MOLECULAR AND EVOLUTIONARY INVESTIGATIONS INTO THE CYANOTOXIN CYLINDROSPERMOPSIN

Rebecca Jane CAMPBELL BSc (Hons)

Department of Pharmaceutical and Biomedical Sciences

The University of South Australia

(Faculty of Health Science)

August, 2009

A thesis submitted for the degree of Doctor of Philosophy

For Mark and Max

Table of Contents

Abstract i

Declaration iii

Acknowledgements iv

Publications in support of this thesis v

Abbreviations, prefixes and symbols vi

1 GENERAL INTRODUCTION 1

1.1 Cyanobacteria 1

1.1.1 Species Diversity and global distribution 1

1.1.2 Blooms and management 2

1.2 Secondary metabolites 4

1.2.1 Polyketide synthases 4

1.2.2 Non ribosomal peptide synthetases 10

1.2.3 Hybrid polyketide and non ribosomal peptide complexes 12

1.3 Cyanobacterial toxic metabolites 13

1.3.1 Hepatotoxins 14

1.3.2 Neurotoxins 16

1.3.3 Cytotoxic 18

1.4 Current identification strategies 21

1.5 Cyanobacteria Gene clusters 22

1.5.1 Microcystin and Nodularins 25

1.5.2 Cylindrospermopsin 26

1.6 Molecular methods for identification of potentially toxic organisms. 27

1.6.1 Conventional PCR 27

1.6.2 Real-time PCR 29

1.7 Phylogenetic analyses 33

1.8 Aims and objectives 36

2 GENERAL MATERIALS AND METHODS 38

2.1 Reagents and Chemicals 38

2.1.1 Antibiotics 38

2.1.2 Buffers 38
2.2 Culture Media 38
2.3 Cultures 39
2.3.1 Maintenance of cultures 39
2.3.2 Toxin analysis 40
2.4 DNA Extraction Protocols 40
2.4.1 Genomica DNA extraction 40
2.4.2 Plasmid extraction 41
2.4.3 Cosmid extraction 41
2.5 Molecular Techniques 41
2.5.1 Enzymes 41
2.5.2 Oligonucleotides 42
2.5.3 Conventional PCR 42
2.5.4 Long template PCR 42
2.5.5 Agarose gel electrophoresis of DNA 43
2.5.6 Bioanalyser 43
2.6 DNA Manipulation 44
2.6.1 Purification of PCR products 44
2.6.2 *In vitro* cloning 44
2.6.3 Restriction enzyme digestion 45
2.6.4 Ligation 46
2.7 Bioinformatics 46
2.8 Phylogenetic Analyses 46
2.8.1 Sequencing 46
3 SEQUENCING AND STRUCTURAL ANALYSIS OF A GENE CLUSTER IMPLICATED IN THE PRODUCTION OF CYLINDROSPERMOSPIN 48
3.1 Introduction 48
3.2 Materials and Methods 50
3.2.1 Cyanobacteria culture conditions 50
3.2.2 P^{32} colony hybridisation 54

3.2.3 Genomic Walking Methods 56
3.2.4 Cylindrospermopsin biosynthetic gene-directed PCR 59
3.2.5 Long template PCR 60
3.3 Results and Disucssion 61
3.3.1 Genome walking 61
3.4 Summary 92
4 BIOINFORMATICS AND PUTATIVE BIOSYNTHETIC PATHWAY OF THE CYLINDROSPERMOPSIN GENE CLUSTER 94
4.1 Introduction 94
4.2 Specific Materials and Methods 96
4.3 Results and Discussion 96
4.3.1 *cynK*-Polyketide synthase (PKS) 98
4.3.2 *cynJ*-Polyketide synthase (PKS) 98
4.3.3 *cynI*-amidohydrolase 98
4.3.4 *cynH*-Prolyl 4-hydroxylase 99
4.3.5 *cynL*-sodium-dependent multidrug extrusion pump 100
4.3.6 *orf 1*, *orf 2* and *orf 3* – IS4 insertion elements 101
4.3.7 *cynF* amidohyrdolase 102
4.3.8 *cynE*-putative aminotransferase/sulfotransferase 102
4.3.9 *cynC*-amidinotransefase 105
4.3.10 *cynD*-Chimeric PKS and nonribosomal peptide synthatase (NRPS) 106
4.3.11 *cynA*-Polyketide synthetase (PKS) 106
4.3.12 *cynB*-Polyketide synthetase (PKS) 107
4.3.13 *cynG*-adenylylsulfate kynase 108
4.3.14 *orf4*-AAA + ATPases 109
4.3.15 Hydrogenase expression cluster *hypC-E* 110
4.4 The proposed biosynthetic pathway 111
4.5 Summary 116
5 INVESTIGATION OF A CYLINDROSPERMOPSIN PRODUCING *CYLINDROSPERMOPSIS RACIBORSKII* BLOOM IN THE LOWER RIVER MURRAY 117
5.1 Introduction 117

5.2 Specific Materials and Methods 119
5.2.1 Location and sampling details 119
5.2.2 Isolation of *C. raciborskii*. CYP 043 119
5.2.3 Toxin Analysis 120
5.2.4 Quantitative real time PCR (qPCR) 121
5.2.5 Universal cyanobacteria hemi-mested 16S *r*RNA gene-directed PCR 123
5.2.6 Denaturing gradient gel electrophoresis 124
5.3 Results and Discussion 125
5.3.1 *C. raciborskii* isolate, CYP043 125
5.3.2 Quantitative real time PCR 126
5.3.3 Environmental conditions within the River Murray 134
5.3.4 Cyanobacterial community composition using DGGE analysis 136
5.4 Summary 143
6 EVOLUTIONARY INVESTIGATIONS INTO THE CYLINDROSPERMOPSIN BIOSYNTHETIC GENE CLUSTER 145
6.1 Introduction 145
6.2 Specific Materials and Methods 149
6.2.1 Summary of isolates 149
6.2.2 Conventional PCR and sequence analysis 149
6.2.3 Cylindrospermopsin biosynthetic gene-directed PCR 149
6.2.4 Long template PCR 156
6.2.5 Phylogenetic analyses 157
6.2.6 Metrics used to determine Horizontal Gene Transfer events. 157
6.2.7 Other Statistical methods 157
6.3 Results and Discussion 157
6.3.1 Sequence and phylogenetic analysis of cylindrospermospin-producing cyanobacteria 157
6.3.2 Codon usage in *C.raciborskii* AWT205 187
6.3.3 Structural comparison between *Aphanizomenon ovalisporum* and *Cylindrospermopsis raciborskii* CYN biosynthetic cluster 191
6.4 Summary 194
7 GENERAL DISCUSSION 196

Bibliography 203

Appendices 219

Appendix 1 *Cylindrospermopsis raciborskii* AWT205 CYN Biosynthetic Gene Cluster Sequence 220

Appendix 2 Detailed results of gene-directed PCR toward the CYN biosynthetic cluster 265

Appendix 3 Distribution of Chi square Probability 267

Appendix 4 Multiple alignment of 16S rRNA global 268

Appendix 5 Multiple alignment of 16S rRNA CYN-producing cyanobacteria only 281

Appendix 6 Multiple alignment of *rpo*C1 286

Appendix 7 Multiple Alignment of *cynC* 294

Appendix 8 Multiple Alignment of *cynD* 303

Appendix 9 Multiple Alignment of *cynA* 310

Appendix 10 Multiple Alignment of *cynB* 317

List of Tables

Table 1.1 Keys to the orders of the Cyanobacteria (blue-green algae) (Waterbury 2006) 2

Table 1.2 Overview of fully sequenced cyanobacterial secondary metabolite biosynthetic clusters 23

Table 2.1 Preparation of ASM-1 and ASM-1 N+Mo (Gorham et al. 1964) 39

Table 3.1 Summary of isolates analysed in Chapter 3 51

Table 3.2 Oligonucleotide primers used to generate probe for cosmid screening expts. 55

Table 3.3 Oligonucleotide primers used for inverse PCR 57

Table 3.4 Primers used in adapted mediated walking PCR 59

Table 3.5 Oligonucleotide primers targeting genes implicated in the synthesis of cylindrospermopsin 59

Table 3.6 The putative biochemical roles of ORFs identified within the CYN biosynthetic cluster. 85

Table 3.7 Chi-square contingency analysis to collaborate CYN toxicity with presence of CYN genes. 88

Table 5.1 Oligonucleotide primers and fluorescent probes used in real-time qPCR 122

Table 5.2 Quantitative real-time PCR data for the duplex CYN and *C. raciborskii-specific* assay (after correction for 50-fold concentration). 128

Table 5.3 The molecular identity of DGGE bands excised from environmental waters of the River Murray 139

Table 6.1 Summary of isolates analysed in Chapter 5. Sequences obtained from Genbank database are indicated in bold 150

Table 6.2 Oligonucleotide primers used to amplify the *rpo*C1 gene locus 156

Table 6.3 The amino acids (excluding termination and initiation codons) and their preferred codon for coding regions of the CYN biosynthetic genes, internal and external to the putative transposase *orf* boundaries and primary metabolism associated genes. 190

List of Figures

Figure 1.1 A comparison between of FAS and generic PKS homologation 5

Figure 1.2 Scheme representing the function and mechanisms of core PKS catalytic domains... 7

Figure 1.3 Representative scheme demonstrating the co-linear and processive nature of modular type I PKSs: 8

Figure 1.4 Reaction scheme illustrating type II PKS synthases, of iteratively acting subunits 9

Figure 1.5 Reaction scheme illustrating type III PKS synthases 10

Figure 1.6 Scheme representing the function and mechanisms of core NRPS catalytic domain 12

Figure 1.7 Chemical structure of microcystin-LR 14

Figure 1.8 Chemical structure of nodularin-R 15

Figure 1.9 The structure of anatoxin-a Acetyl is replaced at C2 with propionyl for homoanatxoin-a 16

Figure 1.10 The structure of anatoxin-a(s) 16

Figure 1.11 The general structure for saxitoxin 17

Figure 1.12 Chemical structure of cylindrospermopsin 18

Figure 1.13 Chemical structure of deoxy-cylindrospermopsin 19

Figure 1.14 Chemical structure of 7-epicylindrospermopsin 19

Figure 1.15 Structures of the nodularin (*nda*) and microcystin (*mcy*) gene clusters 25

Figure 1.16 Real-time melting curve analysis 30

Figure 1.17 Schematic representation of commonly used real-time PCR chemistries 32

Figure 2.1 Map of pCR® 2.1-TOPO® (Invitrogen Corporation) 45

Figure 3.1 Organisation and physical map of the C. raciborskii AWT205, 7671 bp nucleotide sequence contained on the cosmid clone CD7, identified and sequenced by Fergusson (2003). 62

Figure 3.2 Experiment A – P^{32} hybridisation of a 1.1 kb *cynB* probe to a cosmid library constructed from *C. raciborskii* AWT205 chromosomal DNA. 64

Figure 3.3 Diagrammatic representation of the process of inverse PCR. 67

Figure 3.4 Open reading frame organisation of the extended *C. raciborskii* AWT205 CYN biosynthetic cluster nucleotide sequence (2916 bp), obtained by inverse PCR 69

Figure 3.5 Experiment B: P^{32} hybridisation of a 672 clone cosmid library constructed from *C. raciborskii* AWT205 chromosomal DNA. 71

Figure 3.6 Amplification products from a pCR2.1 plasmid insert, obtained from cosmid clone DH5, using the cospro3 and cospro4 primer set, targeting the *ask* gene 72

Figure 3.7 Restriction enzyme digestion of cosmid clone DH5 with *Dra*I, *Eco*R1 and *Not*I. 73

Figure 3.8 Amplification of the *orf4* (putative ABC transporter/AAA+ ATPase) gene from bands excised from *Dra*I (1.3 kb) and *Eco*RI (1.9 kb) restriction enzyme digest of cosmid clone DH5 74

Figure 3.9 Diagrammatic representation of the processes involved in adapted mediated genomic walking strategy. 76

Figure 3.10 Open reading frame organisation of the *C. raciborskii* AWT205 CYN biosynthetic gene cluster, including 2217 bp nucleotide sequence obtained by adapted mediated walking. 78

Figure 3.11 Gene organisation of a 44.2 kb region sequenced from *C. raciborskii* AWT205 80

Figure 3.12 Cylindrospermopsin biosynthetic cluster Gene-directed PCR results 84

Figure 3.13 Negative image of long-templated PCR to verify CYN gene deletion mutant *C. raciborkii* CYP 018A 91

Figure 4.1 Structure of cylindrospermopsin with assigned designation of structural atoms (red) 94

Figure 4.2 Gene organisation of the CYN biosynthetic cluster, with identified domains within the polyketide synthases and nonribsomal peptide synthetase. 97

Figure 4.3 Reaction scheme for amide bond hydrolysis catalysed by dihydrooratase (DHO) 99

Figure 4.4 Structure of guanidinoacetic acid. 103

Figure 4.5 Sequence alignment of 5' PSB-loop motifs from sulfotransferases 105

Figure 4.6 Alignment of sequences for motifs found in AT domains in PKS type I (Haydock et al 1995) 107

Figure 4.7 Sequence comparisons of adenylsulfate kinase 109

Figure 4.8 The proposed biosynthetic pathway for the production of cylindrospermopsin 115

Figure 5.1 Location of River Murray sampling sites used in this study. Lock 9 (not shown) is approximately 450 km further upstream from Morgan. 120

Figure 5.2 River Murray isolate, *Cylindrospermopsis raciborskii*, CYN43. 126

Figure 5.3 Quantitative real-time standard curve for quantification of *pks* gene target copy numbers. 127

Figure 5.4 Quantitative real-time PCR standard curve for quantification of *rpo*C1 gene target copy numbers 127

Figure 5.5 Graphical representation of qPCR for the cylindrospermopsin *pks* and *C. raciborskii-specific rpo*C1 genes and corresponding cell counts of *C. raciborskii* 129

Figure 5.6 DGGE profiles from 5 sampling locations along the River Murray during the summer season 2007/2008. 138

Figure 5.7 Cell numbers for three dominant cyanobacterial species, River Murray, Swan Reach sampling site 2007-2009 141

Figure 5.8 Cell numbers for three dominant cyanobacterial species, River Murray, Walkers Flat sampling site 2007-2009. 142

Figure 5.9 Cell numbers for three dominant cyanobacterial species River Murray, Morgan Sample Pump, sampling site 2007-2009. 142

Figure 6.1 A representative tree and associated terms 148

Figure 6.2 Phylogenetic relationships of selected filamentous cyanobacteria genera and strains using partial 16S rRNA gene sequences. 162

Figure 6.3 The NJ phylogenetic tree reconstructed for the 16S rRNA gene of CYN-producing isolates. 165

Figure 6.4 The NJ phylogenetic tree reconstructed for the *rpo*C1 gene of CYN-producing isolates. 169

Figure 6.5 The NJ phylogenetic tree reconstructed for the Cylindrospermopsin amidinotransferase, *cynC* gene. 173

Figure 6.6 The NJ phylogenetic tree reconstructed for the Cylindrospermopsin acyltransferase domain within the PKS, *cynD* gene. 177

Figure 6.7 The NJ phylogenetic tree reconstructed for the Cylindrospermopsin acyltransferase domain within the PKS, *cynA* gene. 180

Figure 6.8 The NJ phylogenetic tree reconstructed for the Cylindrospermopsin acyltransferase domain within the PKS, *cynB* gene. 186

Figure 6.9 Structural organisation of a region of a CYN biosynthetic cluster from *Aph. ovalisporum* ILC-146 (A) and *C. raciborskii* AWT 205 (B). 192

Figure 6.10 PCR amplification of the genomic region spanning the *cynD* (*aoaB*) and *cynA* acyltransferase (AT) domains. 193

Figure 6.11 Structural organisation of a region of a CYN biosynthetic cluster (in solid) for *Aph. ovalisporum* ILC-146 (A) and *C. raciborskii* AWT 205 (B). 194

Abstract

Cyanobacteria are ubiquitous throughout global water bodies. They have the proclivity to produce a number of potent secondary metabolites that can sully potable waters, with potentially serious implications in respect to public health. Furthermore, these compounds can impact on aquaculture, agriculture and recreational activities. It is therefore imperative that water quality managers are able to rapidly differentiate between problematic and non-problematic cyanobacterial populations, and thus ensure appropriate measures are initiated early to guarantee public health. Molecular based approaches are proving attractive alternatives to current detection strategies, where molecular probes can be designed to be specific, sensitive and lend themselves to greater portable and rapid technology platforms. However, there is a paucity of molecular data available for the biosynthetic mechanisms involved for many of these cyanobacterial secondary metabolites.

Cylindrospermopsin (CYN) is a potent cyanotoxin that inhibits mammalian protein synthesis, leading to toxicity in a range of different organs. CYN is a mutagen and preliminary evidence suggests that long term exposure may have carcinogenic consequences. Identification of CYN in global water bodies is becoming more frequent, perhaps reflective of improved detection strategies in addition to the proliferation of cyanobacteria promoted by eutrophication and global warming. Historically, CYN has had serious repercussions in regard to Australian water quality, and has been implicated as the causative agent for an acute gastrohepatic illnessthat affected a community on Palm Island, QLD. The cyanotoxin has also been attributed to livestock deaths.

To date there have been three candidate genes, correlated with CYN toxicity including an amidinotransferase, a polyketide synthase (PKS) and a peptide synthetase (PS); where the PKS and PS have already been applied for molecular detection strategies. However, without full consideration of the molecular architecture involved in CYN biosynthesis, it is impossible to conclude whether the most explicit molecular target has been selected to unequivocally detect CYN-producing cyanobacteria. Further to this, acquiring the molecular information in respect to the biosynthesis of CYN will facilitate future investigations into the regulation and ecological role of this metabolite.

This thesis describes the detection, sequencing and characterisation of a putative 43KB biosynthetic gene cluster isolated from CYN-producing *Cylindrospermopsis raciborskii* AWT205. A suite of molecular methods were employed to acquire the gene sequence. Through rigorous PCR screening of over 40 CYN- and non-CYN producing cyanobacterial species and strains, the genes involvement in CYN production was implicitly defined. Additionally, the screening permitted identification of suitable candidate genes allowing for unambiguous detection of CYN-producing cyanobacteria in future applications.

The acquired molecular knowledge was also practically applied to track a CYN-toxic cyanobacterial bloom using real-time PCR. The outcomes provided novel insights into the molecular mechanics of bloom-forming *Cylindrospermopsis raciborskii* and the production of CYN in the bloom environment. This knowledge may prove beneficial in guiding water quality managers to target algicide treatments at optimal time points.

Phylogenetic analyses of 21 CYN-producing cyanobacteria, from three distinct genera, was also undertaken to investigate the evolutionary history of the CYN biosynthetic cluster. In addition to multiple alignments, three tree construction methods and codon usage analyses were employed, to investigate the involvement of horizontal gene transfer in disseminating the cluster throughout cyanobacterial genomes.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Rebecca Jane Campbell, August 2009

Signed ..... Date.....

Acknowledgements

First and foremost I sincerely wish to thank my Principal Supervisor Professor Chris Saint for his continued support and encouragement over my eventful candidature. It was your continued belief that I could get there, and I should get there that got me through in the end.
I wish to extend a really big thankyou to my other supervisor Dr Paul Monis, who was always there when I really needed his very cool calm direction and support. I also wish to thank Dr Paul Rasmussen for his mentorship at the beginning of this work.

I acknowledge the scholarship support from the CRC for Water Quality and Treatment, now known as WQRA. I thank Professor Dennis Mulcahy and Ms Carolyn Bellamy for their continued support over this candidture.

I wish to extend a special thanks to the Microbiology R&D group, especially Dan Hoefel for his invaluable input regarding this thesis, Steve Giglio for his help with real-time PCR, and sharing the PhD pain, Brendon King, Alex Keegan, Stella Fanok, Bret Robinson, Suz Froscio, Melody Lau and Dani Tutuka. It has (mostly) been alot of fun! It was great sharing this PhD journey with Lek Laingam. I thank Peter Baker, Jenny House and Mike Burch for their answers to my queries on biological matters. Also big thanks to Drs Anke Stüken and Barbara Sendall for their collaboration and exchange of sequence information.

To my boys; my husband Mark and our beautiful Max, I thank you for your patience and understanding over the last 8 very intense months. I am so proud of how the both of you have pulled together, to get us through this as a unit. To my darling Max, now when you look at me and say “Mummy you don’t go to’ werk’?”, I can now say” No my love, I don’t go to work tonight, lets find possums and chase the moon!”. Let the party truly begin.

I thank my family: Peter and Merdith for their continuous support, especially for their care of Mark and Max in my absence. To my parents, Gary and Helena, despite the distance I could always feel your support and encouragement. To my sisters Alison and Moo, for their constant attention to Max and editorial and formatting assistance, and Nadia for diagram support. To our friends for their laughter, encouragement and support of our family, I thank you all from the bottom of my heart.

Publications in support of this thesis

S. A. Wood, J. P. Rasmussen, P. T. Holland, **R. Campbell**, A. M. Crowe 2007 First report of the cyanotoxin anatoxin-A from *Aphanizomenon issatschenkoi*(cyanobacteria). Journal of Phycology. **43**:356-365

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

A. Stüken, **R. J. Campbell**, A. Quesada, A. Sufenik, P. J. Dadheech, C. Wiedner 2009 Genetic and morphologic characterization of four putative cylindrospermopsin producing species of the cyanobacterial genera Anabaena and Aphanizomenon. Journal of Plankton Research. **31**:465-480

R. J. Campbell, P. T. Monis, J. P. Rasmussen, A. Stüken, C. P. Saint 2009 The long and short of it : a history of cylindrospermopsin in cyanobacterial genomes. Manuscript in preparation.

