauer Associates, Inc.

Gugger, M., C. Lyra, P. Henriksen, A. Coute, J.-F. Humbert and K. Sivonen (2002). Phylogenetic comparison of the cyanobacterial genera *Anabaena* and *Aphanizomenon*. International Journal of Systematic and Evolutionary Microbiology 52: 1867-1880.

Gugger, M. F. and L. Hoffmann (2004). Polyphyly of true branching cyanobacteria (Stigonematales). International Journal of Systematic and Evolutionary Microbiology 54: 349-357.

Guindon, S. and O. Gascuel (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology. 52(5): 696-704.

Hisbergues, M., G. Christiansen, L. Rouhiainen, K. Sivonen and T. Borner (2003). PCR-based identification of microcystin-producing genotypes of different cyanobacterial genera. Archives of Microbiology 180: 402-410.

Ishida, T., A. Yokota and J. Sugiyama (1997). Phylogenetic relationships of filamentous cyanobacterial taxa inferred from 16S rRNA sequence divergence. Journal of Applied Microbiology 43: 237-241.

Iteman, I., R. Rippka, N. T. d. Marsac and M. Herdman (2002). rDNA analyses of planktonic heterocystous cyanobacteria, including members of the genera *Anabaenopsis* and *Cyanospira*. Microbiology 148(481-496).

John, D. M., B. A. Whitton and A. J. Brook, Eds. (2002). The Freshwater Algal Flora of the British Isles. Cambridge, The press syndicate of the University of Cambridge.

Jukes, T. H. and C. R. Cantor (1969). Evolution of Protein Molecules. New York, Academic Press.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111-120.

Komárek, J. and K. Anagnostidis (1986). Modern Approach to the Classification System of Cyanophytes 2. Chroococcales. Archiv fur Hydrobiologie Supplement 73: 157-226.

Komárek, J. and K. Anagnostidis (1989). Modern Approach to the Classification System of Cyanophytes 4. Nostocales. Archiv fur Hydrobiologie Supplement 82: 247-345.

Lange, B. M., T. Rujan, W. Martin and R. Croteau (2000). Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. Proceedings of the National Academy of Sciences 97(24): 13172-13177.

Lewin, R. (1977). *Prochloron*, type genus of the Prochlorophyta. Phycologia 1: 217.

Lewin, R. A. and L. Cheng, Eds. (1989). *Prochloron*, a microbial enigma. New York, Chapman & Hall.

Litvaitis, M. K. (2002). A molecular test of cyanobacterial phylogeny: inferences from constraint analyses. Hydrobiologa 468: 135-145.

Lyra, C., S. Suomalainen, M. Gugger, C. Vezie, P. Sundman, L. Paulin and K. Sivonen (2001). Molecular characterization of planktic cyanobacteria of *Anabaena, Aphanizomenon, Microcystis,* and *Planktothrix* genera. International Journal of Systematic and Evolutionary Microbiology 51: 513-526.

Nei, M. and S. Kumar (2000). Molecular Evolution and Phylogenetics. New York, Oxford University Press.

Palenik, B. and R. Haselkorn (1992). Multiple evolutionary origins of prochlorophytes, the chlorophyll *b*-containing prokaryotes. Nature 355: 265-267.

Palenik, B. and H. Swift (1996). Cyanobacterial Evolution and Prochlorophyte Diversity as seen in DNA-Dependent RNA Polymerase Gene Sequences. Journal of Phycology 32: 638-646.

Pinevich, A. V., S. G. Averina and N. V. Velichko (1997). Another view on the role of photosynthetic pigments in the taxonomy of oxygenic-phototrophic bacteria: proposed rejection of the order Prochlorales Florenzano, Balloni, and Materassi 1986 (Emand. Burger-Wiersma, Stal, and Mur 1989), the family Prochloraceae Florenzano, Balloni and Materassi 1986, and the family Prochlorotrichaceae Burger-Wiersma, Stal, and Mur 1989. International Journal of Systematic Bacteriology 47(4): 1264-1267.

Posada, D. and K. A. Crandall (1998). Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817-818.

Rajaniemi, P., P. Hrouzek, K. Kastovska, R. Williame, A. Rantala, L. Hoffman, J. Komarek and K. Sivonen (2005). Phylogenetic and morphological evaluation of the genera *Anabaena*, *Aphanizomenon*, *Trichormus* and *Nostoc* (Nostocales, Cyanobacteria). International Journal of Systematic and Evolutionary Microbiology 55: 11-26.