Abbreviations, prefixes and symbols

The following abbreviations have been used throughout this thesis.

aa	amino acid
ACP	acyl carrier protein
agrf	Australian Genomic Research Facility
Amp	ampicillin
APS	adenosine 5’-phosphosulfate kinases
ASK	adenylylsulfate kinase
ASM-1	artificial seawater medium
ASM-1 –N +Mo	ASM-1 with nitrogen removed and molybdenum added
AT	acyltransferases
ATP	adenosine 5'-triphosphate
AWQC	Australian Water Quality Centre
BCAT	branched-chain amino acid transferases
bp	base pairs
C_t	cycle threshold
CYN	cylindrospermopsin
dCTP	deoxycytidine triphosphate
DGGE	denaturing gradient gel electrophoresis
DH	dehydratase
DHO	dihydrooratse
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
EDTA	ethylenediaminetetraacetic acid
g	grams
HGT	horizontal gene transfer
HPLC	high performance liquid chromatography
IMVS	Institute of Medical and Veterinary Science
IPCR	inverse PCR
Kb	kilo bases
KR	β-ketoreductase
KS	β-ketoacyl synthase
L	litre
LB	Luria-Bertani
LOD	limit of detection
M	moles L^{-1}
MATE	multidrug and toxic compound extrusion
MDBC	Murray Darling Basin Commission
MIB	methyisoborneol

mL	millilitre
ML	maximum likelihood
mM	millimoles L^{-1}
MP	maximum parsimony
MT	methyltransferase
NCBI	National Centre of Biotechnology Information
ng	nanograms
NJ	neighbour joining
NRPS	nonribosomal peptide synthetase
NSW	New South Wales
NTU	nephelometric turbidity units ,
orf	open reading frame
P	probability
P4H	prolyl 4-hydroxylase
PAPS	3'-phosphoadenosine 5'-phosphosulfate
PAR	photosynthetic active radiation
PCP	peptidyl carrier protein
PCR	polymerase chain reaction
PI	protein inhibition
PKS	polyketide synthase
PLP	pyridoxal 5'-phosphate
Promega	μM deoxynucleoside triphosphates
PS	peptide synthetase
QLD	Queensland
qPCR	quantitative real-time PCR
RE	restriction enzymes
RSCU	relative synonymous codon usage
SA	South Australia
SAM	*S*-adenosylmethionine
SMQ	sterile milli-Q water
SNP	single nucleotide polymorphism
TAE	Tris/acetate (buffer)
TE	thioesterase domain
Tris	tris(hydroxymethyl)aminomethane
Vic	Victoria
v/v	volume/volume
w/v	weight /volume
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactoside
μL	micro litres
μg	micrograms

1 GENERAL INTRODUCTION

Cyanobacteria are renowned for their ability to produce a myriad of secondary metabolites, many of which elicit toxicological responses in both human and animals. The escalating number of cyanobacterial blooms worldwide proves challenging to water quality managers thus it is imperative they are able to determine the potential threat of toxic bloom-forming cyanobacteria at the earliest time point to minimize the risks to human and animal health. This introduction provides a background to cyanobacteria and some of the secondary metabolites they produce that have been characterized at a molecular level. It discusses the implications with unravelling the molecular mechanisms, in particular with regard to development of molecular techniques that permit early detection of toxic threats, along with the evolutionary insights this genetic knowledge can convey. This thesis characterises a putative gene cluster implicated in the biosynthesis of one of the dominant cyanotoxins found in freshwater aquatic systems worldwide, cylindrospermopsin (CYN). It examines the possible evolutionary history of the toxin and investigates the recent occurrence of a CYN toxic bloom in a temperate environment in Australia, by the invasive cyanobacterium *Cylindrospermopsis raciborskii.*

1.1 Cyanobacteria

1.1.1 Species Diversity and global distribution

Cyanobacteria are a large group of photosynthetic prokaryotes that have had an influential contribution toward shaping the course of evolution and atmospheric change throughout the earth's history; being responsible for a large proportion of global CO_2 and N_2 fixation (Kasting and Siefert 2002). They possess a functional plasticity that confers great versatility permitting a wide ecological distribution. Members of the phylum have found niches in a diverse array of aquatic environments including marine, brackish and freshwaters, as well as terrestrial and extreme habitats (polar regions and hot thermal springs). Cyanobacteria form a monophyletic but morphologically diverse group. Traditional taxonomy has been used to classify the phylum into 5 orders; key structural and developmental characteristics of each are described below, with representative genera of each presented in Table 1.1 (Waterbury 2006).

Table 1.1 Keys to the orders of the Cyanobacteria (blue-green algae) (Waterbury 2006)

Order	Key Characteristics	Common Genera
Chroococcales	Unicellular or colonial; Reproduce by binary fission; cell division in one, two or more perpendicular planes; Heterocyst and akinete lacking	*Cyanothece* *Gloeothece* *Microcystis* *Prochlorococcus* *Synechococcus* *Synechocystis*
Pleurocapsales	Reproduce by multiple fission; unicellular or cell aggregates	*Cyanocystis* *Chroococcidiopsis* *Pleurocapsa*
Oscillatoriales	Filamentous with undifferentiated cells; cell division by binary fission in one plane; reproduce by trichome fragmentation	*Geitlerinema* *Leptolyngbya* *Limnothrix* *Lyngbya* *Oscillatoria* *Planktothrix* *Pseudanabaena* *Trichodesmium*
Nostocales	Filamentous with true trichomes; capable of cell differentiation 1) heterocysts 2) akinetes 3) hormogonia (reproductive trichomes) 4) tapered trichomes; divide by binary fission in one plane	*Anabaena* *Aphanizomenon* *Cylindrospermopsis* *Nodularia* *Nostoc*
Stigonematales	Filamentous; true branching; capable of specialised cell differentiation as members of Nostocales; able to divide by binary fission in multiple planes	*Fischerella* *Umezakia*

1.1.2 Blooms and management

Cyanobacterial dominance over phytoplankton populations in aquatic environments has been attributed to several traits that permit a distinct, competitive advantage and have been reviewed by others (Hyenstrand et al. 1998; Havens 2008) . Some of these traits include:

- buoyancy regulation, to enable them to exploit the water column to fulfil their nutrient and photosynthetic requirements
- accessory pigments that assist with chromatic adaption to protect from high surface irradiance or low light due to shading.

- differentiation of heterocysts by several species, for atmospheric dinitrogen (N_2) fixation permitting their existence in nitrogen limited conditions, for example in the ocean.
- differentiation of akinetes by several species, for survival under adverse conditions, for example during periods of cold and desiccation.
- Resistance to grazing, as they appear inedible to herbivorous zooplankton (Bouvy et al. 2001).
- High affinity for nitrogen and phosphorus.

Cyanobacteria proliferate when desirable conditions prevail. Favourable conditions are typically defined by increased stratification of the water column, low total nitrogen to total phosphorus ratios, increases in water temperature and light to growth optima (Fabbro and Duivenvoorden 1996; Paerl et al. 2001; Paerl and Huisman 2009). The resultant explosion in a cyanobacterial population (bloom) leads to increased turbidity, with negative consequences for other aquatic phytoplankton. The bloom can appear as thick surface scums, exist throughout the water column or form mats on the sediment interface. The ensuing light deficient conditions and depletion of oxygen levels can lead to immense biological impacts exemplified by fish kills and, if the bloom persists for significant periods, decreased ecological diversity. In recent years there have been observations of increasing frequency and severity of cyanobacterial blooms. Eutrophication of surface waters driven by anthropogenic activities, natural climatic events (i.e. droughts) and increasing global temperatures have been implicated as causes for these increases in bloom formation (Pitois et al. 2001; Paerl and Huisman 2009).

Due to their propensity to form blooms, aquatic cyanobacteria are intensely studied. The ability of many of the bloom-forming cyanobacterial species to produce bioactive (cyanotoxins) or odorous secondary metabolites threatens the safety and aesthetics of potable water. Further to this, these compounds can pose a significant hazard to agriculture, aquaculture and recreational activities (Chorus and Bartram 1999; Falconer 1999; Chorus et al. 2000). Moreover, the economic consequences are significant. Limiting access to the affected water body may impact on businesses relying on it for revenue. Additionally, restoration practices required to disperse (e.g. mechanical mixing and flushing) (Maier et al. 2001) or eradicate (e.g. algicides) (Steffensen 2008) blooms and water treatment using powered activated carbon (PAC) or ozone to remove and destroy the

toxins or taste and odour compounds greatly increases the fiscal operating costs (Steffensen 2008).

Management strategies can be confounded by environmental heterogeneity, which can lead to the patchy distribution of blooms in respect to intensity, duration and toxin potential (Ohtake et al. 1989; Vezie et al. 1998). Research is driven to understand the complexities behind the mechanics and ecological drivers of bloom formation and the production of toxic secondary metabolites (Bouvy et al. 1999; McGregor and Fabbro 2000; Briand et al. 2008; Briand et al. 2009). A better understanding will ultimately be directed toward the development of proactive restorative management practices, to minimise the ecological, economical and public health impacts that cyanobacteria and their compounds can cause.

1.2 Secondary metabolites

Secondary metabolites are intermediates or by-products of metabolism that are not essential for the primary growth and development of an organism. They are believed to function as beneficial tools that permit an ecological advantage, though the discrete function remains elusive for many of these compounds. It has been suggested that the toxic cyanobacterial metabolites exhibit alleopathic effectsthat influence the microenvironment to protect themselves from predators (Bouvy et al. 2001) and establish founding populations (Jungmann and Benndorf 1994; Schatz et al. 2007). Other theories advocate an intracellular role in sequestering metal ions and facilitating redox activities (Blackburn et al. 1997). Secondary metabolites produced in prokaryotes are predominantly polyketide or non-ribosomal peptide derived.

1.2.1 Polyketide synthases

The biosynthesis of polyketides is enacted by polyketide synthases (PKS), which are viewed as analogous to long chain fatty acid systems (FAS) found in eukaryotes. Parallels exist between the two systems both on a sequence and mechanistic level (Figure 1.1). Differences lie with the degree of processing of the β-carbon; for FAS the β-carbon is fully reduced prior to the cycle resuming with the next condensation of the malonyl-CoA extender unit.

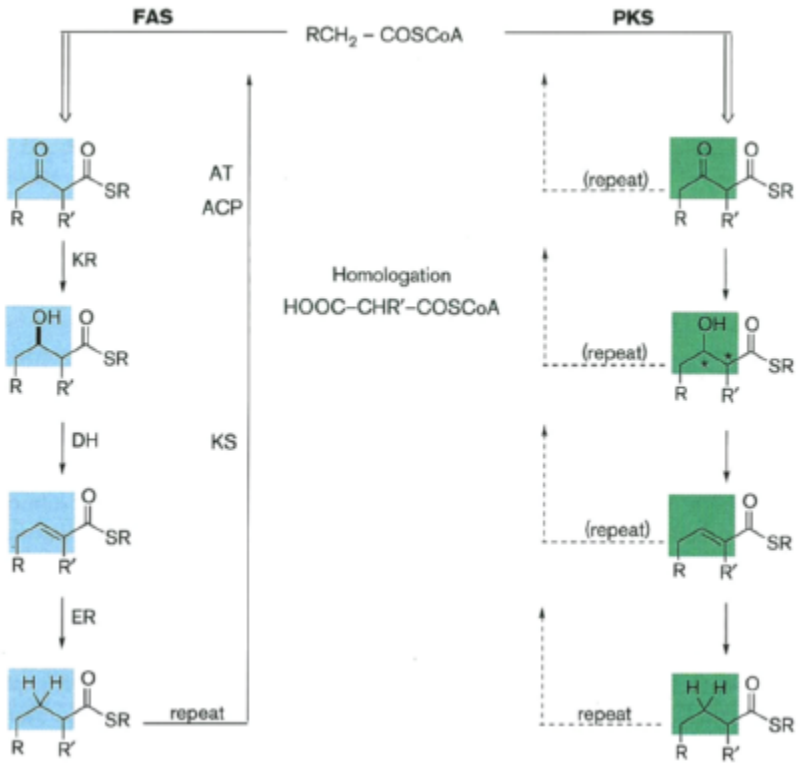


Figure 1.1 A comparison between of FAS and generic PKS homologation

Homologation of acetylCoA with malonyl CoA (and their substituted derivatives with different groups, R and R'), revealing a varying extent of thioacyl β-carbon processing prior to further homologation. KS, β-ketoacyl synthase; AT, acyl transferase; ACP, acyl carrier protein; KR, β-ketoacyl ketoreductase; DH, dehydratase; ER, enoyl reductase *Reproduced from (Townsend 1997)*

However, PKSs set a precedent in regards to both structural diversity and stereochemical integrity. The structural complexities observed within PKS systems, are attributed to their capacity to incorporate different extender units, commonly malonyl-Coenzyme-A (malonyl Co-A) but may also involve methymalonyl-CoA, propionyl-CoA and other derivatives. Furthermore, the presence of auxiliary catalytic domains vary in their extent of processing the β-carbon (Donadio et al. 1991).

Polyketide synthases have been comprehensively reviewed by numerous authors (Hopwood and Sherman 1990; Cane et al. 1998; Khosla et al. 1999; Staunton and Weissman 2001; Khosla et al. 2007). Polyketide synthases are large multi-modular and multi-functional proteins, organised in coordinated groups. Typically, each PKS module is composed of three mandatory domains, two catalytic (acyl transferase, AT and β-ketoacyl synthase, KS) and one carrier (acyl carrier protein, ACP). Additionally, these modules may

also contain auxiliary catalytic domains that further modify the β-carbon functionality. Auxiliary modifications may include reductive β-ketoacyl ketoreductases (KR); dehydratases (DH); enoyl reductases (ER) domains; tailoring domains including methyl transferases (MT) and halogenases (H) or terminating thioesterases (TE)that may also been involved in the lactonisation of the final product. The full length chain is synthesised by the concerted efforts of organised synthase modules somewhat akin to an assembly line. Furthermore, the size and complexity of the ultimate product can be predicted by the number of modules and the order and nature of the catalytic domains within each.

The biosynthesis of polyketides is initiated by the covalent loading of a thio-activated starter unit onto the first module's ACP domain. The carrier protein contains a flexible phosphopantetheinyl binding arm that swings the initiating monomer to a downstream module's KS domain. One elongation cycle then proceeds, with the AT domain priming a chain extending malonyl-CoA (or other acyl-CoA derivatives) onto the consecutive ACP. The primed ACP swings the extender unit toward the KS domain, which proceeds to catalyse the claisen-like decarboxylative condensation of the extender unit to the nascent polyketide chain (Figure 1.2A). Reductive, functional modification then proceeds, dictated by the auxiliary domains contained within each module (Figure 1.2B). The resultant intermediate then acts as the substrate for the subsequent module and the next round of elongation begins. The intermediate remains covalently bound by thioester linkages from the initial condensation of an extending monomer to the starter unit, till the desired chain length is attained. Chain termination is commonly catalysed by a COOH-terminal TE domain in the final module. The core catalytic domains demonstrate narrow tolerances for substrates on which they act, which has assisted in the determination of critical amino acid residues that dictate stereo-specificity (Caffrey 2003) and substrate preferences (Haydock et al. 1995) for KR and AT domains respectively.

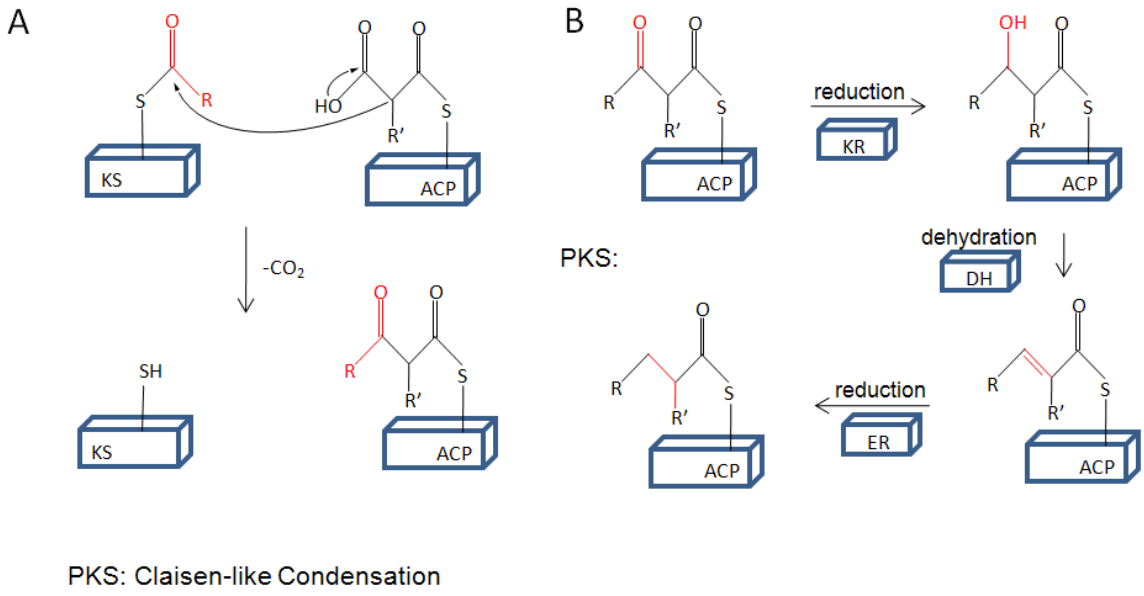


Figure 1.2 Scheme representing the function and mechanisms of core PKS catalytic domains.

A Represents chain elongation and condensation reactions for a minimal PKS module. The ACP is primed by the acyl transferase (not shown) with the malonyl (R=H) or methylmalonyl (R'=Me) chain extender. KS, β-ketoacyl synthase; ACP, acyl carrier protein **B** Reductive modifications between condensation steps catalysed by PKS modules. KR, β-ketoacyl ketoreductase; DH, dehydratase; ER, enoyl reductase. *Adapted from Cane et al (1998)*

Currently, there are considered to be three classical PKS paradigms (Shen 2003):

- Type I PKSs are modular multifunctional complexes with each module being responsible for one round of polyketide chain elongation and subsequent β-carbon processing. These types are exemplified by the 6-deoxyerythromycin (DEBS) B synthase (Figure 1.3). The elucidation of the 6-deoxyerythromycin (DEBS) B synthase (Cortes et al. 1990) served as a model system for some years, permitting an understanding of the mechanisms by which these systems operate. However, the apparent colinear processing once thought to be a requirement of these systems has long been dismissed, as new systems that redefine module and catalytic domain organisation emerge. Recently two unique mechanisms, module skipping and iterative use of a module, have been additionally identified for type I PKSs (Moss et al. 2004). Type I biosynthetic clusters have also been identified that lack a cognate AT domain, which has previously been accepted as a minimal requirement within a module (Cheng et al. 2003).

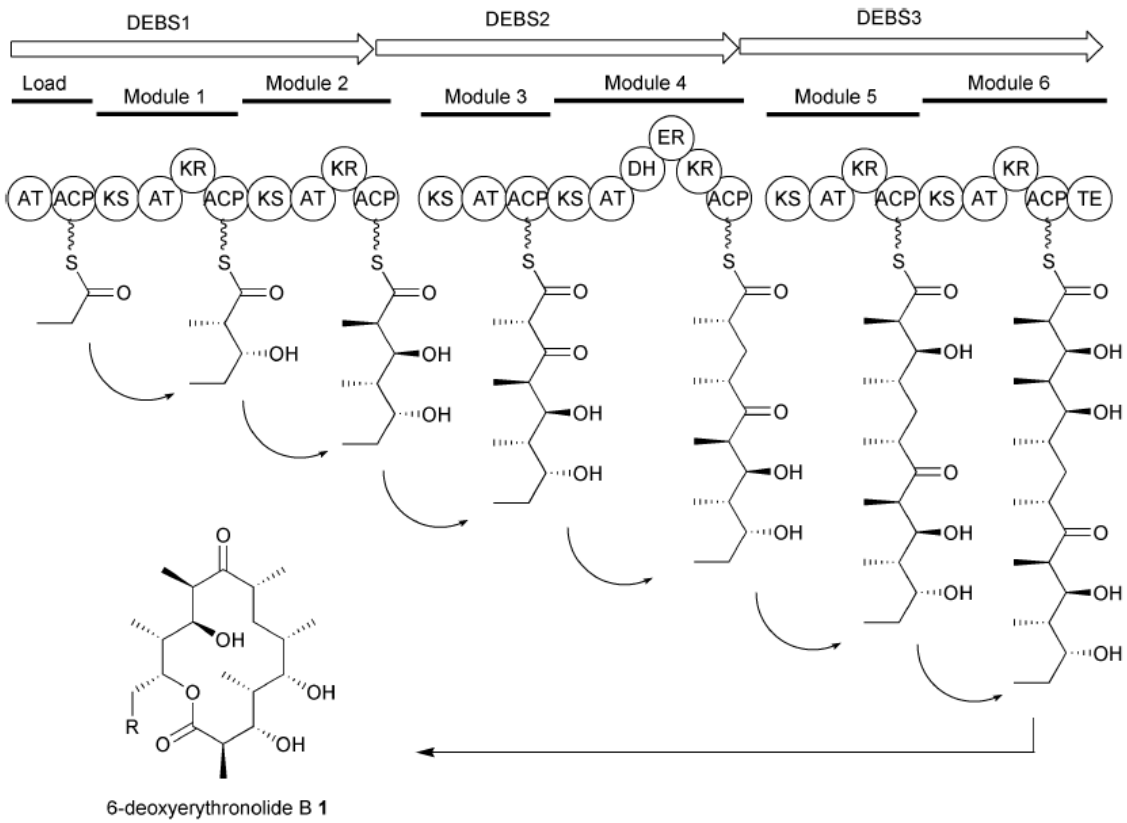


Figure 1.3 Representative scheme demonstrating the co-linear and processive nature of modular type I PKSs:

6-deoxyerythromycin B (**1**) biosynthesis by DEBS1, DEBS2 and DEBS3 *Reproduced from Moss et al (2004)*

- Type II PKSs (Figure 1.4), are a class of non-modular proteins that predominantly biosynthesise cyclic and aromatic polyketides. Multiple rounds of elongation are performed, through the iterative use of a set of domains which are present on discrete proteins. No reductive processing of the β-keto carbon occurs till the intermediate reaches the desired chain length, whereby the resultant chain undergoes reduction, cyclisation, aromatisation and other auxiliary enzymatic modifications.

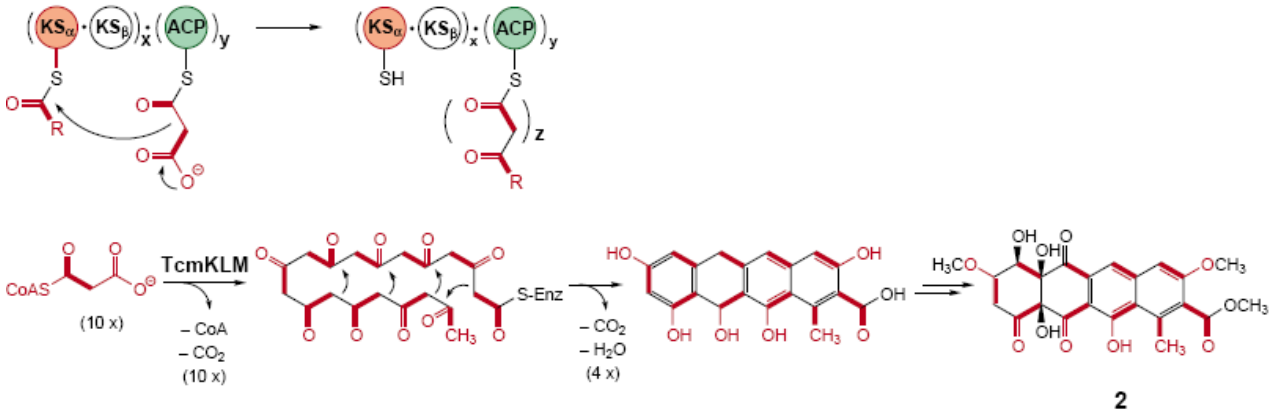


Figure 1.4 Reaction scheme illustrating type II PKS synthases, of iteratively acting subunits

As exemplified by TcmKLM for tetracenomycin (2) biosynthesis. Atoms that were incorporated intact from the acyl-CoA precursors to resultant polyketides are shown in bold. *Reproduced from Shen (2003)*

- Type III PKSs (Figure 1.5) contrast to type I and II by lacking an ACP domain to tether the extending chain. Instead acyl-CoA thioesters are utilised and enacted on at a single site. In all other aspects type III PKSs functions similarly to type II, being iterative and postponing the reductive processing till the required chain length has been attained. Again, predominantly aromatic products are biosynthesised.

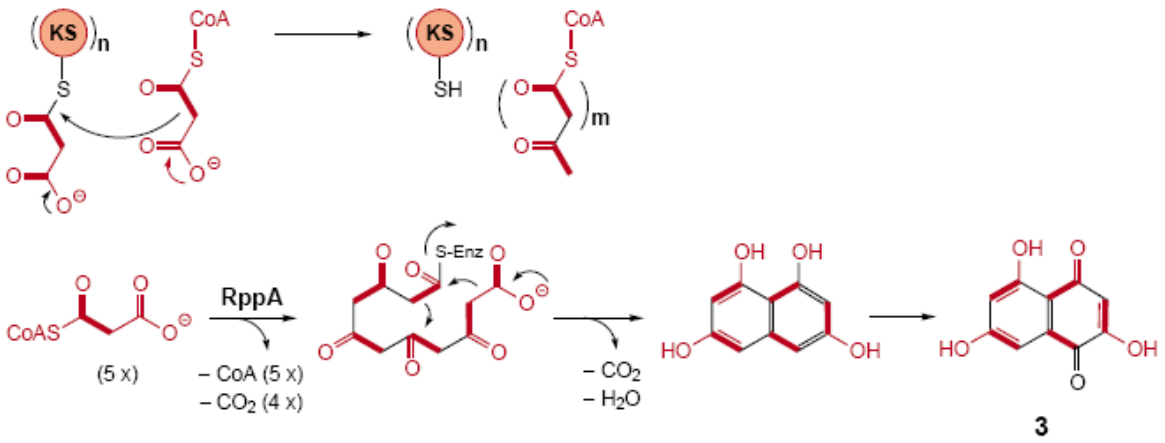


Figure 1.5 Reaction scheme illustrating type III PKS synthases

Consisting of an ACP independent and iteratively acting subunit as exemplified by RppA for flavolin (3) biosynthesis. Atoms that were incorporated intact from the acyl-CoA precursors to resultant polyketides are shown in bold. *Reproduced from Shen (2003)*

1.2.2 Non ribosomal peptide synthetases

Much like modular PKSs, non ribosomal peptide synthetases (NRPS) are also mutimodular and multifunctional megasynthases, which participate in the biosynthesis of complex bioactive macrocyclic peptides. Synthesis precedes independently of a nucleic acid-derived template (Grunewald and Marahiel 2006). Whilst there are recognised differences between PKSs and NRPSs; in monomeric building blocks and mechanisms of biosynthesis; there are also enormous parallels with respect to organisation and the basic manner of assemblage (Cane and Walsh 1999).

The building blocks for NRPSs involve simple amino acid monomers. Each module of NRPS is responsible for the incorporation and functional modification of one amino acid monomer. Much Like PKSs, the number and sequential order of the NRPS modules in addition to the functional modifying catalytic domains incorporated in a module, dictate the resultant product. Minimal module requirements for a NRPS include the presence of adenylation domains (A) that recognise and activate the amino acid substrate; a condensation domain (C), responsible for the elongation of the growing peptide chain and peptidyl carrier proteins (PCP) that are analogous to the PKSs ACP (Figure 1.6A) (Cane et al. 1998). Additionally, a NRPS module may also house auxiliary domains responsible for the structural modification of the intermediate. Auxiliary domains may include epimerisation of the α-carbon (E), N-methylation (M) or cyclisation (C) to the peptidic backbone, the resultant chemistry of each is represented in Figure 1.6B.

Currently there are over 400 different amino acid monomers identified that are recognised by substrate specific A-domains in NRPSs (Kleinkauf and von Döhren 1996; Marahiel and Essen 2009). A "specificity-conferring code" was established, from review of biochemical, phylogenetic and crystal structure data available from several characterised A-domains (Stachelhaus et al. 1999; Challis et al. 2000). This permitted the resolution of the preferred substrate for an A-domain in most cases. The NRPS uses a comparable elongation strategy to PKS modular complexes. An activated peptide unit, loaded on an upstream PCP, is attacked by an amino acid monomer loaded on a downstream module's PCP arm, a C-domain catalysing the condensation and forming the peptide bond. The biochemical mechanisms employed by NRPS have been extensively reviewed by several authors (Cane et al. 1998; Cane and Walsh 1999; von Döhren et al. 1999; Finking and Marahiel 2004; Walsh 2004; Grunewald and Marahiel 2006; Marahiel and Essen 2009).

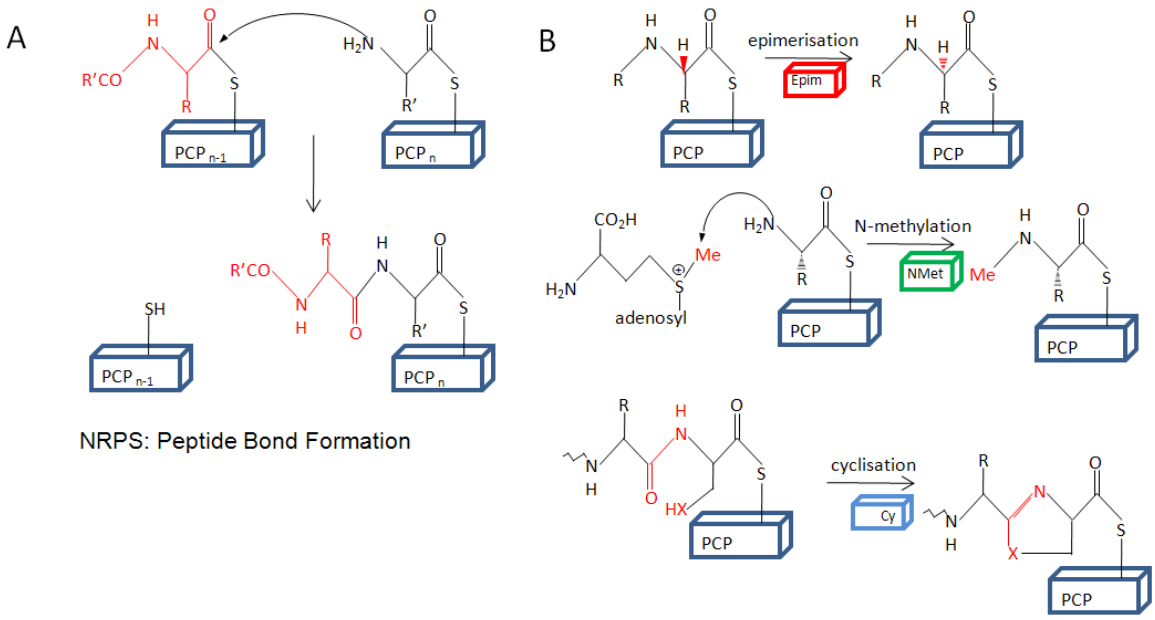


Figure 1.6 Scheme representing the function and mechanisms of core NRPS catalytic domains

A Peptide bond formation catalysed by a NRPS. Peptidyl carrier protein (PCP).

B Modifications between condensations steps catalysed by NRPSs. Epimerase, Epim; *N*-methyl transferase, NMet; cyclase, Cy. *Adapted from Cane et al (1998)*

1.2.3 Hybrid polyketide and non ribosomal peptide complexes

Perhaps due to the structural and basic biosynthetic similarities they share, there are natural examples of hybrid PKS and NRPS systems. The evolution of PKS/NRPS hybrids further contributes to the diversity of natural products produced by these systems. There are at least two naturally occurring forms of mixed PKS/NRPS systems. Firstly, there are examples where the biosynthetic organisation of the cluster comprises a combination of PKS and NRPS encoding modules. This type is exemplified by rapamycin synthase. Whilst being predominantly polyketide in origin; *rapP* is an independent NRPS-like gene that encodes the protein that mediates peptide bond formation between a pipecolate unit and the polyketide backbone, followed by cyclisation and subsequent cleavage of the resultant product (Staunton and Wilkinson 1997).

Additionally, integrated genes encoding both PKS and NRPSs domains on a single polypeptide have also been identified and constitute the second type. In these instances a C-C bond elongation step is mediated by a KS domain of the PKS, through

decarboxylative condensation to the amino acid monomer. The KS domain accepts the peptidyl intermediate tethered to the upstream NRPS PCP and catalyses the condensation in conjunction with the acyl-CoA located on a downstream ACP (Du et al. 2001). Many of the cyanobacterial secondary metabolites that have been characterised to date include integrated PKS/NRPS systems (Nishizawa et al. 2000; Chang et al. 2002; Chang et al. 2004).

The most intensively studied cyanobacterial secondary metabolites are those that elicit a toxicological response, the cyanotoxins. They have been implicated with episodes of acute poisoning and death in humans and both wild and domesticated animals (Stewart et al. 2008). Further to this there is preliminary evidence that several cyanotoxins lead to tumour promotion upon long term exposure (Falconer et al. 1988; Falconer and Humpage 2001; Zhou et al. 2002) . However, not all cyanotoxins are considered detrimental to humans; there are potential leader compounds that could be developed into anti-viral and anti-tumour agents (Burja et al. 2001), or as potential herbicide and insecticides of the future (Berry et al. 2008).

1.3 Cyanobacterial toxic metabolites

The cyanotoxins have been comprehensively reviewed, discussing their chemical structure, modes of action and the potential short and long-term consequences to water bodies (Carmichael 1992; Falconer 1999; Carmichael 2001; Falconer 2001; Rao et al. 2002; van Apeldoorn et al. 2007). There are four distinct classes of toxic secondary metabolites that have been classified across all orders of cyanobacteria; hepatotoxins (injurious to liver), neurotoxins (nerve damaging), cytotoxins (generalized cell damage), and endotoxins (general irritants). Those commonly identified in freshwater systems will be briefly discussed below.

1.3.1 Hepatotoxins

1.3.1.1 Microcystin and Nodularin

Microcystins are a remarkably diverse class of heptapeptides with over 65 known variants. Microcystin-LR (Figure 1.7) is considered the most frequently encountered of all the structural variants and also the most toxic (Chorus and Bartram 1999). The others variants differ in respect to their amino acid composition, structural arrangement and methylation states. Nodularins (Figure 1.8) are pentapeptides which also exhibit a degree of structural variation with several known analogues. Microcystins and nodularins share a number of structural features including the unique 3-Amino-9-methoxy-2, 6, 8-trimethyl-10-phenyldeca-4, 6-diene acid (adda) (Chorus and Bartram 1999).

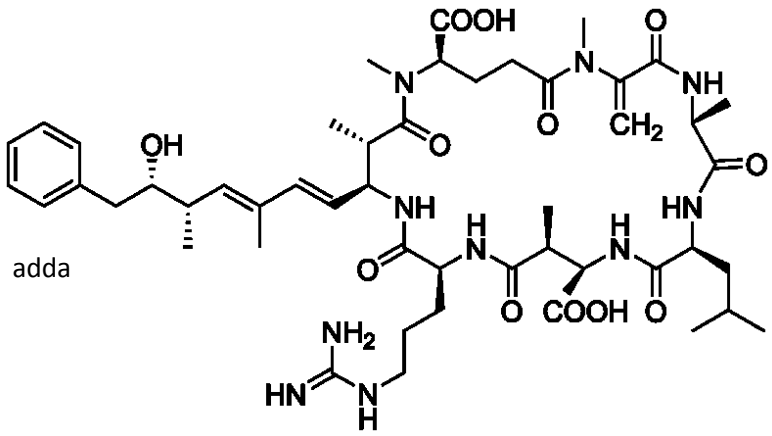


Figure 1.7 Chemical structure of microcystin-LR

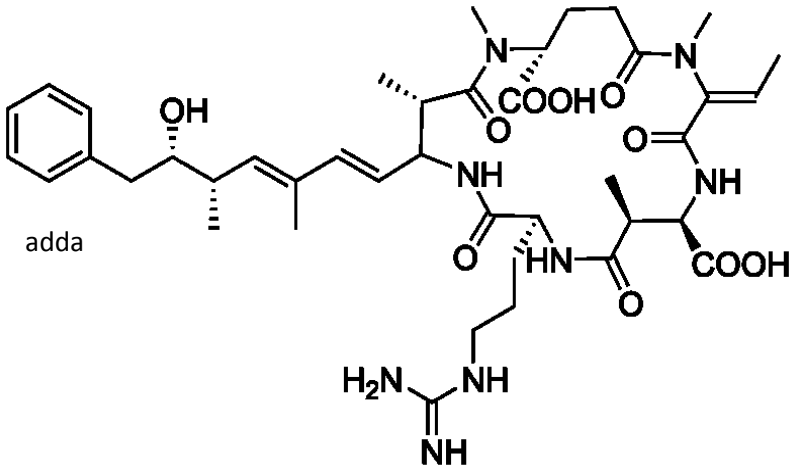


Figure 1.8 Chemical structure of nodularin-R

Both microcystin and nodularin are potent mammalian protein phosphatase type 1 and type 2A inhibitors (Runnegar et al. 1993). The activity appears primarily mediated by the adda group. The LD_{50}'s of microcystin LR, the most abundant analogue, and nodularin are 50 µg/kg and 60 µg/kg (intraperitoneal i.p, mice, in 24 hr) respectively. The primary acute effects experienced are intra-hepatic haemorrhage, as the liver cells shrink and subsequently separate allowing blood to accumulate. Death may result from hypovalemic shock and initial symptoms are reflective of blood loss (vertigo, vomiting, breathlessness, peripheral shutdown) (Runnegar and Falconer 1986; Runnegar et al. 1993).

On a global scale, microcystins are perceived as the greatest threat to public health in potable and recreational waters, a provisional guideline established by the World Health Organisation (WHO) of 1 µg L^{-1} for microcystin-LR has been accepted by most water authorities (WHO 2004). In Australia, the recommendation is 1.3 µg L^{-1} of total microcystins, expressed as toxicity equivalents of microcystin-LR for drinking water (NHMRC/NRMMC 2004).

.

Microcystins are produced by a number of freshwater species including *Microcystis aeruginosa*, *Planktothrix agardhii*, genera of *Anabaena* (Harada et al. 1991; Sivonen et al. 1992a), and *Nostoc* (Sivonen et al. 1992b) . Benthic cyanobacteria, *Haphalosiphon* (Fewer et al. 2007) and *Oscillatoria* spp. (Izaguirre et al. 2007) have also been identified as

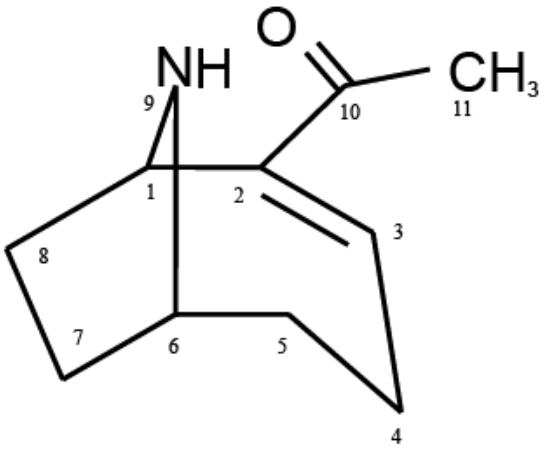
producers. The brackish water cyanobacteium *Nodularia spumigena,* has to date been the only identified producer of nodularins (Moffitt and Neilan 2001).

1.3.2 Neurotoxins

1.3.2.1 Anatoxin-a (homoanatoxin-a) and anatoxin-a(s)

Anatoxin-a (homoanatoxin-a) are alkaloid toxins (Figure 1.9) with the structure, 2-acetyl-9-azabicyclo (4-2-1) non-2-ene (acetyl at C2 is substituted for a propionyl). These toxins are nicotinic acetylcholine receptor antagonists, which lead to loss of muscle coordination, convulsions and death due to synaptic depolarisation. The LD_{50} (i.p. mice) for these toxins is 200 µg/kg and they have been attributed to dog deaths (Wood et al. 2007a). They have been identified in *Aphanizomenon issatschenkoi* (Wood et al. 2007b), *Anabaena* spp. (Bruno et al. 1994), *Planktothrix rubescens* (Viaggiu et al. 2004), *Rhaphidiopsis mediterranea* (Namikoshi et al. 2004) and benthic *Phormidium favosum* (Gugger et al. 2005).

Anatoxin-a(s) is a phosphorylated cyclic N-hydroxyguanine (Figure 1.10). An acetylcholinease inhibitor, it has an LD_{50} (i.p. mice) of 20 µg/kg and elicits similar effects to that of anatoxin-a. It is also associated with increase in salivation and lacrimation. The main producer species are *Anabaena flos-aquae* and *Anabaena lemmermanni* and the toxin has been implicated in the death of wild birdlife (Onodera et al. 1997).



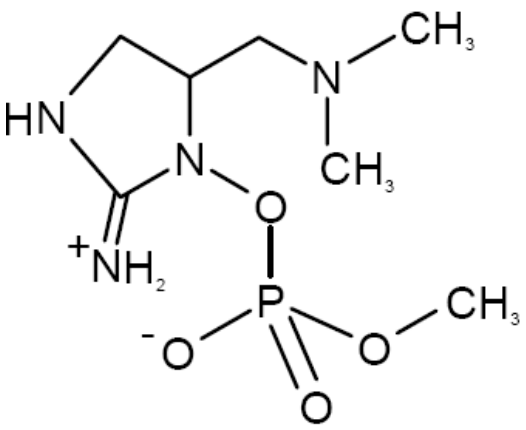


Figure 1.9 The structure of anatoxin-a
Acetyl is replaced at C2 with propionyl for homoanatxoin-a

Figure 1.10 The structure of anatoxin-a(s)

1.3.2.2 Saxitoxin

Saxitoxins are alkaloids with over 30 different structural variants (Figure 1.11). The structural backbone is based on a 1,3,4 trialky tetrahydropurine which can undergo various functional modifications (sulphated, carbamylated or N-sulphocaramylated) (Schantz et al. 1975).

The LD_{50} (i.p. mice) for saxitoxin is 10 µg/kg, though the potency of the toxin varies between analogues. The toxins are sodium ion channel antagonists that disrupt communication between neurons and muscle cells, resulting in numbness, paralysis and subsequent death by respiratory arrest (Humpage 2008).

Several freshwater cyanobacteria have been identified as producers, including *Anabaena circinalis* from Australia (Humpage et al. 1994), *Lyngbya wollei* from North America (Onodera et al. 1997), *Cylindrospermopsis raciborskii* from Brazil (Lagos et al. 1999), *Aphanizomenon* spp. (Mahmood and Carmichael 1986; Pereira et al. 2000) and *Planktothrix* spp. (Pomati et al. 2000).

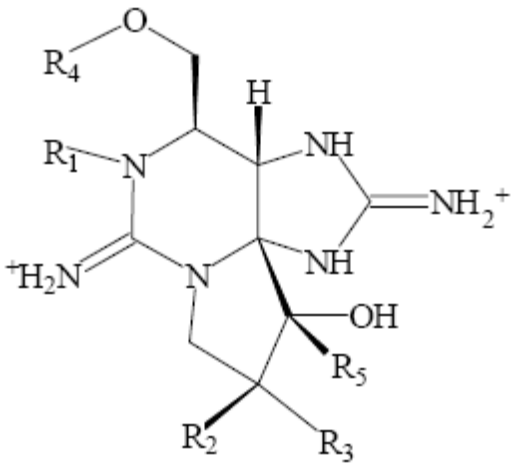


Figure 1.11 The general structure for saxitoxin

R_1 is H or OH; R_2 is H or OSO_3^-; R_3 is H or OSO_3^-; R_4 is H, $CONH_2$, $CONHSO_3^-$; R_5 is H or OH

1.3.3 Cytotoxic

1.3.3.1 Cylindrospermopsin

Cylindrospermopsin (CYN) is a novel alkaloid compound incorporating a tricyclic guanidine group in conjunction with a hydroxylated uracil moiety (Ohtani et al. 1992) (Figure 1.12). Naturally occurring analogues have since been isolated, including deoxy-cylindrospermopsin (Norris et al. 1999) (Figure 1.13) and 7-epi-cylindrospermopsin (Banker et al. 2000) (Figure 1.14). Whilst preliminary investigations suggested the hydroxyl on the uracil bridge was crucial for toxicity, all analogues have since been revealed to be either equivalent or comparable in toxicity to that found for CYN itself (Banker et al. 2000; Looper et al. 2005). These findings suggest the need to look at total toxicity as additive when referring to cylindrospermopsin and its congeners.

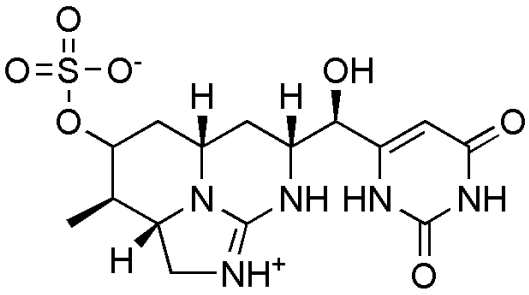


Figure 1.12 Chemical structure of cylindrospermopsin

Indicating correct stereochemistry of C7 hydroxyl on the uracil bridge (Heintzelman et al. 2002)

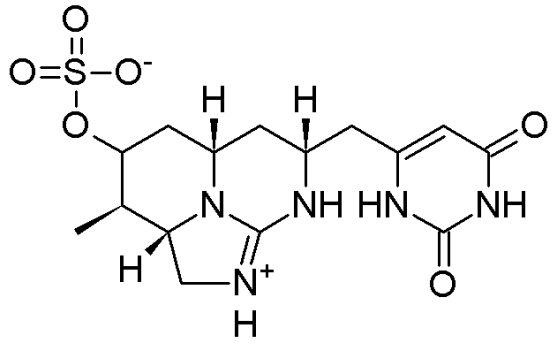


Figure 1.13 Chemical structure of deoxy-cylindrospermopsin

The hydroxyl on the uracil bridge is lost at position C7.