Rippka, R., J. Deruelles, J. B. Waterbury, M. Herdman and R. Y. Stanier (1979). Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. Journal of General Microbiology 111: 1-61.

Rocap, G., D. L. Distel, J. B. Waterbury and S. W. Chisholm (2002). Resolution of *Prochlorococcus* and *Synechococcus* Ecotypes by Using 16S-23S Ribosomal DNA Internal Transcribed Spacer Sequences. Applied and Environmental Microbiology 68(3): 1180-1191.

Rodriguez, F., J. F. Oliver, A. Marin and J. R. Medina (1990). The general stochastic model of nucleotide susbstitution. Journal of Theoretical Biology 142: 485-501.

Seo, P.-S. and A. Yokota (2003). The phylogenetic relationships of cyanobacteria inferred from 16S rRNA, *gyr*B, *rpo*C1, and *rpo*D1. Journal of General Applied Microbiology 49: 191-203.

Svenning, M. M., T. Eriksson and U. Rasmussen (2005). Phylogeny of symbiotic cyanobacteria within the genus *Nostoc* based on 16S rDNA sequence analyses. Archives of Microbiology 183: 19-26.

Swofford, D. L. (2001). PAUP* Phylogenetic analysis using parsimony and other methods, Sinauer.

Urbach, E., D. L. Robertson and S. W. Chisholm (1992). Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. Nature 355: 267-269.

Watson, S. B. (2004). Aquatic taste and odour: a primary signal of drinking water integrity. Journal of Toxicological and Environmental Health, Part A 67: 1779-1795.

CHAPTER 5

SUMMARY

The objectives of this study were as follows:

1. **Delineate various known geosmin-producing and non-geosmin-producing cyanobacterial isolates to the species level using the 16S rRNA gene and morphological characteristics**
2. **Develop a DGGE standard marker with isolates from the previous objective using a variable region of the 16S rRNA gene**
3. **Compare the DGGE standard marker to environmental samples collected from Lake Ontario in late August early September 2005 to identify the presence of possible geosmin producing cyanobacteria.**
4. **Determine the phylogenetic relationships among all isolates of geosmin-producing and non-producing cyanobacteria using 16S rRNA gene sequences to determine if geosmin production may have evolved from a common ancestor and compare these phylogenies with current taxonomy**

This study has provided insight into the genetic and evolutionary aspects of geosmin-producing and non-producing cyanobacteria. A molecular approach to the identification of geosmin-producing cyanobacteria using DGGE is promising but will require future development. The 16S rRNA-V3 region used in the DGGE standard marker cannot resolve to the species level and does not distinguish geosmin-producing and non-producing strains of *An. lemmermannii.* As a result, this region should not be used alone to infer identities of geosmin-producing species/strains in environmental samples. However with additional data to support the presence of a geosmin-producing strain at the sampling site in the DGGE profile, samples with *An. lemmermannii* bands were consistent with

182

locations in which this species is common. Presence of a potential geosmin-producing species band in DGGE did not correlate statistically to an increase in geosmin concentrations at the sampling site, however they did correspond with regions where *An. lemmermannii* is common. Additionally, average geosmin concentrations were higher at sites with *An. lemmermannii* in the DGGE profile than those without the presence of this species. These results indicate that using the 16S rRNA-V3 region in DGGE alone may not be useful for identifying geosmin-producing species/strains, however with additional data can be increasingly accurate. In the future, other genetic markers should be investigated for a more accurate identification of geosmin-producing cyanobacteria such as a combination of the 16S rRNA-V3 and V5 regions, RNA polymerase C1 (*rpo*C1), the subunit B protein of DNA gyrase (*gyr*B), Rubisco (*rbc*LX), and the *isp* genes from the MEP pathway. The ITS-1 region examined in this study was not useful for identifying individual species from environmental samples due to the high variability and multiple amplicons among strains.

The phylogenetic analyses have given some insight into the evolutionary origins of the trait of geosmin production. Since geosmin-producing strains did not group monophyletically, it was not possible to state that a single evolutionary event has led to the acquisition of the gene(s) involved in its metabolism. It is therefore likely that these genes were acquired from an ancestral bacterial species (perhaps of the actinomycetes) and have been lost throughout the course of time. Because this trait is not species specific, it is possible that members of the same species possess the genes required for enzymes involved in geosmin biosynthesis but they may be repressed throughout the life cycle of certain strains. The mechanisms behind expression of these genes are unknown as are the origins of the genes involved however it seems possible that the genes involved in geosmin

biosynthesis may be plasmid encoded rather than nuclear encoded. Future phylogenetic studies using other genes such as ITS, rpoC1, *gyr*B, Rubisco, and *isp* may be more informative in determining the origins and relationships among geosmin-producing cyanobacteria. Investigations into potential plasmid encoded genes involved in geosmin biosynthesis may also lead to the elucidation of the metabolic pathway(s) and may involve genes useful in determining the relationships among these strains of cyanobacteria.