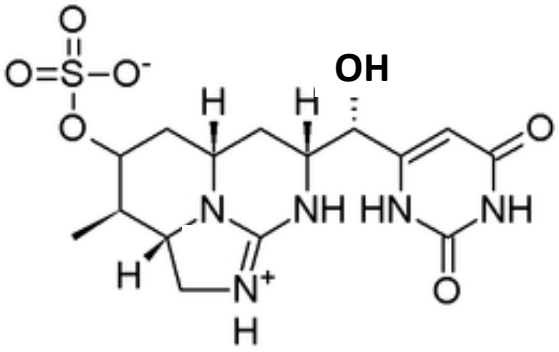


Figure 1.14 Chemical structure of 7-epicylindrospermopsin

Indicating correct stereochemistry of C7 hydroxy on the uracil bridge (Heintzelman et al. 2002)

Biochemical and toxicological investigations reveal CYN to be a potent inhibitor of protein synthesis (Terao et al. 1994) and glutathione (Runnegar et al. 1995), leading primarily to heptacellular degeneration and variable degrees of necrosis. CYN poisoning is also characterized by effects on the kidney, heart, lungs, spleen and thymus. The LD_{50} of purified CYN, after intraperitoneal injection into mice, was demonstrated to be 2.1mg/kg over a 24hr period and 0.2 mg/kg over 6 days (Ohtani et al. 1992). Intriguingly, the administration of cyanobacterial extracts containing CYN revealed higher toxicity in both mice and *Artemia salina*, than equivalent doses of purified CYN, suggesting either the

presence of additional unidentified toxins or synergistic effects (Metcalf et al. 2002). Cylindrospermopsin is regarded as a mutagen and preliminary evidence suggests it causes carcinogenicity in mice (Falconer and Humpage 2001). Cylindrospermopsin has additionally demonstrated a potential to be a dermatological irritant, with studies in mice revealing a delayed-type hypersensitivity when exposed to purified CYN and cyanobacterial extracts containing CYN (Stewart et al. 2006a) . Preliminary human studies were complimentary to these findings, and showed that whilst the cutaneous reactions were atypical within the sample set, risks were evident for people with an established history of skin hypersensitivity (Stewart et al. 2006b). This may translate to an additional exposure route requiring consideration by regulatory bodies for users of recreational waters.

Cylindrospermopsin was first isolated from *Cylindrospermopsis raciborskii* (Woloszynska) Seenaya Subba Raju, order Nostocales. The discovery followed a comprehensive post mortem, delving into the "Palm Island mystery" (Byth 1980; Bourke et al. 1983; Hawkins et al. 1985). This incident details an acute human poisoning episode following ingestion of potable water from the local supply. The reservoir had recently experienced an algal bloom with subsequent algicidal treatment. The ingestion of the contaminated water resulted in 148 inhabitants being hospitalised with acute gastrohepatitis. This incident highlighted the impact that this potent toxin could have on human health. Furthermore, CYN has also been associated with mortality in both wild and domesticated animals (Saker et al. 1999). Increasingly, additional species are catalogued as CYN producers worldwide, most found within the order of Nostocales; *Aphanizomenon ovalisporum* (Banker and Carmeli 1997), *Anabaena bergii*, *Rhadiopsis curvata* (Li et al. 2001), *Anabaena lapponica* (Spoof et al. 2006) and *Aphanizomenon flos aquae* (Preussel et al. 2006), with the exception of one from the order Stignomentales; *Umezakia natans* (Harada et al. 1994) and more recently the benthic Oscillatoriales *Lyngbya wollei* (Seifert et al. 2007). Currently, there are limited data to sufficiently recommend guideline values for CYN in potable and recreational waters. Conservatively, the guidelines for CYN have been recommended to be reflective of WHO guidelines for microcystin-LR (Humpage and Falconer 2003).

1.4 Current identification strategies

An important requirement for ensuring consumer safety from exposure to cyanotoxins is the availability of robust monitoring and analytical protocols, to both identify and quantify toxins. Screening strategies to identify problematic cyanobacterial populations currently employ skilled microscopists, who can readily identify common cyanobacterial genera and species. However, the production of toxin is strain-specific not species-specific, which poses a challenge to water quality managers, as inspection by microscopy is not able to assess the threat of toxicity (Ohtake et al. 1989). Further to this, confusion and misidentification can arise, as taxonomical classifications can be subjective decisions and morphological traits can demonstrate degrees of plasticity and ambiguity, often related to the environment they are isolated from (Wacklin et al. 2009). Indeed, there are mounting calls for revisions of cyanobacteria genera and species classifications via a more polyphasic approach, to amend discrepancies in taxonomy arising from phenotypic instability as well as conflicts observed between genotype and phenotype (Suda et al. 2002; Stüken et al. 2009; Wacklin et al. 2009).

Currently, quantitative analytical methods; high-performance liquid chromatography (HPLC) and liquid chromatography-mass spectrometry (LC-MS) and matrix laser desorption/ionisation time-of-flight (MALDI-TOF); offer sensitive and high throughput approaches. However, they are not economical or readily available to all water industry managers (Chorus and Bartram 1999; Sivonen 2008). Furthermore, sample collection and preparation continues to be an intensive process, along with a lack in availability of standards (Burch 2008; Sivonen 2008). At present, detection and quantification of CYN is primarily achieved through HLPC coupled with UV (Harada et al. 1994) detection or HPLC with tandem mass spectroscopy (HPLCMS-MS) (Eaglesham et al. 1999).

Enzyme-linked immunosorbent assays (ELISA) for saxitoxin and microcystins (Smith and Kitts 1994; Micheli et al. 2002; Yu et al. 2002) and *in vitro* assays (protein phosphatase inhibition (PPI) analysis) for microcystins and nodularins (Meriluoto et al. 2000; Metcalf et al. 2001) are also available. Whilst these assays are amenable to high throughput and are sensitive, they are not able to differentiate between compounds and quantification can be complicated by cross-reactivity with the various analogues (Carmichael and An 1999;

Sivonen 2008). Furthermore, some analogues of microcystin are not detected; therefore risks of underestimating toxicity are a possibility. A quantitative protein inhibition (PI) assay is also available to detect CYN. Deoxy-CYN is also detected and the end point is discussed as total toxin equivalence. The assay is at times sensitive to other protein inhibitors; however the slope for CYN in quite distinctive and diagnostic, therefore this does not present as issue (S. Froscio, personal communication). Whilst the assay is published it is still in the process of validation prior to establishing it as an alternative for routine use (Froscio et al. 2001).

There is a requirement for new methods to be developed that are both specific to the toxin and producer organism, rapid, sensitive and economical. Genetic methods have the potential for the development of specific screening tools to aid water quality authorities determine the toxin potential of a cyanobacteria bloom. However there is a prerequisite of prior molecular knowledge of the biosynthetic cluster to permit the development of primers or probes. Prior to commencing this work there was a dearth of fully characterised cyanobacterial biosynthetic clusters. Because of the potentially significant impacts on human health and stringent water quality requirements globally, molecular investigations have begun in earnest to identify the biosynthetic genes implicated for many of the toxic secondary metabolites.

1.5 Cyanobacteria Gene clusters

At the commencement of this thesis, of all the cyanotoxins with implications to public health in potable waters, only the microcystins (Nishizawa et al. 1999; Nishizawa et al. 2000; Tillett et al. 2000a; Christiansen et al. 2003; Rouhiainen et al. 2004) and the structurally related nodularin (Moffitt and Neilan 2004) had been fully characterised at the molecular level. In fact, microcystin was the first cyanobacterial peptide synthetase identified and verified by mutagenesis (Dittmann et al. 1997). Since then, several other cyanobacterial secondary metabolites have been characterized at the molecular level and all have been found to belong to structurally diverse polyketides and/or non ribosomal peptides (Table 1.2).

Table 1.2 Overview of fully sequenced cyanobacterial secondary metabolite biosynthetic clusters

Toxin	Effect	Size of biosynthetic cluster (Kb)	Structural type	Species*	Reference
Barbamide	molluscicidal	26	12 open reading frames (ORFs) encoding 4 NRPS, 1 integrated NRPS/PKS, thioesterase type II; tailoring enzymes; flanking transposases	*Lyngbya majuscula* L19 (marine)	(Chang et al. 2002)
Curacin A	cytotoxic	64	14 ORFs encoding 9 PKS; 1 integrated NRPS/PKS; tailoring enzymes	*Lyngbya majuscula* (marine)	(Chang et al. 2004)
Jamaicamide A	Neurotoxic (sodium channel blocking)	58	17 ORFs encoding 6 PKS; 1 NRPS; 1 integrated NRPS/PKS; tailoring enzymes including halogenase	*Lyngbya majuscula* JHB (marine)	(Edwards et al. 2004)
Lyngbyatoxin	dermatoxic	11.3	4 ORFs encoding 1 NRPS; tailoring enzymes	*Lyngbya majuscula* (marine)	(Edwards and Gerwick 2004)
Microcystin	hepatotoxic	55	10 ORFs encoding 5 NRPS; 1 PKS (2 modules); 2 integrated NRPS/PKS; tailoring enzymes; P. agardhii contains additional	*Microcystis aeruginosa* K139; *Microcystis aeruginosa* PCC 7806; *Planktothrix agardhii* CYA126/8; *Anabaena* 90	(Nishizawa et al. 1999; Nishizawa et al. 2000; Tillett et al. 2000a; Christiansen et al. 2003; Rouhiainen et al.

Table 1.2 Continued.

Toxin	Effect	Size of biosynthetic cluster (Kb)	Structural type	Species*	Reference
			thioesterase II		2004)
Nodularin	hepatotoxic	47	9 ORFs encoding 1 PKS (2 modules), 2 NRPS, 2 integrated NRPS/PKS; tailoring enzymes; flanking transposases	*Nodularia spumigena* NSOR10	(Moffitt and Neilan 2004)
Anabaenopeptilide	Serine protease inhibitor	30	4 ORFs encoding 3 NRPS	*Anabaena* 90	(Rouhiainen et al. 2000)

*Indicates which Genera the biosynthetic gene cluster has been characterised for.

1.5.1 Microcystin and Nodularins

The characterisation of the complete microcystin gene cluster from three distinct genera, along with the structurally related nodularin synthase, makes possible comparisons between their structural organisations (Figure 1.15 microcystin synthetase, *Anabaena* not shown). Furthermore, with the molecular architecture of the biosynthetic cluster being readily available, it optimises the selection of gene regions that can be developed into primers and probes to be either genus specific and/or toxin specific. There are a plethora of non-ribosomal peptides additionally produced by *Microcystis*, including aeruginosins (Ploutno et al. 2002), cyanopeptolins (Martin et al. 1993) and microviridins (Erhard et al. 1999). Therefore, when considering probe design, these similar compounds must be considered to ensure a gene unique to the microcystin structure is selected.

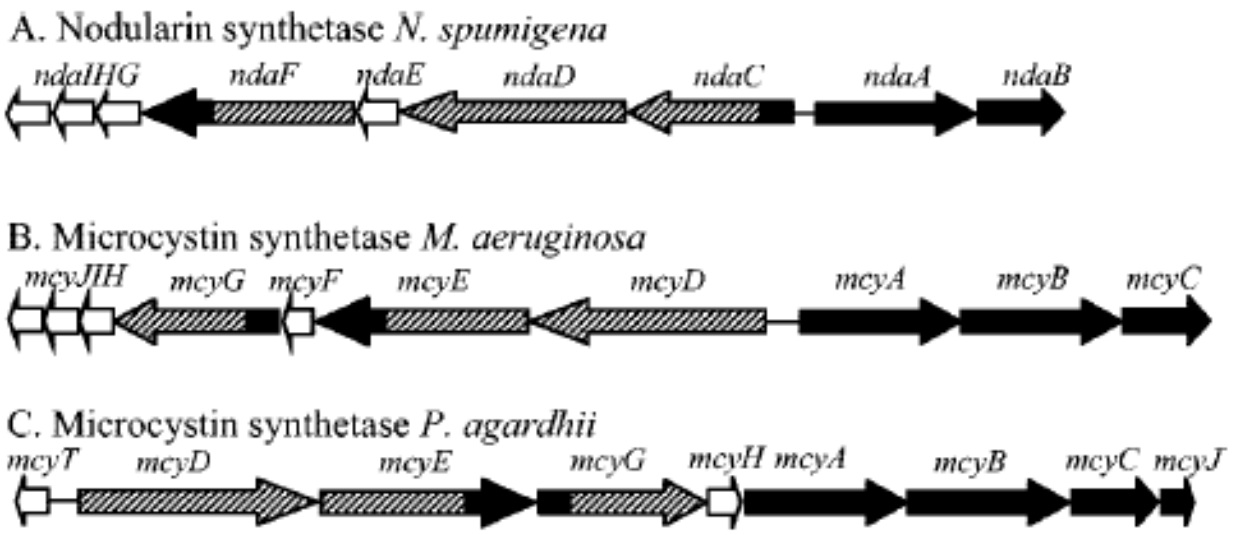


Figure 1.15 Structures of the nodularin (*nda*) and microcystin (*mcy*) gene clusters

PKS (hatched), NRPS (solid black) and tailoring enzymes (white). **A** Structure and gene organisation of the *nda* gene cluster isolated from *Nodularia spumigena* NSOR10 (Moffitt and Neilan 2004) **B** Structure and gene organisation of the *mcy* gene cluster isolated from two *Microcystis aeruginosa* strains (Nishizawa et al. 1999; Nishizawa et al. 2000; Tillett et al. 2000a) **C** Structure and gene organisation of the *mcy* gene cluster isolated from *Planktothrix agardhii* CYN126/8 (Christiansen et al. 2003). *Reproduced from Moffitt and Neilan (2004).*

1.5.2 Cylindrospermopsin

The identification and characterisation of many of the cyanobacterial biosynthetic clusters was supported by precursor feeding studies, whereby the incorporation of labelled putative substrates were able to resolve the origin of the natural products (Moore et al. 1991; Verdier-Pinard et al. 1999; Sitachitta et al. 2000). Precursor studies undertaken by Burgoyne et al (2000), in an effort to understand the biosynthesis of CYN, established that CYN was of polyketide origin because five acetate units were incorporated into its structure. This finding implicated the involvement of a polyketide synthase in assembling the CYN structural backbone. Furthermore, the precursor studies also identified a guanidinoacetic acid starter unit.

Subsequent work by Schembri et al (2001) searched for molecular markers that would definitively detect the CYN toxic phenotype and identified two genes fragments encoding a PKS and a PS polypeptide. These gene fragments demonstrated a strong correlation to the toxic CYN phenotype for the 13 *C. raciborskii* strains analysed by gene-directed polymerase chain reaction (PCR). Whilst the researchers were comfortable in implicating the involvement of the PKS, as indicated by the earlier precursor study, they remained ambivalent about the actual role of the PS in CYN production. Some resolution was accomplished when sequence analysis of an 11.3KB fragment isolated from a CYN-producing *Aphanizomenon ovalisporum* revealed a PS gene adjacent to a novel amidinotransferase gene and a PKS (Shalev-Alon et al. 2002). The amidinotransferase (designated *aoaA*) is responsible for the amidination of glycine in other systems, and appeared likely to be responsible for the incorporation of glycine into the guanidinoacetic acid starter unit of CYN. The close proximity of the PS and PKS to the *aoaA* gene provided indirect evidence that they were likely participants in CYN production, however without gene inactivation analysis this was not an unequivocal conclusion. Nevertheless, the benefits of identifying the PKS and PS gene fragments and their strong correlation to CYN toxicity showed the potential for their application to nucleic acid-based assays to facilitate the rapid detection of CYN-producing cyanobacteria.

1.6 Molecular methods for identification of potentially toxic organisms.

The potential benefit of nucleic acid-based testing is to reduce analysis time, allow high throughput, and provide accurate, sensitive and specific results. The clinical diagnostic field has long embraced these methodologies and real-time PCR and microarray-based technologies are routine in many clinical laboratories (Csako 2006; Weile and Knabbe 2009). Recently, with the availability of the microcystin synthetase sequence, the water industry has begun to explore these techniques as an alternative approach to current detection methods. Some of the more routine methodologies currently being developed are briefly detailed below.

1.6.1 Conventional PCR

The polymerase chain reaction (PCR) has advanced the study and analysis of molecular genetics (Saiki et al. 1985). Oligonucleotide primers and probes can be designed from known DNA sequences, promoting specificity for a region of interest. The short single stranded primer pair binds to a DNA target and facilitates the amplification of the intervening region between them, using thermostable DNA polymerase enzymes (*taq* DNA polymerases) (Saiki et al. 1988). There are now a myriad of *taq* DNA polymerases available with varying fidelities, speed and proof reading abilities; the latter dependant on the presence or absence of 3'→5' exonuclease activity. The instrumentation required to perform thermocycling and gel electrophoresis, are readily accessible to most routine laboratories.

However, the limitation of conventional PCR is the requirement of end-point analysis to visualise the amplified DNA, normally through gel electrophoresis. This increases the likelihood of laboratory contamination and furthermore band intensities do not permit accurate or easy quantification (quantification can be achieved by conventional PCR but it generally has a narrow dynamic range). There is also the issue of DNA extraction and purification from environmental samples. Environmental samples can contain inhibitors, which potentially generate false negative results; this ideally needs to be addressed within

the assay to enhance its applicability as a screening tool (Nonneman and Zimba 2002). Regardless, conventional PCR has been instrumental in developing rapid and specific molecular-based screening assays toward the detection of toxigenic cyanobacteria in environmental waters.

Molecular methods using conventional PCR have primarily been centred on microcystin synthetase and its producer organisms, due to the availability of the sequences. One of the initial assays targeted the N-methyltransferase region in *mcyA* (Tillett et al. 2001). The *mcyA* gene is responsible for the activation of N-methyl-dehydroalanine, and homologous regions in other organisms have been identified as providing protection of the peptide against proteolytic cleavage (Marahiel et al. 1997). Whilst the initial screening of over 38 cultures of *Microcystis* sp., along with several field samples, demonstrated correlation between the PCR and toxicity, later field studies challenged its robustness (Baker et al. 2001). The field trials demonstrated that PCR fragments were also detectable from non-microcystin producing strains. This perhaps is suggestive that the NMT region of *mcyA* is not the best candidate for determination of microcystin potential.

Another assay has targeted the *mcyE* and *mcyD* genes, which play an essential role in the biosynthesis and incorporation of the adda and glutamic acid into the structure (Rantala et al. 2004). These components are strongly linked to the potency of the compound and thus the nucleotide sequences of these regions are relatively conserved across genera. As adda is also essential to nodularin, these genes did not discriminate between microcystin and nodularin producers. However, as the purpose of this study was to investigate the phylogenetic origin of microcystin in the genome, it was aiming to detect all producers of microcystin and nodularin.

A multiplex assay to screen for more than one microcystin toxigenic marker was trialled on whole cells of *Microcystis* by (Ouahid et al. 2005). Overall, six primer sets were designed, three of the sets targeting the NRPS genes, *mcyA*, *mcyB*, *mcyC* and three targeting the PKS genes, *mcyD*, *mcyE*, *mcyG*. Simultaneous detection was possible for any combination of primers pairs as sets of three. All primers selected showed specificity to the toxic phenotype. However, the assay lacked the use of internal controls to assist in its detection

of false negatives when using environmental samples. Moreover, other microcystin-producing species were not examined to ascertain the specificity of the probes across other genera.

Fergusson and Saint (2003) developed a multiplex PCR based on the work of Schembri et al (2001) and (Wilson et al. 2000) to design species-specific primers for *C. raciborskii*. The assay was extended to target both the toxic-related *pks* and *ps* genes and a region of DNA dependant polymerase *rpoC1* that showed high specificity to the cyanobacterium *C. raciborskii*, a dominant CYN producer in Australian water bodies. Furthermore, the methodology incorporated an internal control, which is vital in detecting PCR inhibition when dealing with environmental samples. The assay proved robust and was able to determine and discriminate between CYN-producing cyanobacteria and the presence of *C. raciborskii* from both culture and environmental samples.

However, as was evident in molecular-based assays developed toward microcystin-producers, care must be taken when selecting target regions, so that they are unequivocal in their detection of toxicity. Knowledge of all genes enhances the potential to find the most discriminate marker. Prior to the commencement of the work undertaken in this thesis, the mechanism of CYN biosynthesis remained poorly understood, leaving open the possibility that the most robust molecular markers for CYN toxicity are yet to be found.

1.6.2 Real-time PCR

The development of fluorescent probes and double-stranded DNA (dsDNA) specific fluorescent intercalating dyes, coupled with amplification methodologies, pioneered the dawn of quantitative real-time PCR (qPCR). Quantitation by q PCR is possible due to the proportional relationship between starting template and the cycle at which exponential amplification is detected. In the case of dsDNA dyes and some probe technologies, it is able to discriminate sequences when DNA melting analysis is employed. Amplification products will melt at different temperatures, based on their length, base composition and the absolute order of bases (Figure 1.16), which can lead to unique melt peak profiles

(Monis and Giglio 2006). The melt peaks that result are analogous to the endpoint visualisation of conventional PCR, whilst removing post amplification handling.

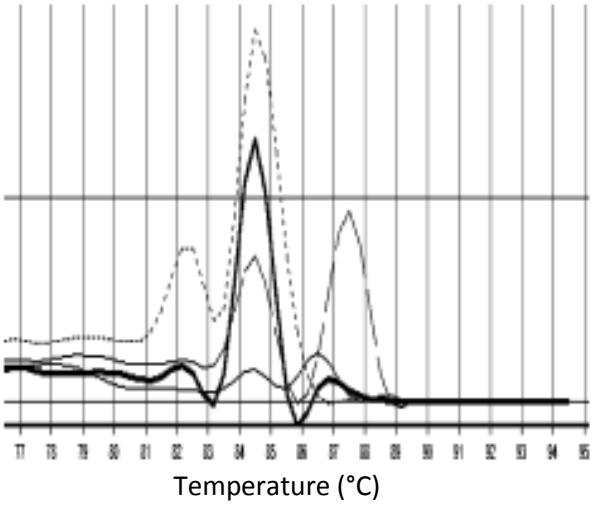


Figure 1.16 Real-time melting curve analysis

Illustrating the differntiated melting curves of geosmin synthase fragments amplified from *Anabena circinalis* (solid line), *Anabaena laxa* (dotted), *Phormidium sp*. (large dashes) *Geitlerinema* (thin line) *Reproduced from (Giglio et al. 2008)*

Several types of fluorescent chemistries are available, including molecular beacons, fluorescence resonance energy transfer (FRET), minor groove binding and taqman probes. The most common probe formats used to detect cyanotoxins to date are intercalating dyes such as SYBER Green I (Garner et al. 1994) and SYTO9 (Monis et al. 2005) and taqman methods.

Intercalating dyes bind to the minor groove of double DNA. They are not sequence specific therefore they bind to both desired products and non specific amplification, forming primer dimers. This probe format is often selected when using melt curve analysis as the endpoint analysis. (Figure1.18a and b)

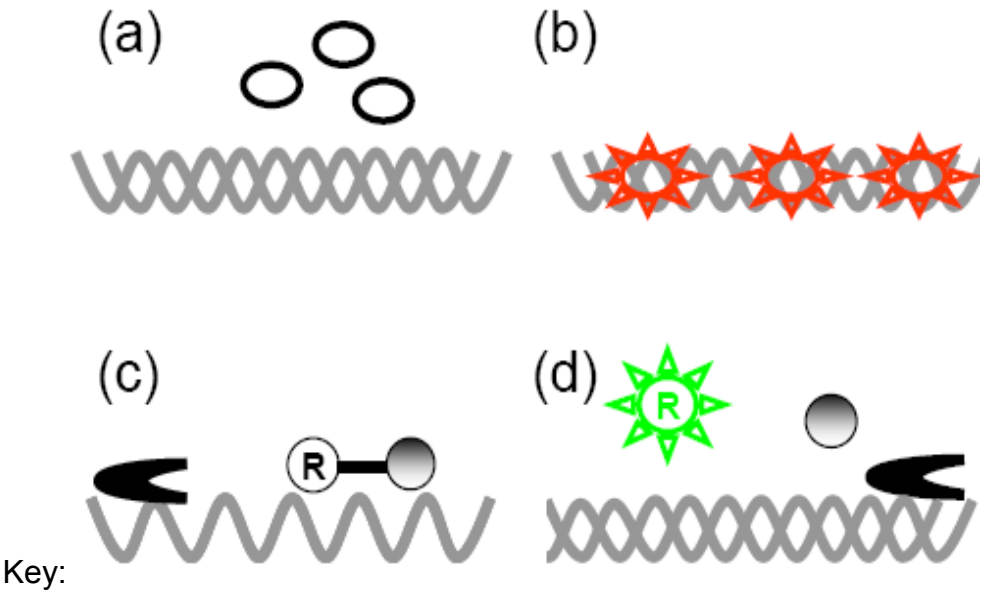
Taqman probes are regarded as more specific. They are dual-labelled hydrolysis probes that are initially not fluorescent. By utilising the 5' nuclease activity of *taq* polymerase,

they bind to the internal strand of the target sequence and are degraded during primer extension. This releases the fluorophore from the quenching molecule and fluorescence is emitted (Figure 1.17c and d)

Quantitative PCR has been employed to generate a profile of a microcystin-producing bloom and provide insight to its toxic potential over the blooms time course. Vaitomaa et al (2003) were able to take advantage of the sensitivities of qPCR and develop probes targeting the *mcyE* gene. Further to this, by exploiting the molecular knowledge, sequence differences between *Microcystis* and *Anabaena* microcystin synthetase supported the design of genus-specific probes. The use of the specific sequence enabled the relative contribution of each toxic species to be calculated against the total microcystin and cyanobacterial cell count for the water bodies. *Microcystis* was shown to be 30-fold more abundant than *Anabaena*, which is important with respect to selecting restoration strategies to reinstate ecological balance.

A similar qPCR assay was also used to determine the toxin potential of a *Microcystis* sp. bloom in a German water body (Kurmayer and Kutzenberger 2003). A multiplex methodology was adopted to compare *mcyB* gene copy numbers to a *Microcystis* specific phycocyanin sequence, in order to establish the toxic potential of the bloom. The findings by PCR were compared to cell counts to estimate the accuracy of qPCR and a good correlation was observed.

Quantitative real-time PCR promises to provide three needs for water quality managers. Firstly, implementation of fluorescent strategies facilitates multiplexing. This permits the development of assays that can be both genus specific as well as targeting a toxic marker. Secondly, the quantitative capability of qPCR permits not only the identification of the toxin but can estimate the toxicity potential of the blooms based on ratios between toxin-producing cells and cyanobacterial cell numbers. Finally, the closed-tube methodology is time efficient, sensitive and lends itself to be transferred to portable technology.



Key:

Intercalating dye

Fluorescing intercalating dye

Taq DNA polymerase

Reporter molecule

Quenching molecule

Fluorescing report molecule

Figure 1.17 Schematic representation of commonly used real-time PCR chemistries

Unbound intercalating dye is not fluorescent (a) but will produce a quantum yield increase in fluorescence upon binding to a dsDNA. Taqman probes cannot fluoresce when intact due to the proximity of the reporter and quencher molecule (c) but will produce fluorescent signal following hydroylsis by *Taq* polymerase and release of reporter molecule (d). *Reproduced from Monis and Giglio (2006).*

1.7 Phylogenetic analyses

Further to the development of molecular tools for detection of toxic species, the genetic information can be used to reconstruct the evolution and confirm the taxonomy of an organism. The molecular code contains vast heterogeneity, including base composition variation, coding preferences and hierarchies of repeats which can be exploited to resolve phylogenetic relatedness between genera, species and strains. The resolution obtained is dependent on the genetic markers selected. Often the taxonomy based on morphology does not correspond to the molecular phylogeny.

A number of genes have been used as evolutionary markers in the delineation of cyanobacterial phylogenetic relatedness. Generally, phylogenetic keys used for identification are "housekeeping" genes that are involved in processes critical to cell survival and therefore are under functional constraints and selective pressure. Woese et al (1987) identified 16S rRNA as a powerful phylogenetic marker molecule, or "molecular chronometer", and selected the gene to organise microorganisms into a phylogenetic system, according to their evolutionary descent. The first global phylogenetic analysis of the cyanobacteria comparing 16S rRNA gene sequences was put forward by (Giovannoni et al. 1988). However, due to its conservation, 16S rRNA lacks the resolving power to discriminate between species of a genus or strains of a species (Wilmotte 1994). Therefore, to investigate beyond the genus level, alternative markers for identification purposes have been explored with varying degrees of success. Within cyanobacterial phylogeny, alternative markers that have provided greater resolution at species and strain level include the phycocyanin gene (*cpc*) (Neilan et al. 1995; Haande et al. 2007), the 16S-23S internal transcribed spacer (ITS) and 23S rRNA gene (Iteman et al. 2000; Boyer et al. 2001; Gugger et al. 2005; Haande et al. 2007), DNA-dependent RNA polymerase gene (*rpoC1*) (Palenik and Haselkorn 1992; Palenik and Swift 1996; Wilson et al. 2000) and the *nifH* gene (Tamas et al. 2000; Dyble et al. 2002). These have previously been used in both phylogenetic evaluations and nucleic acid-based assays.

Cyanobacterial blooms are frequently a composite of toxin producing and non-toxic strains. There is evidence to suggest that a toxic strain can be distinguished by the presence of a biosynthetic gene cluster, which is lacking within genomes of non-toxic strains

(Meißner et al. 1996; Schembri et al. 2001). How these secondary metabolite-encoding clusters are disseminated through cyanobacterial genomes is of interest, and may provide insight into resolving the possible ecological roles that these toxins fulfil, as well as improving our understanding of toxic bloom formation.

Current evolutionary investigations regarding cyanobacterial biosynthetic gene clusters have largely focussed on genes encoding the microcystin and nodularin synthetases. This is principally due to the abundance of available molecular data (Tillett et al. 2000a; Nishizawa et al. 2001; Christiansen et al. 2003; Moffitt and Neilan 2004; Rouhiainen et al. 2004). Phylogenetic evidence suggests that the genes encoding the microcystin synthetase were present in the last common ancestor of all microcystin producing species (Rantala et al. 2004). However, others have supported active intragenic recombination (Tanabe et al. 2004) or horizontal gene transfer (Tillett 2000b; Moffitt and Neilan 2004) as putative mechanisms of disseminating the microcystin and nodularin biosynthetic cluster through cyanobacterial genomes. Certainly sequence analysis of whole genomes of cyanobacteria reveals a high degree of repeat sequence, up to 100 base pairs in length, which confers plasticity to the genome and may serve as hot spots for intragenic recombination events (Frangeul et al. 2008).

Horizontal (or lateral) gene transfer (HGT) has been recognised as an influential force in shaping the evolutionary history and molecular landscape of prokaryotic genomes (Koonin et al. 2001; Boucher et al. 2003). These mechanisms permit the acquirement of niche-defining attributes and provide genetic variation, potentially leading to enhanced fitness of the organism/s involved. Agents facilitating HGT of naked DNA include infection by bacteriophages, mobile genetic elements such as plasmids or transposons; often the exchange is not unidirectional (Gophna et al. 2006). There are various methods available to investigate the probability that a HGT event has occurred within a genome. Observations of incongruence within gene phylogenies, evidence of patchy distribution of homologous genes throughout related genera, species and strains, and uncharacteristic nucleotide composition are often indicative of HGT events (Ochman et al. 2000).

Overall, it is felt the inherent flaws within methodologies used can severely underestimate the incidence of HGT. Eventually amelioration of acquired genes occurs, resulting in compositional homogeneity and subsequent loss of the HGT signature. In addition, the success of HGT events would be better facilitated between intraphylum rather than interphylum members, due to similar environmental preferences and constraints through physiological or genomic mechanisms (Kunin et al. 2005; Zhaxybayeva et al. 2006).

There is ongoing debate about the extent that HGT plays in disseminating secondary metabolite biosynthetic clusters through cyanobacterial lineages. Currently there is little direct evidence of mechanisms that could arbitrate the transfer of naked DNA between freshwater cyanobacterial genomes. Interestingly, most biosynthetic clusters have redundant insertion-like sequences flanking them which have been surmised as the mechanism of transferring these clusters (Tillett et al. 2000a; Moffitt and Neilan 2004). Despite this, bloom-forming cyanobacteria have proved recalcitrant to transformation *in vitro*, with the resistance attributed to the number and activity of restriction endonucleases (Takahashi et al. 1996; Lyra et al. 2000). Recently, evidence of a putative DNA transfer mechanism was identified in the naturally transformable cyanobacteria, *Microcystis aeruginosa*, with the recent identification of a pilT proteinthat may mediate the uptake of naked DNA (Nakasugi et al. 2007).

1.8 Aims and objectives

From review of the literature it was apparent that rapid methods for detection of CYN toxic cyanobacteria are warranted. Early detection of the toxic potential of a cyanobacterial bloom facilitates the implementation of water management strategies at the earliest time to ensure public health remains secure. Molecular methods have the potential to offer toxin and species-specific approaches that are amenable to rapid and portable technologies. However, the molecular architecture for CYN biosynthesis has not been fully elucidated. The molecular knowledge of the biogenesis of CYN will enable the determination of gene targets that unambiguously detect toxin potential, in addition to providing a greater understanding of the importance of this compound to the organisms that produce it. The availability of the CYN molecular code may also aid in providing an understanding of the ecological drivers that promote its production, which will in turn facilitate future directions in establishing sufficient regulatory guidelines in regard to this toxin and advance proactive management strategies.

Therefore the general objectives of this thesis were:

1. To fully characterise the CYN biosynthetic cluster from *C. raciborskii*.

2. To use gene-directed PCR to interrogate the biosynthetic cluster for the most reliable molecular marker that defined CYN toxicity across genera of CYN-producing cyanobacteria.

3. To gain preliminary insights into the molecular mechanics of a CYN-producing cyanobacteria.

4. To investigate the evolutionary history of the CYN biosynthetic cluster within cyanobacterial lineages

A further six chapters are presented in this thesis. The second chapter presents the general material and methods applicable to the whole thesis. The third and fourth chapters describe

the characterisation of the CYN biosynthetic cluster and the identification of a suitable CYN determinant for application in future molecular strategies. The fifth chapter tracks a CYN-producing *C. raciborskii* bloom in the River Murray. The sixth chapter explores the mechanism of dissemination of the CYN biosynthetic cluster in cyanobacterial genome, using phylogenetic analyses to examine the congruency between primary metabolism and the CYN secondary metabolite genes. Finally the seventh chapter discusses the significance of the results presented in this thesis.

2 GENERAL MATERIALS AND METHODS

The materials and methods outlined in this chapter describe procedures that are general methods used throughout this work. Material and Methods that pertain to only one of the following research chapters are described in the specific material and methods of the relevant chapter.

2.1 Reagents and Chemicals

All reagents and chemicals used throughout this thesis were of molecular or analytical grade unless otherwise specified.

2.1.1 Antibiotics

Ampicillin (Amp) was purchased from the Sigma-Aldrich Chemical Company (St Louis, MO, USA). Amp was prepared as a stock solution of 100 mg/ mL in Sterile Milli-Q water (SMQ), filter sterilised through a 0.2 μm filter (Pall Life sciences, Ann Arbor, Mi, USA) and stored at -20 °C until required. Antibiotics were added to liquid or solid agar as specified in relevant sections.

2.1.2 Buffers

All buffers were prepared in Milli-Q water (Millipore Ultra-Pure System) and were either steam sterilised at 121 ºC for 20 min or filter sterilised through a 0.2 μm filter, unless otherwise indicated.

2.2 Culture Media

All media was prepared in Milli-Q water (Millipore Ultra-Pure System) and were steam sterilised at 121 ºC for 20 min.

2.3 Cultures

2.3.1 Maintenance of cultures

All cyanobacterial cultures used during these experiments were grown in a 50 cm^2 Greiner Bio-one Tissue culture Flask with vented lid (Frickenhausen, Germany), containing 20-30 mL of either Artificial Seawater Medium (ASM-1) or ASM-1 with nitrogen removed and molybdenum added (ASM-1 –N +Mo), pH adjusted to 7.6 (Table 2.1). Cultures were incubated at 20 °C under photon irradiance of 20 $\mu Mm^{-1}s^{-1}$ provided by cool white.

Table 2.1 Preparation of ASM-1 and ASM-1 N+Mo (Gorham et al. 1964)

Compound	Formula	Final Conc. mg/L	
Sodium nitrate	$NaNO_3$	170	ASM-1
Sodium molybdate	Na_2MoO_4	0.01	ASM-1-N+Mo
Potassium phosphate	K_2HPO_4	17.4	
Di sodium hydrogen phosphate	Na_2HPO_4	14.2	
Magnesium chloride	$MgCl_2.6H_2O$	40.62	
Magnesium sulphate	$MgSO_4.7H_2O$	49.33	
Calcium chloride dihydrate	$CaCl_2.2H_2O$	29.4	
Ferric chloride	$FeCl_3.6H_2O$	1.0835	
Boric acid	H_3BO_3	2.47	
Manganese (II) chloride	$MnCl_2.4H_2O$	1.3683	
Zinc chloride	$ZnCl_2$	0.44	
Sodium EDTA	Na_2EDTA	6.64	
Cobaltous sulphate	$CoSO_4.7H_2O$	0.0216	
Copper (II) chloride	$CuCl_2.2H_2O$	0.00013	

2.3.2 Toxin analysis

Cyanobacterial cultures for CYN analysis were prepared as cell free extracts, obtained through sonication. Sonication was performed on ice using a Branson sonifier (Branson Ultrasonics Corporation, Danbury, Ct, USA) at an amplitude of 60 % and constant output for 20 s, followed by a 40 s rest and filtration through a 0.2 µm nitrocellulose filter (Pall Life sciences, Ann Arbor, Mi, USA).

Determination of cylindrospermopsin in cell free extracts were performed at either the Queensland Health Scientific Services using reverse phase HPLC/MS-MS under protocols established by (Eaglesham et al. 1999) or at the AWQC Laboratories, SA Water Corporation utilising LC/MS/MS following in-house methods similar to those described in (Torokne et al. 2004). Alternatively, Protein Inhibition (PI) studies were performed to analyse CYN toxicity in some cyanobacterial extracts. Protein inhibition assays were performed by Melody Lau at the AWQC utilising protocols established by (Froscio et al. 2001).

2.4 DNA Extraction Protocols

2.4.1 Genomica DNA extraction

Cyanobacterial cultures ranging from 20 to 50 mL, or environmental waters from the River Murray, were pelleted by centrifugation on a Beckmann Coulter Allegra X-12R at 3270 × g, 4 °C for 20 min after which the supernatant was removed by gentle suction. Genomic DNA purification was performed using the Qiagen DNA minikit (Qiagen GmbH, Hilden, Germany) following the manufacturer's protocol. Genomic final DNA was eluted in 50 µL SMQ water.

2.4.2 Plasmid extraction

Following the cloning reactions described in section 2.6.2 transformed *E. coli* were cultured in 10 mL of LB broth [10 g/L tryptone, 5 g/L of yeast extract, 10 g/L of NaCl, SMQ-H_2O to 1L (pH 7.5)] , including 100 µg / mL of Amp., overnight at 35 °C with shaking at 180 rpm. Cells were then pelleted by centrifugation using a Beckmann Coulter Allegra X-12R at 3270 × g, for 20 min at 4 °C and the resulting supernatant removed by gentle suction. Plasmid DNA was extracted using the Qiaprep Spin Mini prep Kit (Qiagen GmbH, Hilden, Germany) following the manufacturer's protocol. The final DNA was eluted in 50 µL SMQ water.

2.4.3 Cosmid extraction

E.coli containing cosmid DNA were cultured overnight at 35 °C with shaking (180 rpm), in 50 -100 mL of prewarmed LB broth containing 100 µg/mL of Amp. The *E. coli* cells were pelleted by centrifugation using a Beckmann Coulter Allegra X-12R at 3270 × g, for 20 min at 4 °C and the resulting supernatant decanted. Purification of cosmid DNA was then performed using the MOBIO Ultraclean Plasmid Prep Kit (Carlsbad, CA, USA) following the manufacturer's protocol. The DNA was eluted in the elution buffer supplied with the kit.

2.5 Molecular Techniques

2.5.1 Enzymes

Ampli*Taq*® Gold DNA polymerase, 25 mM $MgCl_2$ and buffer (Geneamp 10 × Buffer II, containing 100 mM Tris-HCl pH 8.3, and 500 mM KCl) used within the amplification reactions was supplied by Roche Applied Biosystems (Foster City, CA, USA).

Elongase® Enzyme Mix, 5 × Buffer A (containing 300 mM Tris-SO_4 pH 9.1, 90 mM $(NH_4)_2SO_4$ and 5 mM $MgSO_4$) and Buffer B (containing 300 mM Tris-SO_4 pH 9.1, 90 mM $(NH_4)_2SO_4$ and 10 mM $MgSO_4$) was purchased from Invitrogen (Carlsbad, CA, USA).

Platinum® *taq* DNA polymerase High Fidelity system, 50 mM $MgCl_2$ and buffer (10 × High Fidelity Buffer containing 600 mM Tris-SO_4 pH 8.9, 180 mM $(NH_4)_2SO_4$) was obtained from Invitrogen (Carlsbad, CA,USA).

T4 DNA ligase and 10 × buffer (300 mM Tris-HCl (pH 7.8 at 25 °C), 100 mM $MgCl_2$, 100 mM DTT and 10 mM ATP) was purchased from Promega (Madison, WI, USA).

Restriction endonucleases (*Acc*I, *Eco*RV, *Hae*III, *Hin*dII, *Hin*dIII, *Hpa*I, *Rsa*I, *Sma*I and *Ssp*I) were purchased from either Promega (Madison, WI, USA) or New England Biolabs (MA, USA) and used with the appropriate buffer as supplied by the manufacturer.

2.5.2 Oligonucleotides

All synthetic oligonucleotides used for PCR or sequencing purposes were purchased from Geneworks, Adelaide Australia, and are listed in the relevant chapters. Primers were initially rehydrated in sterile Nuclease free water (Promega, Madison, WI, USA) to 100 μM and stored at -20°C.

2.5.3 Conventional PCR

All conventional PCR reactions were run on a Perkin-Elmer Gene Amp 2400 series thermal cycler. PCR amplifications were performed as 50 μL reactions containing 0.5 μM of forward and reverse primers (Geneworks, Adelaide Australia), 1 × reaction buffer II, 2.5 mM of $MgCl_2$, 1 U of Ampli *Taq* gold DNA polymerase, 200 μM deoxynucleoside triphosphates (Promega) and approximately 10-15 ng of genomic DNA template. Negative controls consisted of an equal volume of nuclease free water with no DNA template. Thermal cycling protocols varied in accordance to the gene targets and primer sets used and are listed within the relevant sections. Upon completion of the amplification reaction, the product was visualised on 1.0 % agarose gel as described section 2.5.5.

2.5.4 Long template PCR

All PCR reactions that were anticipated to generate an amplification product of 5.0 kb or greater were performed using the Elongase *taq* polymerase system. This incorporated proof reading (3' to 5' exonuclease) activity, permitting editing of the nascent strand and correction for nucleotide misincorporation. Each 50 μL reaction included 0.5 × buffer A, 1.75 mM $MgCl_2$ within buffer B, 0.2 μM each of forward and reverse primer (Geneworks), 200μM of deoxynucleoside triphosphates (Promega, Madison, WI, USA), 1 μL of Elongase *taq* DNA polymerase, 2-5 μL of DNA template. Negative controls consisted of an equal volume of nuclease free water with no DNA template. Thermal cycling protocols varied in accordance with the gene targets and primers sets used and are outlined in the relevant sections.

Visualisation of amplified products was performed using 0.7 to 1.0 % agarose gel electrophoresis, as described in section 2.5.5.

2.5.5 Agarose gel electrophoresis of DNA

Visualisation of PCR products, restriction digests or genomic extracts was performed at room temperature using 0.7 % or 1.0 % (w/v) agarose gels, prestained with 1 × SYBR® safe DNA gel stain (Invitrogen Molecular Probes™, Eugene, Oregon USA) . Prior to loading, an aliquot of DNA product to be analysed was combined with an equimolar concentration (v/v) of 2 × loading buffer (10 % glycerol, 2 mM EDTA, 0.1 % xylene cyanol, 0.1 % bromophenol blue, Geneworks, Adelaide, Australia). Electrophoresis was then conducted in 1 × TBE buffer (0.1 % (w/v) Tris, 0.93 % (w/v) EDTA, 0.055 % (w/v) boric acid) at 100 V utilizing an OWL model B3 electrophoresis chamber (Thermo Fisher Scientific, Rochester, New York) connected to an EC250-90 Electrophoresis power supply (E-C Apparatus Corporation, St Petersburg, Florida, USA). Products were then visualised using a DR-45M Dark Reader transilluminator (Clare chemical Research Inc., Dolores, Co, USA) and the image captured using a Kodak DC290 digital camera fitted with a 49 mm lens. Documentation and further image analysis was performed using Kodak ID version 3.5 Image analysis Software (Eastman Kodak Company, Rochester, NY, USA).

The sizes of PCR products and restriction enzyme fragments were estimated by comparing the relative mobility with that of a 100bp, 500bp, SPP1/*Eco*R1, or λ-*Eco*RI/*Hin*dIII Molecular weight markers (Geneworks, Adelaide, Australia).

2.5.6 Bioanalyser

When required, PCR products were visualised and quantified using an Agilent Bioanalyser (Agilent Technologies, Waldbronn, Germany) consisting of the 1000 DNA chip kit and Agilent 2100 Expert software (Version B.02.06 S1418).

2.6 DNA Manipulation

2.6.1 Purification of PCR products

PCR reactions yielding single products of interest for subsequent cloning or sequencing purposes were purified using either the MOBIO Ultraclean DNA purification spin columns (MOBIO Laboratories Carlsbad, CA,USA) or the Montage PCR centrifugal filter devices (Millipore Corporation, Bedford, MA, USA) following the manufacturer's protocol. When multiple PCR products were evident, bands of interest were excised from the agarose gels and purified using the QIAEX II gel extraction kit (Qiagen GmbH, Hilden, Germany) following the manufacturer's protocols, except the DNA was eluted in 20 - 50 µL of SMQ water to minimise the possibility of salts interfering with downstream sequencing of the DNA.

2.6.2 *In vitro* cloning

Cloning of PCR products was performed using the pCR® 2.1 TOPO vector (Figure 2.1) as part of the TOPO TA Cloning® kit (Invitrogen) following the manufacturer's protocols. Briefly, cloning reactions was incubated on ice with α-select chemically competent *E. coli* (Bioline USA) for 30 min followed by a heat shock at 42 °C for 45 s and immediately returning to ice for 2 min. The cells were then resuscitated with 950 µL of SOC medium (2 % tryptone, 0.5 % yeast extract, 0.4 % glucose, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ (Invitrogen Carlsbad, CA, USA)) and incubation at 35 °C for 1 h with horizontal shaking at 180 rpm. Transformants were plated onto LB agar (10 g/L tryptone, 5 g/L of yeast extract, 10 g/L of NaCl, 15 g/L bacteriological agar, SMQ-H_2O to 1L (pH 7.5)) containing 100 µg/ mL amp and 100 µg/ mL X-gal and cultured overnight at 35 °C. Blue/white screening of colonies permitted identification of transformants containing an insert.