The large phylogenetic analyses carried out in this study have shown that the taxonomy for this class of prokaryotes requires a large amount of revision. None of the orders (Chroococcales, Oscillatoriales, Nostocales and Stigonematales) or sections (I, III, IV, and V) were monophyletic. The Nostocales (Subsection IV) appeared to be the most resolved of the groups with low divergence and a near-monophyletic grouping. Within this order/section however the genera were not resolved, as is the case with all of the studied taxa. The heterocystous cyanobacteria were distinct from the non-heterocystous and showed less divergence and polyphyly in comparison to the non-heterocystous cyanobacteria. It is apparent that the non-heterocystous cyanobacteria require a greater amount of revision to the classification owing to the intermixing of the orders/subsections and greater divergence among taxa. Although both systems of classification used in this study are similar, nomenclature and groupings are occasionally different among the groups. As a result, it demonstrates the incongruity between bacteriologists and phycologists and emphasizes the need for a consensus system of classification for the Cyanobacteria.

To date, studies have focused on the biochemical and ecological characteristics of geosmin-producing cyanobacteria to gain an understanding of geosmin biosynthesis as well as the physiological and environmental factors that lead to its production. DGGE studies

have been used to examine and monitor cyanobacterial communities under different ecological conditions but not during peak geosmin concentrations. Additionally, the phylogenetic relationships of geosmin-producing cyanobacteria and the evolutionary origins of geosmin-production have not been inferred. Taste and odour episodes in Lake Ontario and the world have not been proactively managed due to the lack of knowledge into the causes of geosmin production resulting in the inability to predict episodes. This study has introduced DGGE as a novel approach to monitoring cyanobacterial communities in a water body to predict and prepare for taste and odour episodes caused by geosmin. The evolutionary origins of geosmin-production are not well characterized however the data presented in this study has raised a number of questions to be addressed. With an increased knowledge into the genetic and evolutionary aspects of geosmin-producing cyanobacteria in addition to biochemical and ecological data, the future management of taste and odour episodes may turn proactive resulting in consumer confidence and satisfaction in Lake Ontario drinking water quality and drinking water around the world.

Appendix A. Primers used in the amplification and sequencing of the 16S rRNA gene and the 16S rRNA-V3 region in cyanobacteria isolates

Primer Name	Direction Forward/Reverse	Sequence 5'-3'	Number of Nucleotides
27F1	Forward	AGAGTTTGATCCTGGCTCAG	20
1494Rc	Reverse	TACGGCTACCTTGTTACGAC	20
400	Forward	AAGGCTCTTGGGTTGTAAACC	21
861	Forward	TAACGCGTTAAGTATCCC	18
521	Forward	GCGTAAAGGGTCCGCAGGT	19
1061	Forward	CGCGAAGAACCTTACCAAG	19
Cya781R (a)	Reverse	GACTACAGGGGTATCTAATCCCTTT	25
Cya781R (b)	Reverse	GACTACTGGGGTATCTAATCCCATT	25
Cya781F (a)	Reverse	AATGGGATTAGATACCCCAGTAGTA	25
Cya781F (b)	Reverse	AAAGGGATTAGATACCCCTGATGTC	25
Cya350F	Forward	GGGGGATTTTCCGCAATGGG	20
Cya359F	Forward	CGCCCGCCGCGCCCCGCGCCGGTCCCGCCGCCCCCGCCCGGGGGAATYTTCCGCAATGGG	60
2000R	Reverse	CTCTGTGTGCCTAGGTATCC	20
cya-b-F371	Forward	CCTACGGGAGGCAGCAGTGGGGAATTTTCCG	31
cya-R783	Reverse	GACTACWGGGGTATCTAATCCCW	23
F357GC	Forward	CGCCCGCCGCGCCCCGCGCCCGGCCCGCCGCCCCCGCCCCCCTACGGGAGGCAGCAG	57
R518	Reverse	ATTACCGCGGCTGCTGG	17