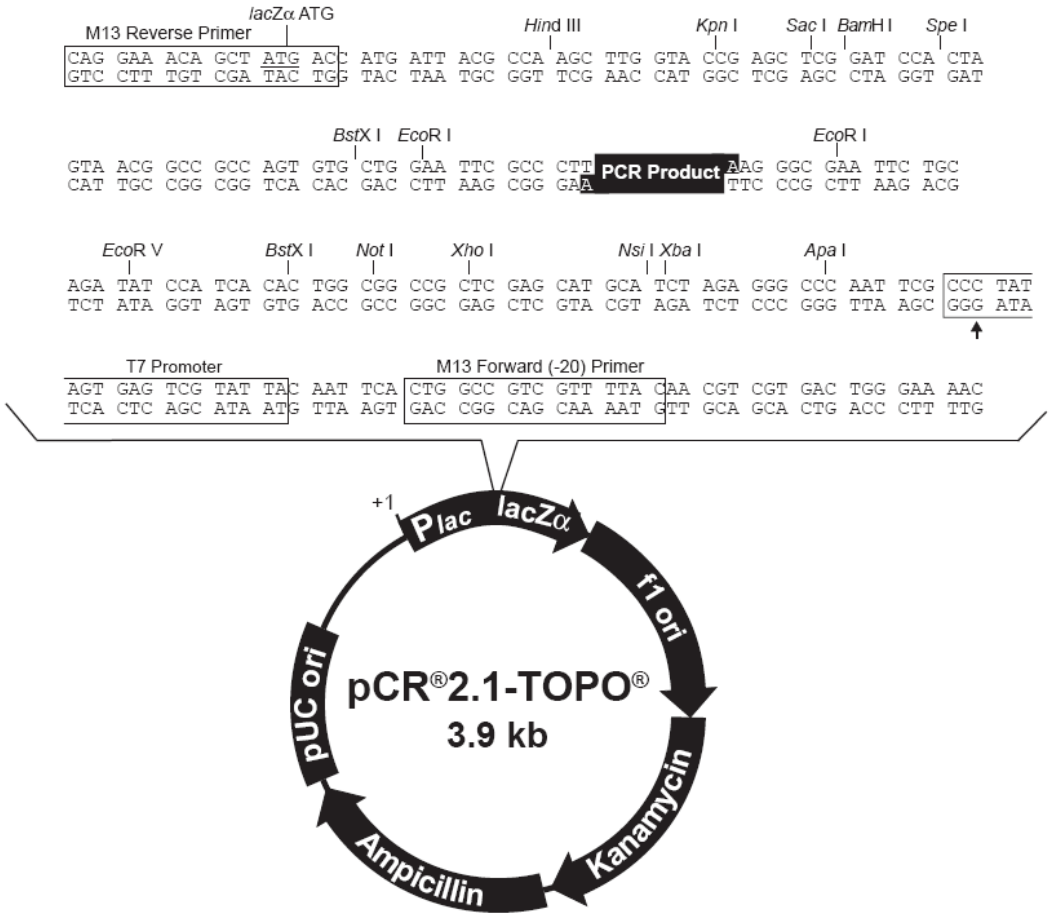


Figure 2.1 Map of pCR® 2.1-TOPO® (Invitrogen Corporation)

*LacZ*α fragment: bases 1-547. M13 reverse priming site: bases 205-221. Multiple cloning site: bases 234-357. T7 promoter/priming site: bases 364-383. M13 Forward (-20) priming site: bases 391-406. f1origin: bases 548-985. Kanamycin resistence ORF: bases 1319-2113. Ampicillin resistence ORF: bases 2131-2991. pUC origin: bases 3136-3809.

2.6.3 Restriction enzyme digestion

Genomic extracts (minimum 1 µg/reaction), obtained from the cyanobacterium *C. raciborskii* AWT 205, were digested for 2-3 hours at 37 ºC with a range of individual restriction endonucleases (0.25 U/reaction). Restriction enzymes used are detailed in the relevant chapters. Additional components of the reaction, comprised of the recommended buffer (1 ×) for each enzyme as supplied by the manufacturer, 0.1 µg of acetylated BSA and SMQ water to a final volume of 20 µL.

The reactions were inactivated at 65 ºC - 80 ºC for 20 mins; dependent upon the manufacturer's recommendations for individual enzymes; visualised on 1.0 % agarose for quantification and qualification purposes followed by purification using MOBIO Ultraclean DNA purification spin columns per kit protocols (MOBIO Laboratories Carlsbad, CA,USA). The DNA was eluted in 50 µL SMQ water.

2.6.4 Ligation

Purified, digested DNA (5 µL) was religated overnight at 10 ºC for overhanging cut sites or 15 ºC for blunt ended cuts; dependant on the restriction enzyme used; with 1 µL of T4 ligase buffer, 1 µL of T4 ligase DNA and SMQ water to a final volume of 10 µL.

2.7 Bioinformatics

Contiguous nucleotide sequences were assembled using KODON® version 2.0 software after manual editing. Sequence and amino acid analysis was performed within the KODON® program and additional analysis was also performed using NCBI (National Centre of Biotechnology Information) Blast server (www.ncbi.nlm.nih.gov/BLAST) and Tigr (www.tigr.org) for homologous genes and putative polypeptide functionality.

2.8 Phylogenetic Analyses

DNA sequences were aligned using CLUSTAL W embedded within the MEGA analysis software version 4 (Tamura et al. 2007). Alignments were edited manually to resolve ambiguous alignments. Pair wise deletion was used to account for gaps and sequence length variance. Neighbour joining trees were constructed using the Log Det, corrected algorithm (Tamura and Kumar 2002). Maximum Parsimony Trees were constructed within MEGA 4, utilising the close neighbour interchange with 11 Random tree additions search method. Maximum Likelihood Trees were constructed using FastDNAML obtained within the PHYLIP phylogenetic program (Felsenstein). Bootstrap analyses were performed for all trees to affirm branching confidence using 1000 replicates (Felsenstein).

2.8.1 Sequencing

Nucleotide sequencing was performed using Big Dye III terminator chemistry with the majority of sequencing being performed at the Australian Genomic Research Facility (agrf)

Brisbane, Australia however some initial sequencing was undertaken at the Institute of Medical and Veterinary Science (IMVS) Adelaide, Australia.

3 SEQUENCING AND STRUCTURAL ANALYSIS OF A GENE CLUSTER IMPLICATED IN THE PRODUCTION OF CYLINDROSPERMOSPIN

3.1 Introduction

Cylindrospermopsin (CYN) is a toxic, secondary metabolite produced by numerous, aquatic cyanobacterial species. Its presence in surface waters, threatens the integrity and utility of municipal, recreational and agricultural water supplies (Falconer and Humpage 2005; Falconer and Humpage 2006). Historically, the presence of CYN within Australian water bodies has resulted in dire consequences, with accounts of acute gastro-hepatic illness affecting a community on Palm Island, Queensland (QLD) (Byth 1980; Bourke et al. 1983) and livestock deaths (Thomas et al. 1998). In addition to the acute impacts on health, CYN is a mutagen (Humpage et al. 2000) and preliminary evidence suggests that long term exposure may have carcinogenic consequences (Falconer and Humpage 2001). It is therefore imperative that the presence of CYN and CYN-producing cyanobacteria are detected rapidly and unambiguously, to facilitate effective water management strategies and ensure minimal impact on the consumers. For this purpose, many water bodies are routinely monitored for bloom-forming, toxigenic cyanobacteria.

Current monitoring procedures toward identifying the presence of problematic cyanobacteria are through conventional microscopy. However, some morphological characters used to classify genera and species, can be broad and ambiguous complicating taxonomic assignment (Rippka 1988; Wilmotte 1994). Furthermore, toxin and non-toxin producing strains are not morphologically discernible, so additional analyses using physico-chemical methods are required to detect the presence of toxic compounds. The existing methodologies prove time consuming, labour intensive and expensive. Molecular-based methods are promising alternatives that offer rapid, sensitive, definitive and high throughput approaches, which lend themselves to portable devices that permit *in situ* analysis. However, to ensure the applicability of molecular-based approaches, robust molecular markers must be selected that are implicit with respect to toxicity.

Prior to the commencement of this work, no biosynthetic pathway had been definitively outlined for the production of cylindrospermopsin. Candidate polyketide synthase (PKS) and non ribosomal peptide synthetase (NRPS) genes had been elucidated from *C. raciborskii,* and proved to be synonymous with CYN-producing species and strains (Schembri et al. 2001). Homologues of both the PKS and NRPS were additionally identified in an Israeli CYN-producing *Aphanizomenon ovalisporum,* adjacent to a putative amidinotransferase, whose role is thought to involve the uploading of the guandinoacetic starter unit (Shalev-Alon et al. 2002). To date, probes targeting the PKS and PS genes have already been applied to molecular detection strategies, to determine the presence of CYN-producing cyanobacteria (Fergusson and Saint 2003; Rasmussen et al. 2008b). Moreover, these assays were designed to be specific toward a dominant producer in Australian waters, *Cylindrospermopsis raciborskii*.

However, without full consideration of the molecular architecture of CYN biosynthesis, it is impossible to ascertain whether the molecular probes selected represent an unequivocal toxigenic determinant. Whilst false positives may be better tolerated than false negatives in respect to potential impact on public health, a probe that is targeted toward a gene that is redundant within a genome may result in the overestimation of the toxin potential of a bloom. Thus, the motivation behind this research was to extend upon current knowledge of the putative CYN biosynthetic cluster through additional sequencing and construction of a gene assembly. As each new gene was identified, gene-directed PCR was initiated, with the principal aim being the eventual seclusion of indiscriminate toxigenic marker/s that could be adapted to molecular technologies to afford more rapid and accurate detection strategies in the future.

This chapter details the experimental approaches attempted, in an effort to fully determine the cylindrospermopsin biosynthetic gene cluster, from a dominant producer organism in Australian waters, *C. raciborskii*. The approaches employed, in preference to knockout studies, are also described to demonstrate the clusters implicit involvement in the biogenesis of cylindrospermopsin. Interrogation of the sequence data through gene-directed PCR, permits the identification of potential candidate genes toward unequivocal detection of CYN-producing cyanobacteria.

3.2 Materials and Methods

3.2.1 Cyanobacteria culture conditions

All cultures used during these experiments were maintained as described in Chapter 2 section 2.3.1. A summary of the isolates used in this chapter is outlined in Table 3.1.

Table 3.1 Summary of isolates analysed in Chapter 3

Species	Strain	CYN	Geographical origin	Reference	Notes
Anabaena bergii	ANA283A	+	Glenbawn Dam,NSW	AWQC	HPLC/MS-MS
	ANA283A*	-	Glenbawn Dam,NSW	AWQC	HPLC/MS-MS
	ANA283B	-	Glenbawn Dam,NSW	AWQC	HPLC/MS-MS
	ANA283D	-	Glenbawn Dam,NSW	AWQC	
	ANA360H	-	Tanunda, SA	AWQC	Fergusson (2003)
	ANA360B	-	Tanunda, SA	AWQC	HPLC/MS-MS
	ANA366A	-	Thornden Pk, SA	AWQC	Fergusson (2003)
	ANA366B	+	Thornden Pk, SA	AWQC/	Fergusson (2003)
	ANA372E	-	Bundaleer Reservoir	AWQC	HPLC/MS-MS
	ANA372J	-	Bundaleer Reservoir	AWQC	HPLC/MS-MS
Anabaena lapponica	966	+	Lake Takajärvi, Finland	Jarkko Rapala/Lisa Spoof et al. (2000)	HPLC/MS-MS and LCMS
Anabaena aphanizomenoides	ANA259C	-	Darling Anabranch, NSW	AWQC	HPLC/MS/MS and PI
	ANA262A	-	Darling Anabranch, NSW	AWQC	HPLC/MS/MS
Aphanizomenon issatschenkoi	CAWBGO2	-	New Zealand	Wood et al. (2007c)	Anatoxin-a producer
	16B	-	Torrumbarry Weir, Vic	AWQC	HPLC/MS/MS
Aphanizomenon ovalisporum	031E	+	Mildura, Vic	AWQC	

Table 3.1 Continued.

Species	Strain	CYN	Geographical origin	Reference	Notes
	031G	+	Mildura, Vic	AWQC	Fergusson (2003)
	033B	+	Thornden Pk,SA	AWQC	PI assay
	033C	+	Thornden Pk,SA	AWQC	Fergusson (2003)
	033E	+	Thornden Pk,SA	AWQC	
	035C	+	Thornden Pk,SA	AWQC	PI assay
	035D	+	Thornden Pk,SA	AWQC	PI assay
	035E	+	Thornden Pk,SA	AWQC	
	035F	+	Thornden Pk,SA	AWQC	
	ACO284	+	Rockhampton,QLD	L. Fabbro	LC/MS
Aphanizomenon sp	10E6	+	Melangsee, Germany	A. Stuken,	Preussel et al (2006)
Cylindrospermopsis raciborskii	CYP005		Murray/ DarlingJunction	AWQC	
	CYP014A	-	Coen Dam, QLD	AWQC	Fergusson (2003)
	CYP018A	-	Coen Dam, QLD	AWQC	HPLC/MS/MS
	CYPLI-19	-	Melangsee, Germany	J Fastner	Fergusson (2003)
	CYP020A	+	Solomon Dam, QLD	AWQC	Fergusson (2003)
	CYP020B	+	Solomon Dam, QLD	AWQC	Fergusson (2003)
	CYP023A	+	Coen Dam, QLD	AWQC	Fergusson (2003)

Table 3.1 Continued.

Species	Strain	CYN	Geographical origin	Reference	Notes
	CYP023E	+	Coen Dam, QLD	AWQC	Fergusson (2003)
	CYP024C	+	Fitzroy River, QLD	AWQC	Fergusson (2003)
	CYP025B	+	Fitzroy River, QLD	AWQC	Fergusson (2003)
	CYP026J	+	Bourke, NSW	AWQC	HPLC/MS-MS
	CYP028A	+	Cania Dam, QLD	AWQC	HPLC/MS/MS
	CYLI-29	-	Melangsee, Germany	J. Fastner	Fergusson (2003)
	CYP030A	+	Malparinka Farm,NSW	AWQC	HPLC/MS/MS
	CYLI-31	-	Melangsee, Germany	J. Fastner	Fergusson (2003)
	CYP032	+	Fitzroy River, QLD	AWQC	HPLC/MS/MS
	CYP034	+	Billabong, QLD	AWQC	HPLC/MS/MS
	CYP040	-	Brazil	S. Avevedo	Saxitoxin producer
	CYP041	-	Brazil	S. Avevedo	Saxitoxin producer
	CYP042	-	Brazil	S. Avevedo	Saxitoxin producer
	CYN43	+	Swan Reach,SA	AWQC	HPLC/MS/MS
	CYLI-53	-	Melangsee, Germany	J. Fastner	Fergusson (2003)
	CYLI-75	-	Melangsee, Germany	J. Fastner	Fergusson (2003)
Lyngbya wollei†		+	Yabba Creek, QLD	QHCaSS, Barbara Sendall/(Seifert et al. 2007)	(Seifert et al. 2007)

Sequences obtained from Genbank database are indicated in bold.
*Loss CYN toxicity through course of this study; †Culture maintained and PCRs performed by B Sendall; Blank not tested

3.2.2 P^{32} colony hybridisation

A replicate of the original cosmid library, constructed by Fergusson (2003), was achieved through the subculture of the 7 plates each consisting of 96 cosmid clones, in 200 µL of LB broth and 50 µg/mL of Amp. The plates were incubated at 37 °C overnight on a plate shaker with gentle shaking.

3.2.2.1 Preparation of membranes

Seven membranes were prepared by the transfer of 100 µL of each cosmid clone from each 96-well plate, onto a corresponding well within a slot blot apparatus (Bio-Rad Laboratories, Hercules, CA, USA). The clones were then captured by gentle vacuum onto Hybond™ N+ (Amersham Biosciences, Bucks, UK) membrane that was pre-wet beforehand with 200 µl of Sterile Milli-Q water. The cosmid CD7, containing the majority of genes belonging to the *C. raciborskii* AWT 205, cylindrospermopsin cluster, including the region where the *cynB* probe was designed from, was spotted onto each membrane at the top (centre) and on the right hand side of each membrane as a positive control for the *cynB* probe only. The cosmid DF7 which does not contain the region from which the *cynB* probe was designed from, was spotted onto the bottom (centre) and left hand side of each membrane as a negative control and each top left hand corner of the membranes was removed to permit orientation. The plates were replenished with 100 µL of LB and 30 % glycerol solution and stored at -80 °C.

The membranes were lysed in 10 % SDS for 3 min followed by denaturation for 5 min (0.5 M NaOH, 1.5 M NaCl) then neutralised for 5 min (1.5 M NaCl, 1.0 M Tris, pH 7.5). Membranes were washed in 2 × SSC, pH 7.0 for 5 min then allowed to air dry at RT for 30 min before alkali fixing with 0.4 M NaOH for 20 min. The preparation of the membranes was completed with a final rinse step of 5 × SSC, pH 7.0 for 1 min with gentle agitation, followed by drying at RT overnight before storage waiting hybridisation.

3.2.2.2 Preparation of P^{32} labelled probe

The PCR reactions were prepared as outlined in conventional PCR (Chapter 2 section 2.5.3). Thermal cycling conditions consisted of an initial denaturation temperature of 94 ºC for 10 min followed by 35 cycles of 94 ºC for 30 s , annealing of 56 ºC for 30 s and extension of 72 ºC for 2 min with a final extension of 72 ºC for 7 min, 4 °C hold. Results were verified by 1.0 % agarose gel electrophoresis (Chapter 2 section 2.5.5) at the completion of the cycling

protocol. Oligonucleotide primer sets used to amplify probes used in colony hybridisation experiments are listed in Table 3.2.

PCR products were purified by using MO BIO Ultraclean DNA purification spin columns as per kit protocols. Experiment A using a *cynB* purified probe (143 ng) was P^{32} dCTP labelled using the Amersham Megaprime™ DNA Labelling System (GE Healthcare Biosciences, Piscataway, NJ, USA) following the manufacturer protocols. The labelled probe was purified in a Qiagen purification mini column and eluted in the Elution buffer supplied with the kit. Experiment B with a *cynG, ask* probe (75 ng) was labelled by an in-house method developed by Zahra Shoari from the Plant Genomic Centre (Urrbrae, South Australia), and purified by a Sephadex G-100 column (Sigma-Aldrich) using the Geiger to determine the collection point (> 2000 clicks).

Table 3.2 Oligonucleotide primers used to generate probe for cosmid screening expts.

Primer	Sequence (5'-3')	Description	Product (bp)
B1	TTCCAAACGCCGCATCA	Reverse primer, cosmid probe, Experiment A	1.1 (kb)
B3	CCTCCGAGTTCAGACATACCAC	Forward primer, cosmid probe, Experi Experiment A	
B4	TGTGAGAACTGCTGAGGTATC	3', Reverse internal PCR primers to B1/B3 cosmid probe	259
B5	AATGTTACTGCGCCATCTAC	3', Forward internal PCR primers to B1/B3 cosmid probe	
cospro3	ATCCGGTAGTCGAGCTGTGG	Reverse primer, cosmid probe, Experiment B	749
cospro4	AGCTGTTGTAAGCTGGCCTC	Forward primer, cosmid probe, Experiment B	

3.2.2.3 P^{32} hybridisation and screening of membranes

The 7 membranes were divided between two hybridisation chambers and washed with 5 × SSC for 3 successive washes. The prepared membranes were initially incubated for 4 hr at 65 °C in a rotating oven with pre-warmed pre-hybridisation solution containing 15 mL of 10 × SSC, 3 mL of Denhardt's III, 1.5 mL of salmon sperm carrier DNA (5 mg/mL) and sterile Milli-Q water to a final volume of 30 mL. Upon completion of 4 hr pre-hybridisation the solution was decanted off and replaced with 10 mL of pre-warmed hybridisation buffer (3 mL 5 × HSB, 3 mL of Denhardts III, 3 mL of 25 % Dextran sulphate, 0.5 mL of salmon sperm

carrier DNA and sterile Milli-Q water). Meanwhile the P^{32} dCTP labelled probe was combined with 250 µL of salmon sperm denatured at 100 °C for 5 min and then immediately placed on ice for a further 5 min prior to the addition to the hybridisation buffer. The membranes were incubated overnight at 65 °C with rotation.

The hybridisation buffer was decanted off and the probe diluted out with 3 successive washes of 2 × SSC and 0.1 % SDS (w/v). Subsequent washes were then performed in a shaking water bath set to 65 °C. First a low stringency wash was performed, 2 × SSC and 0.1 % SDS for 5 min followed by a high stringency wash of 0.5 × SSC and 0.1 % SDS for 15 min. The washed membranes were sealed in plastic clingfilm and left for a week at -80 °C in a phosphorimager cassette containing an x-ray film.

The x-ray was removed and the image was scrutinised for evidence of hybridisation on each of the membranes, experiment A (Figure 3.2) and experiment B (Figure 3.5)

Membranes were stripped for future use, by incubation at 65 °C in 0.1 % SDS, 12 mM EDTA for 30 min, washed in 2 × SSC followed by a wash in Mill-Q water then air dried.

3.2.3 Genomic Walking Methods

3.2.3.1 Inverse PCR

Genomic DNA extracted from *C. raciborskii* AWT205 was digested with the Restriction Enzymes (RE), *Acc*I, *Eco*RV, *Hae*III, *Hin*dII, *Hpa*I, *Ssp*I as described in Chapter 2 (section 2.6.3). Selection of appropriate RE's was determined through the the RE cutter tool within the KODON™ software package. The digested DNA was re-ligated in accordance to protocols outlined in chapter 2 (section 2.6.4), followed by PCR amplification. Oligonucleotide primers sets used during inverse PCR are outlined in Table 3.3 alongside the corresponding RE digests.

All Inverse polymerase chain reactions were performed utilizing the Elongase *Taq* polymerase system (Invitrogen) and reactions were prepared as outlined for long template PCR in Chapter 2 (section 2.5.4). Thermal cycling protocols consisted of an initial denaturation step of 94 ºC for 30 s followed by 35 cycles of 94 ºC for 30 s , 59 ºC for 30 s and 68 ºC for 7 min all reactions were preformed on a Perkin-Elmer Gene Amp 2400 series.

The PCR endpoint was visualised by agarose gel electrophoresis as described in chapter 2 (section 2.5.5).

Table 3.3 Oligonucleotide primers used for inverse PCR

Primer	Sequence 5'→ 3'	RE digests
B6	GAGCCAGATACCTCAGCAG	***Hpa* II**/*Hin*d II/*Ssp* I
B7	TGTATGTAGATGCCAAGCTCC	***Hpa* II**
B8	GCTCTACAATGGCAGTTGC	*Hin*d II
B9	TCCAATAGTGTCTGCCAGTTG	*Ssp* I
B10	AAGGACAGAAGTACCAGCAC	*Eco*RV(1st digest)/Hae III/Acc I
B11	CATTTTGTGGTAGCGCCTG	*Eco*RV (1st digest)
B12	TCTGTAGATGGCGCAGTAAC	*Hae* III
B13	GCTGAGGTATCTGGCTCAG	*Acc* I
B18	AGCCAGATACCTCAGCAGTTCTCAC	***Eco*RV**(2nd digest) amplification primers
B19	GAATAGCCGGCTTAACTTCCTCAAGG	***Eco*RV(**2nd digest) amplification primers
B22	AGTAAATCTGGTGTTCTCGGTG	***Eco*RV** internal sequencing primer
B24	CTCTGGACTAATGGCCTGTTC	***Eco*RV** internal sequencing primer
B25	CTGATCCTGATAATCCAGTGAACC	***Eco*RV** internal sequencing primer
B26	TAACACCTGAGTGGCTTGAC	***Eco*RV** internal sequencing primer
B29	GAACAGGCCATTAGTCCAGAGG	*AccI, Hin*dIII,
B30	GCATACCAATTAACCCCACCGAATG	*Hin*dIII
B32	CGTTCTAAAGTGCCACCACC	*AccI*

Bold denotes Restriction enzymes IPCR that results in extending the CYN biosynthetic cluster.

3.2.3.1.1 Adapted mediated walking

The construction of adapters for ligation and thermocycling protocols was adapted from those described on the University of North Carolina Biology department website: www.bio.unc.edu/Faculty/copenhaver/Lab/Protocols.

Construction of adapter library

Genomic DNA extracted from *C. raciborskii* AWT205 was digested with the restriction enzymes, *Dra* I, *Eco*RV, *Hpa* I, *Hin*dII, *Rsa* I, *Sma* I, *Ssp* I as described in Chapter 2 (section 2.6.3).

Adapters required for ligation onto the ends of DNA fragments generated from restriction digests were prepared by combining 50µL each of 100µM oligonucleotides

ENDADAP, 5'-PO_4-ACCAGCCC-NH_4-3'
ADAPT, 5'GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGT -3'

The solution was boiled in a water bath for 2 mins followed by cooling in the deactivated bath for 1 hr. The prepared adaptor was stored at -20°C till required.

A 10 µL aliquot of purified, blunt ended, digested genomic DNA from *C. raciborskii* strain AWT205 was ligated with an excess of adaptor. The reaction mixture contained 5 µM of adapter (Geneworks), 6 U of T4 ligase (Invitrogen), 1 × T4 ligase buffer and Nuclease free water (Promega) to the final volume of 20 µL. The reaction was vortexed followed by incubation at 16°C for 16 hours then diluted 10-fold in 180 µL of sterile 1 × TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA) and stored at -20 °C till required.

Primary PCR

Primary PCR amplifications were conducted utilizing long template PCR as described in Chapter 2 (section 2.5.4), with the exception being the replacement of genomic DNA template with 2 µL of the dilute ligated DNA. Touchdown thermal cycling parameters involved an initial denaturation at 94 °C for 30 s, followed by 5 cycles of 94 °C for 30 s, annealing at 60 °C for 30 s and extension of 68 °C for 7 min the annealing was decreased 1 °C each cycle, a further 30 cycles of 94 °C for 30 s, 55 °C for 30 s and 68 °C for 7 min, 4 °C hold.

Secondary PCR

Secondary PCR conditions were as described for primary PCR with the exception that the template was the purified product from the primary PCR and the primers used were hemi nested in respect to the primer pairs utilised in the initial PCR (see Table 3.4).

Table 3.4 Primers used in adapted mediated walking PCR

Primer Name	Sequence	Reaction	Reference
AP1	GTAATACGACTCACTATAGGGC	Primary PCR, adaptor specific	Siebert et al 1995
GSP3	TTAGTATGGAACAGGCCATTAGTCCAGAGG	Primary PCR, gene specific	This work
AP2	ACTATAGGGCACGCGTGGT	Secondary PCR, adaptor specific	Siebert et al1995
GSP4	TGGGTGTTAGTTCATGTCGGTTTTGC	Secondary PCR, gene specific	This work

3.2.4 Cylindrospermopsin biosynthetic gene-directed PCR

All PCR reactions were prepared as outlined for Conventional PCR in Chapter 2 (section 2.5.3). Oligonucleotide primer sets used to target genes implicated in the biosynthesis of cylindrospermopsin are outlined in Table 3.5.

Thermocycling conditions consisted of an initial denaturation temperature of 94 ºC for 10 min followed by 35 cycles of 94 ºC for 30 s , annealing of 55 ºC for 30 s and extension of 72 ºC for 5 min with a final extension of 72 ºC for 7 min, 4 °C hold. Results were verified by agarose gel electrophoresis (Chapter 2 section 2.5.5) at the completion of the cycling protocols.

Table 3.5 Oligonucleotide primers targeting genes implicated in the synthesis of cylindrospermopsin

Primer	Sequence (5'-3')	Gene target	Product (bp)
AB1	AACCATCCTATTGGCCAGAG	*orf4*(putative ABC transporter/ AAA+atpase)	466
AB2	ACTAACACCTGAGTGGCTTG		
ASK1	AGTAAATCTGGTGTTCTCGGTG	*cynG*, Adenylysulfate kinase (*ask*)	333
ASK2	ATGGTGACAATATCCGTCATGG		
CB2	AGGAAATTGACCAGACGG	*cynB*, *pks* (AT domain), *aoaC* homologue	609
CB1	CGAACAAGACAGGATTGC		
CA2	CACCTCAAGCACTACAAGAAC	*cynA*, *pks* (AT domain)	689
CA1	CTCTCGGCTGACAGATTG		
CD4	CTGGCTTACAGGCACAAC	*cynD*, *ps/pks* (AT domain), *aoa*B homologue	689
CD3	GTCTAACATTGCTTCCACTAGC		
CC2	TCGAAGCAACTATGTCACG	*cynC* amidotransferase, *aoaA* homologue	752
CC1	GAAGTAATCCTCACAGGTTCC		

Table 3.5 cont.

Primer	Sequence (5'-3')	Gene target	Product (bp)
AMT1	GAATCGCAATGTGGTCTGTG	*cynE*, aminotransferase/ sulfotransferase	689
AMT2	AACTTGCTACGGTGCTGTAC		
ATZ2	AGCTCCTATGTTCGATGCAC	*cynF*	616
ATZ1	AGCCATACGCAGGACTTC		
TZ1	TGAACTTAGACCGCATCCTG	*orf3*	585
TZ2	TCATGGAGTAGCCGTTTGTC		
N1	AGGTGCTAACTGTTCTGTGG	*cynL* (Na^+ Efflux)	724
N2	GTACCACTGCGTTAACAGCA		
PHF	TCAAGAAGACGCACCCAC	*cynH*, amidohydrolase	323
PHR	GCCATTCATCCCAAAGCAC		
CF2	CATTTACTCGCAGTGCCATG	*cynJ* (*cyrF*), *pks* (AT domain)	1037
CF1	AGATAGAGACCTCTCCCGATG		
CDY1	GTGACGACATTCTGCATCC	*cynK (cyrD, pks)* (AT domain)	1007
CDY2	CTGATGGTAGCCGGATCA		
AL3	TGCAGATGGCTCGAGAACT	*cynB, pks, A.lapponica* only,	2.3 (kb)
AL4	ACGATAGCCGGCACATTGC		
CYLATF	ATTGTAAATAGCTGGAATGAGTGG	*cynC, A. lapponica* only Kellman et al 2006	1.1 (kb)
CYLATR	TTAGGGAAGTAATCTTCACAG		

3.2.5 Long template PCR

Long template PCR utilising the Elongase *taq* polymerase system was prepared as previously described in Chapter 2 (section 2.5.4) and was conducted with assistance from Paul Rasmussen. The reactions were performed using oligonucleotides designed described in Rasmussen et al (2008).

ATint CAAAATTTAGGGATCGAGTCCATTCTCTC

***CYNA*BCint** AACCTTTAACTCTATGCCCATCACTACCTT

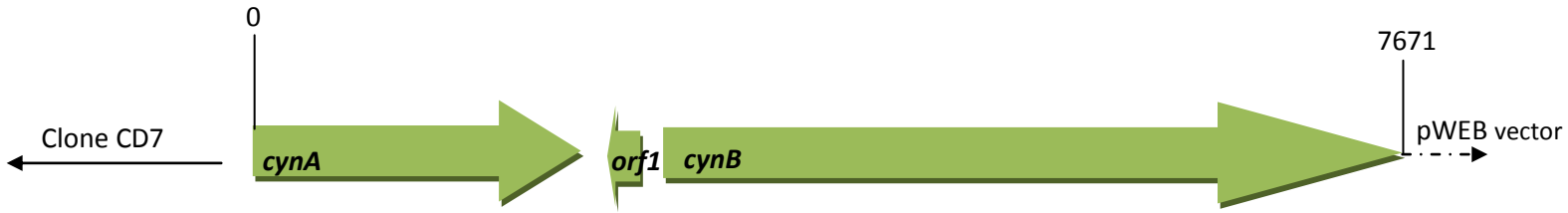
Thermal cycling protocols consisted of an initial denaturation of 94 °C for 2 min followed by 35 cycles of 94 °C for 30 s, 60 °C for 30 s and finally 68 °C for 20 min, Products were verified by 0.8 % agarose gel electrophoresis as described in Chapter 2 (section 2.5.5). The product of interest was excised from the agarose gel and cloned as described in Chapter 2 (section 2.6.2) to facilitate sequencing.

3.3 Results and Disucssion

3.3.1 Genome walking

3.3.1.1.1 *C. raciborskii* AWT 205 cosmid library

The sequence of the CYN biosynthetic cluster was primarily achieved through directly sequencing the CD7 cosmid. The CD7 cosmid was obtained through the construction of a cosmid library incorporating the genome of *C. raciborskii* AWT 205 utilising the pWEB™ Cosmid Cloning Kit (Epicentre Technologies, WI, USA) as outlined in Fergusson (2003). Positive clones containing inserts with regions homologous to *cynB*, (*pks*) and also *cynD* (*ps*) genes were identified through P^{32} hybridisation experiments also conducted by Fergusson (2003). An alternate cosmid identified was that of DF7; however this cosmid was shown to have a truncated open reading frame (*orf*) for the *cynB pks* gene. The sequence obtained and annotated by Fergusson (2003) is represented by Figure 3.1. The putative genes that were discovered throughout the course of the current work described below, were named following the convention adopted by Fergusson (2003) and were named in the order of discovery, not in relationship to those described by Shalev-Alon et al (2002). Some of the sequencing work was conducted by Dr Sarah Baker.



Scale 500 bp

Figure 3.1 Organisation and physical map of the C. raciborskii AWT205, 7671 bp nucleotide sequence contained on the cosmid clone CD7, identified and sequenced by Fergusson (2003).

Bold arrows are indicative of the open reading frames annotated by Ferguson (2003), *cynA* and *cynB* encode polyketide synthases and *orf1* demonstrates homology to IS4 transposase sequence (76% similarity to *Synechocystis* sp. PCC 9413).

3.3.1.2 P^{32} colony hybridisation cosmid library – Experiment A

It became evident that the insert within cosmid CD7, contained an incomplete *cynB orf*, upstream of the original *cynB* probe utilised by Fergusson (2003). Sequence pertaining to the pWEB vector was identified with routine vector screening of contiguous nucleotide sequences amplified from the CD7 cosmid DNA template.

To identify an alternate cosmid candidate that would facilitate the contiguous sequencing of the cylindrospermopsin biosynthetic gene cluster, upstream of *cynB*, P^{32} colony hybridisation was undertaken. A 1.1 kb cosmid probe (Table 3.2) was amplified by conventional PCR, in accordance to methods outlined in Chapter 2 (section 2.5.3) by primers targeting the 3' end of the *C. raciborskii* AWT205 *cynB* gene. Replication of the original cosmid library and preparation of membranes are detailed in sections 3.2.2. and 3.2.2.1. Of the 672 original clones, only 628 were successfully resuscitated. The conditions undertaken for P^{32} hybridisation of the membranes with the *cynB* probe are described in section 3.2.2.3. The X-ray images of the membranes relevant to this experiment are presented in Figure 3.2.

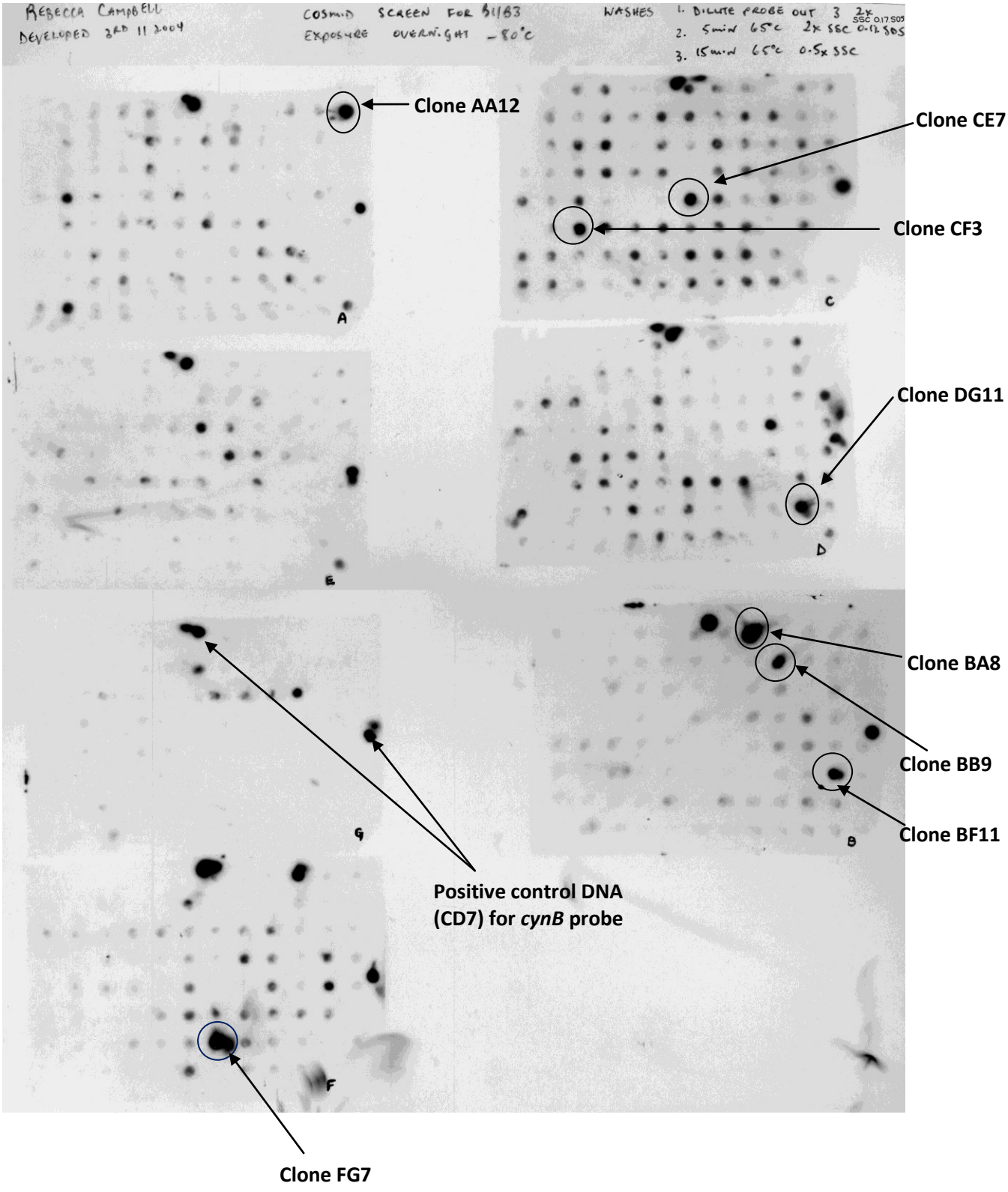


Figure 3.2 Experiment A – P ^{32} hybridisation of a 1.1 kb *cynB* probe to a cosmid library constructed from *C. raciborskii* AWT205 chromosomal DNA.

Positive control DNA (cosmid DNA from CD7) was applied top (centre) and right hand side of each membrane. Cosmid clones that represented possible candidates and were selected for further screening are indicated.

The positive controls consistently demonstrated effective hybridisation with the *cynB* probe, across all 7 membranes, which verified that the probe was distributed evenly around the whole membrane. Initially four cosmids (AA12, BA8, DG11 and FG7), showing strong hybridisation to the probe, were selected for further analysis, to determine whether one may contain an insert that pertained to the *cynB* region. Fresh cultures of the relevant clones were grown and the cosmid DNA extracted (Chapter 2 section 2.4.3). PCR screening was performed using the B1 and B3 primers, which were used to amplify the hybridisation 1.1 kb probe. However, analysis of the amplification reaction indicated that it was unlikely that any of the clones selected contained the insert of interest. There was either no amplification product or, the obtained product was considerably larger (3.0 kb) than positive control of 1.1 kb (data not shown). A further four clones (BB9, CE7, CF3, and BF11), with moderate hybridisation intensities, were additionally selected for screening. Similarly, the amplified PCR products were larger than anticipated or absent (data not shown). A final PCR attempt was undertaken using new primers, B4 and B5 (Table 3.2), designed internal to, and focusing on the furthermost 3' end of the region bound by the hybridisation probe primers (B1/ B3). No PCR amplification of products was observed from any of the template DNA isolated from the selected cosmids (data not shown). This outcome is perhaps suggestive that the positive hybridisation observed, was a result of homologous regions within the genome. The probe was designed from a region encoding a polyketide synthase, these genes are ubiquitous throughout cyanobacterial genomes, and therefore it is possible that similar sequences exist within their coding regions that led to false positives. These results could also be suggestive that a "hole" may potentially exist within the cosmid library. Fergusson's (2003) calculations indicated that the genome of *C. raciborskii* AWT205 was represented by four genome equivalents within the cosmid library. Whilst this should be adequate in ensuring that any given sequence would be found within the library, factors like preferential cloning may generate sequence gaps. Research groups undertaking genome assemblage projects, facilitated by clone libraries, have often experienced these discontinuities in sequence (Kohara et al. 1987; Deckers and Voordouw 1994), even when they have obtained 6-12 fold coverage of the genome (Goff et al. 2002; Eichler et al. 2004). Additional sequencing strategies are often employed to complement clone libraries and acquire the missing sequence data (Burton et al. 2004; Garber et al. 2009).

3.3.1.2.1 Inverse PCR

In light of the failure to identify a suitable clone (section 3.3.1.2), it was necessary to pursue the remaining sequence of *cynB* through an alternative genome walking strategy, inverse PCR. Inverse PCR (IPCR) enables the characterisation of unknown sequence by a method first described by (Ochman et al. 1988) and is diagrammatically represented in Figure 3.3. In short, genomic DNA is digested by restriction enzymes, which have been determined from regions of known sequence, and anticipated to generate 1-3 kb fragments. The cleaved DNA is then re-ligated to produce circular templates that are amplified by PCR. Oligonucleotide primers are designed from homologous regions of the known sequence and are transcribed in opposite directions, resulting in a linear product that contains both 5' and 3' sequence from the known genic region. These remnant sequences facilitate anchoring the new sequence within the known region, confirming its validity of as an adjacent region. These steps are able to be carried out repeatedly, with different enzymes, permitting genomic walking into flanking regions of unknown sequence.

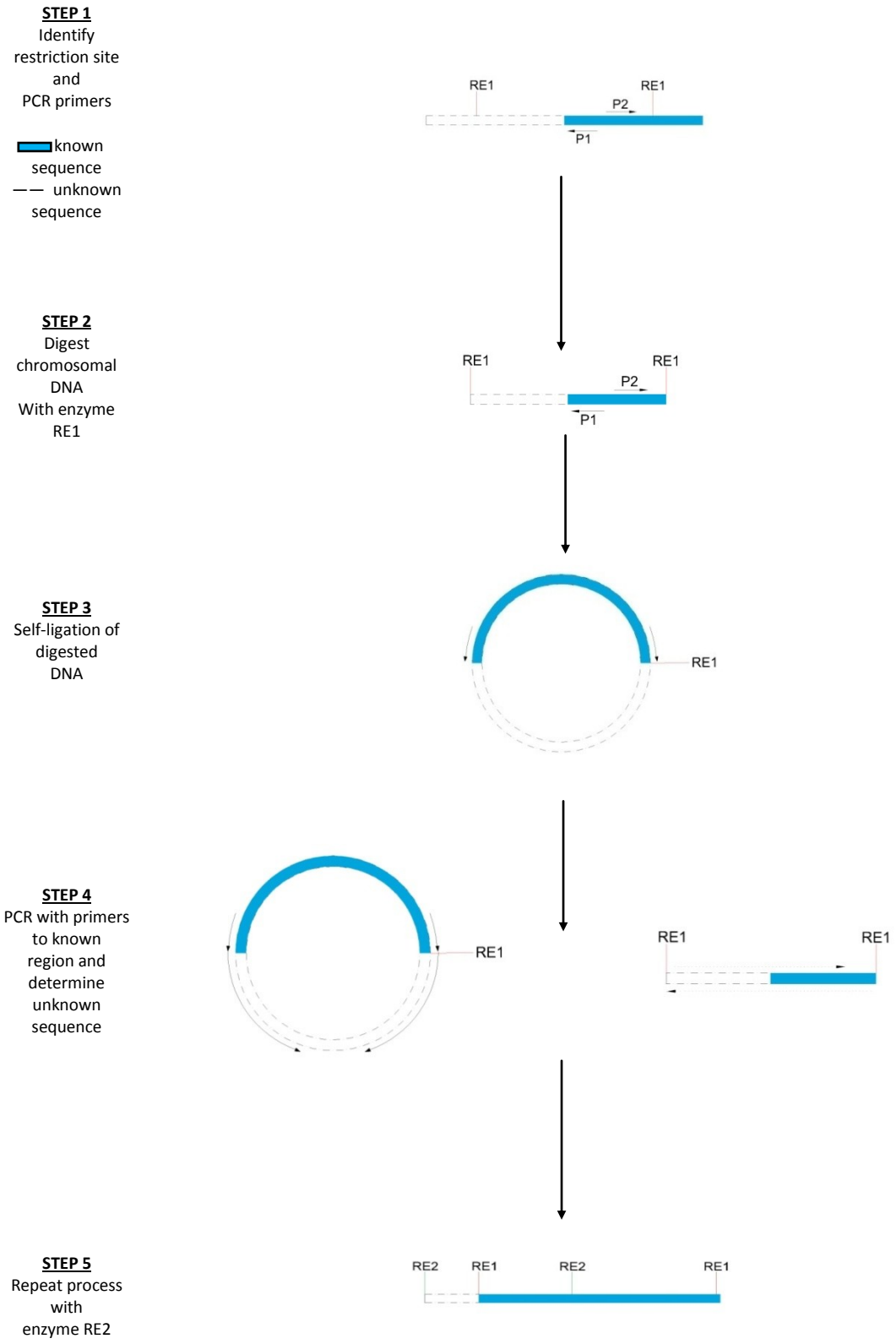


Figure 3.3 Diagrammatic representation of the process of inverse PCR.

Adapted from K.Fergusson (2003).

Overall, IPCR was able to extend the sequence by approximately 3.2 kb. The additional sequence finalised the *cynB* ORF and identified three additional ORF's, *orf2,* a putative IS4 transposase, an adenylysulphate kinase (*ask*) and *orf4,* a putative ABC transporter/AAA^+ ATPase that are all potentially implicated in the CYN biosynthetic cluster (Figure 3.4). However, there is potential with IPCR to generate fragments, which are not contiguous with the region of interest. This was evident for one 600 bp fragment, amplified from an inverse primer set designed and amplified from an *Eco*RV digest of the *C. raciborskii* genomic DNA. The errant fragment revealed a sequence that was not able to be aligned into the genome assemblage and encoded an ORF that demonstrated high homology to a permease, which didn't extend and finalise the sequence of the *cynB* ORF. For this reason improved methods for genomic walking were explored, as more efficient alternatives.

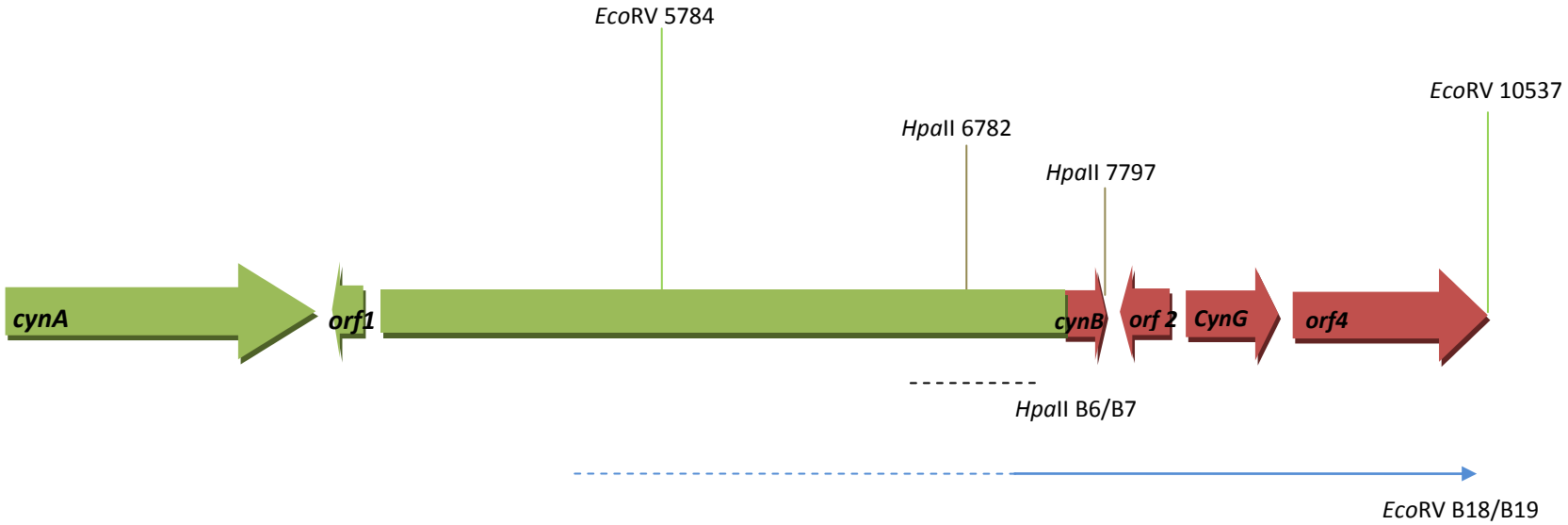


Figure 3.4 Open reading frame organisation of the extended *C. raciborskii* AWT205 CYN biosynthetic cluster nucleotide sequence (2916 bp), obtained by inverse PCR

Sequence obtained from the cosmid CD7 is shown in green (Fergusson 2003), sequence extended by inverse PCR is shown in red (This work). Enzyme sites and corresponding amplification primers are indicated. Solid lines represent new sequence, dashed lines known sequence prior to IPCR. ORF *cynG* encodes adenylsulphate kinase (*ask*); *orf2* encodes IS4 insertion element, transposase; *orf4* encodes a putative ABC transporter/AAA+ ATPase

.

3.3.1.2.2 P^{32} colony hybridisation cosmid library – Experiment B

Due to inverse PCR identifying genes being perceived as potentially more "unique" within a genome as polyketide synthases, reprobing of the cosmid library was reattempted to expedite the sequencing process. This subsequent attempt utilised a 749 bp cosmid probe designed within a region of the *ask* gene. The method used is detailed in section 3.2.2 and conditions were similar to experiment A, with the exception of the P^{32} probe preparation (outlined in section 3.2.2.2). The membranes pertaining to this attempt are presented in Figure 3.5. Strong hybridisation with the *ask* probe is evident and several candidates were isolated for further screening and are indicated on Figure 3.5.

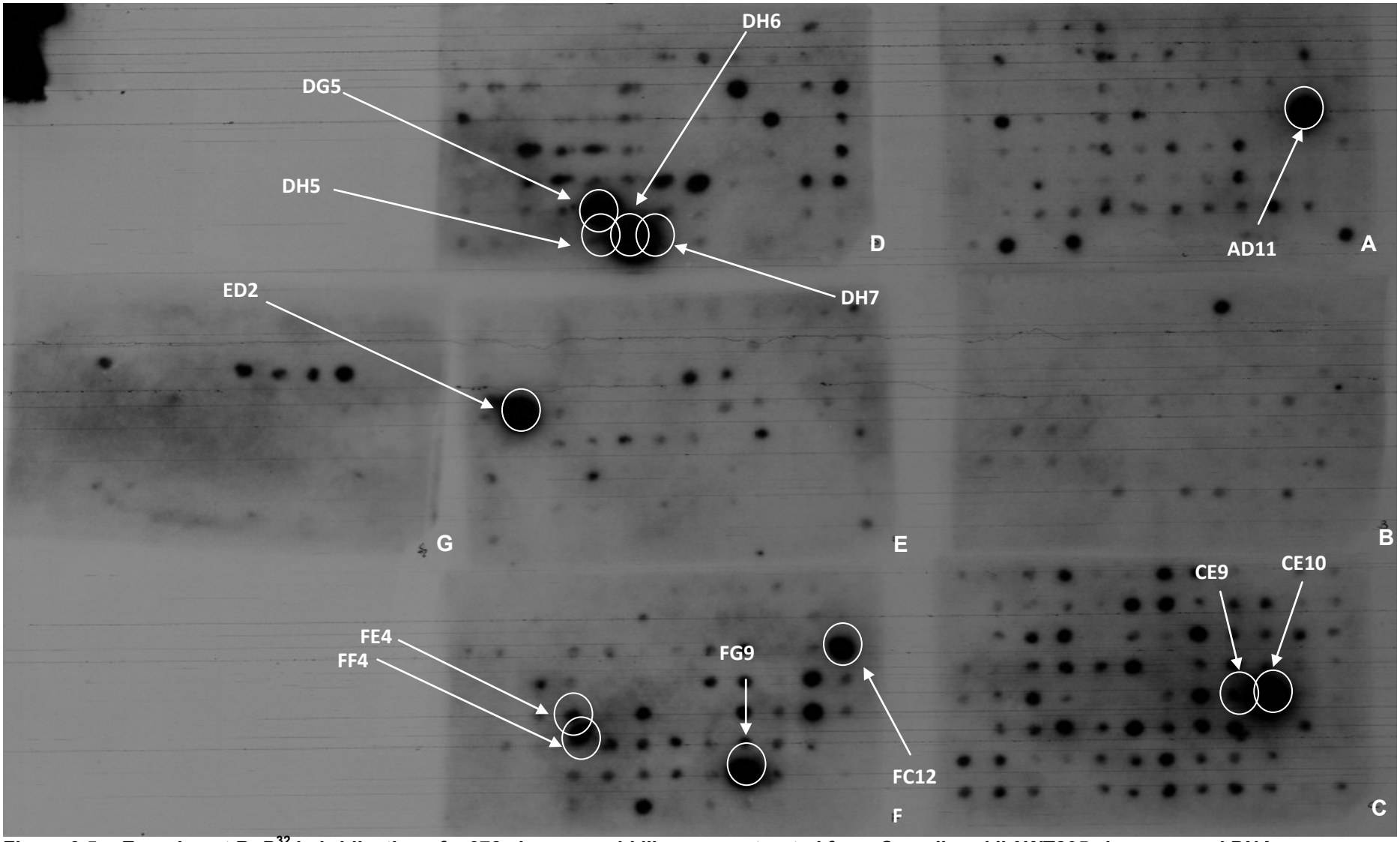


Figure 3.5 Experiment B: P^{32} hybridisation of a 672 clone cosmid library constructed from *C. raciborskii* AWT205 chromosomal DNA.

P^{32} hybridisation was performed with a 749 bp probe targeting the *ask* (*cynG*) ORF. Cosmids selected for further screening are circled. Of the 22 cosmids selected for addditonal investigation only 12 of the most promising candidates are indicated.

Initially, 11 potential candidates were screened by PCR, to ascertain whether they, in fact, contained the *ask* gene. The PCR amplification reactions were performed with the primers, cospro3 and cospro4 (749 bp *ask* hybridisation probe). The PCR conditions were as described in Chapter 2 (section 2.5.3) for conventional PCR protocols and the endpoint verified by gel electrophoresis (Chapter 2 section 2.5.5). Analysis of the PCR products revealed that the majority were either significantly shorter (≤ 300bp) or exhibited considerable amplification of multiple, non-specific products (data not shown). A further 10 cosmids, selected from surrounding regions demonstrating intense hybridisation, were additionally screened by PCR. Again either no amplification or non-specific binding was observed, with one exception, cosmid DH5 which amplified a PCR product that was approximately 150 bp shorter in length than the positive control of 749 bp (data not shown). The amplification efficiency for the DH5 product was poor, so the products were purified (Chapter 2 section 2 6.1) and cloned (Chapter 2 section 2 .9.2) followed by re-amplification of the insert using the cospro3 and cospro4 primer set (Figure 3.6). Subsequent sequencing of the CP1 plasmid insert, revealed a product that was neither homologous to *ask* or had a cyanobacterial precedent in the database (69% to 68%, similarity to *Rhodospirillum centenum and Sphingopyxis alaskensis,* thiamine-phosphate pyrophosphorylase). Interestingly, the 20 bp primer sequence of cospro4 was identifiable in the 5' end of the obtained sequence. This outcome suggests that there was a bacterial contaminant either in the PCR reaction itself or in the DNA extract isolated from DH5. It is also suggestive that within this contaminant, there was a coincidental homologous sequence that demonstrated 100% sequence identity with primer cospro4, which facilitated the amplification of the 600 bp product.

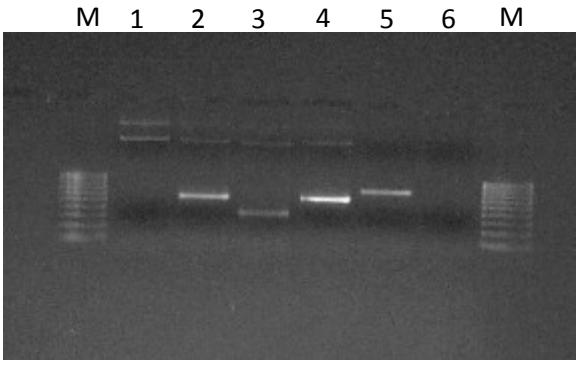


Figure 3.6 Amplification products from a pCR2.1 plasmid insert, obtained from cosmid clone DH5, using the cospro3 and cospro4 primer set, targeting the *ask* gene.

Lane 1 - Plasmid CP1; Lane 2 - Plasmid CP2; Lane 3 - Plasmid CP3; Lane 4 - Plasmid CP4; Lane 5 - *C.raciborskii* AWT205 positive control; Lane 6 - negative control; Lane M - molecular weight marker 100 bp ladder (bp): 1000, 900, 800, 700, 600, 500, 400, 300, 200, 100.

In parallel to the cloning and sequencing of the DH5 PCR product, restriction enzyme digestion was performed on the DH5 cosmid clone DNA. This was to investigate whether this shorter amplification product was potentially a truncated *ask* region, as a result of either recombination or incomplete insertion. The available sequence of the biosynthetic cluster was submitted to the KODON™ restriction enzyme analytical tool, which enabled selection of suitable enzymes that bound the *ask* loci. Restriction enzymes, *Eco*RI and *Dra*I, were selected with predicted fragments of 1.3 kb and 1.9 kb, respectively. Both bands were predicted to contain the recently identified genes, *ask* and the region downstream, *orf4*. The DH5 clone was digested, in separate reactions, with *Eco*RI, *Dra*I and *Not*I, a restriction site that linearises the pWEB vector, and visualised by agarose gel electrophoresis as outlined previously (Chapter 2 section 2.5.5). The restriction enzyme profiles of clone DH5, revealed a single band with *Not*I and the predicted bands of approximately, 1.3 kb and 1.9 kb within the *Eco*RI and *Dra*I digests (Figure 3.7). The bands of interest, from the *Eco*RI and *Dra*I digests, were excised and purified from the gel followed by cloning as described in Chapter 2 (sections 2.6.1 to 2.6.2).

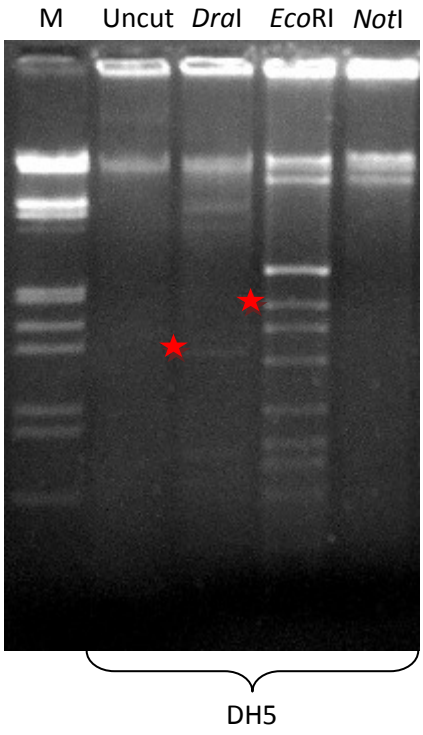


Figure 3.7 Restriction enzyme digestion of cosmid clone DH5 with *Dra*I, *Eco*R1 and *Not*I.

The red star indicates the 1.3kb, 1.9kb bands as predicted by KODON restriction enzyme analysis tool, and linearization of the cosmid. Lane M molecular weight marker λ *Hin*dIII/*Eco*RI (bp): 21226, 5148, 4978, 4268, 3530, 2027, 1904, 1584, 1375, 947, 831, 564, 125.

Gene-directed PCR amplification was performed using primers that were targeted toward *orf4* (AB1 and AB2, 466 bp product see Table 3.5). Confirmation of the presence of this ORF was highly desirable, as it resembled the furthermost region of the known 3' biosynthetic sequence. PCR conditions were as described for conventional PCR, Chapter 2 (section 2.5.3) and products verified by gel electrophoresis (section 2.5.5). Analysis of the PCR using the *orf4* directed primer set, revealed products of the same size as the positive control, for both the *Eco*RI and *Dra*I band (Figure 3.8). Sequence analysis of the product isolated from the PCR performed on the *Eco*RI band revealed homology to *orf4* and aligned into the contig.

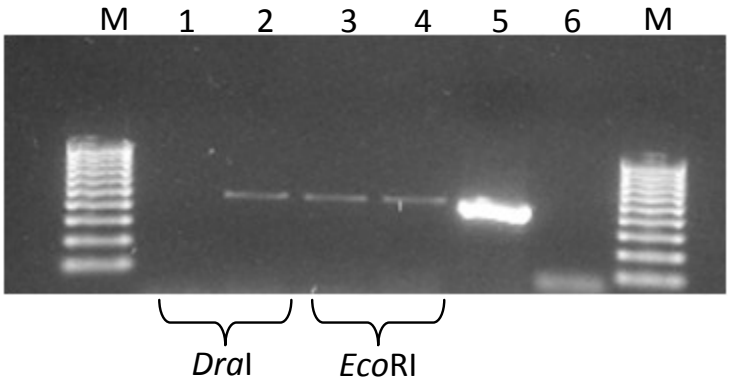


Figure 3.8 Amplification of the *orf4* (putative ABC transporter/AAA+ ATPase) gene from bands excised from *Dra*I (1.3 kb) and *Eco*RI (1.9 kb) restriction enzyme digest of cosmid clone DH5

Lane 1 *Dra*I, 1.3 kb band, replicate 1

Lane 2 *Dra*I, 1.3 kb band, replicate 2

Lane 3 *Eco*RI, 1.9 kb band, replicate 1

Lane 4 *Eco*RI, 1.9 kb band, replicate 2

Lane 5 *C.raciborskii* AWT205, positive control

Lane 6 negative control

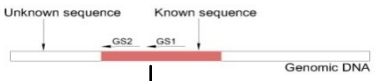
Lane M molecular weight marker 100 bp ladder (bp): 1000, 900, 800, 700, 600, 500, 400, 300, 200, 100

Therefore it appears probable that cosmid DH5 was a contiguous clone. However, whilst difficulties were being experienced toward identification of the clone's potential, an alternative genomic method was being run in parallel. This alternate method, adapted mediated walking, had amplified a promising fragment and therefore it was decided to pursue sequencing of the biosynthetic cluster via this approach.

3.3.1.3 Adapted mediated walking

A ligation mediated suppression PCR described by (Siebert et al. 1995) was utilised to extend the cylindrospermopsin biosynthetic cluster sequence from DNA extracted from the *C. raciborskii* isolate AWT205. This approach was undertaken when P^{32} hybridisation of the cosmid library (section 3.3.1.2.2), had not readily identified a suitable candidate, to complete the sequencing of the cylindrospermopsin cluster. The ligation mediated suppression approach is designed to improve confidence and efficacy with walking in uncloned genomic DNA, more so than inverse PCR. The method incorporates the ligation of an adaptor onto DNA fragments generated by restriction digestion and use of adaptor specific primers that facilitate the suppression of non-specific DNA synthesis during PCR amplification. The process is represented diagrammatically in Figure 3.9

STEP 1
Design hemi nested, gene specific primers (GS1, GS2)



STEP 2
Restriction digest with appropriate six-base blunt ended enzymes

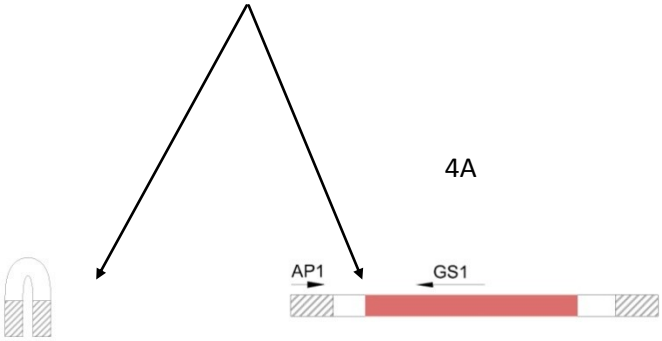


STEP 3
Ligation of adaptor to digested DNA fragments



STEP 4A
Primary PCR using gene specific primer 1 and adaptor primer 1

STEP 4B
Repeat PCR protocol Using step 4A product as template. Secondary PCR using gene specific primer 2 and adaptor Primer 2



No primer binding: Inverted terminal repeats sequences in adaptor lead to panhandle formation and PCR suppression.

Primer binding PCR proceeds with DNA amplification

Figure 3.9 Diagrammatic representation of the processes involved in adapted mediated genomic walking strategy.

Two of the seven digests attempted (*Ssp*I and *Hin*dII), produced fragments that were able to extend the sequence by a further 2.2 kb. The sequence revealed three new ORFs, which were all transcribed in the same direction and encoded hydrogenase expression genes (*hypC, hypD, hypE*) (Figure 3.10). The role of theses ORFs appeared unlikely to be explicitly involved in the biosynthetic pathway of cylindrospermopsin, therefore on this premise, sequencing of the 5' region of the biosynthetic cluster was concluded.

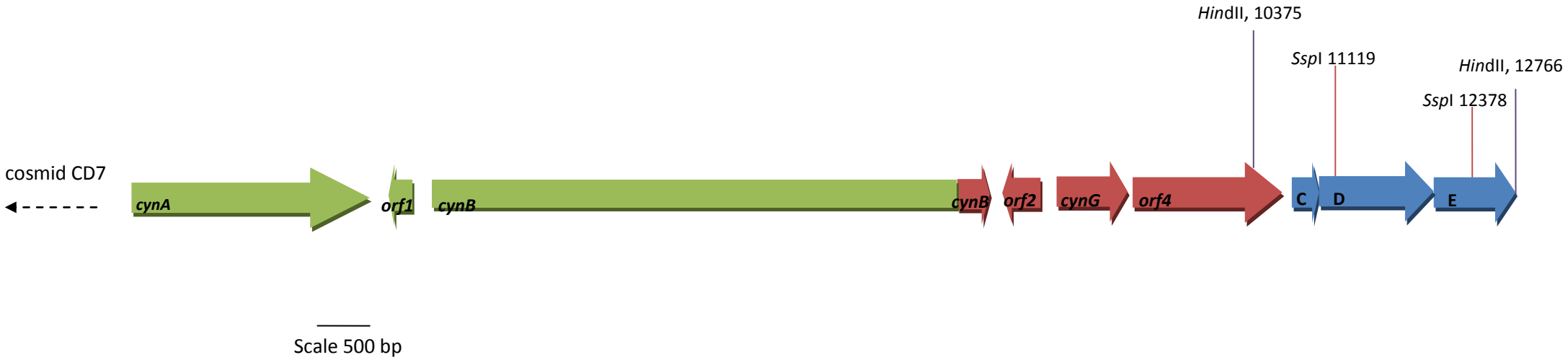


Figure 3.10 Open reading frame organisation of the *C. raciborskii* AWT205 CYN biosynthetic gene cluster, including 2217 bp nucleotide sequence obtained by adapted mediated walking.

Sequence extended by adapted mediated walking is shown in blue (this work), sequence extended by inverse PCR is shown in red (this work), and sequence obtained from the cosmid CD7 is shown in green (Fergusson 2003). The approximate restriction enzyme sites are indicated. The cluster includes hydrogen expression genes, *HypE* (E), *HypD* (D), *HypC* (C), a putative ABC transporter/ AAA^+ ATPase *(orf4)*, adenylysulphate kinase (*cynG*), Transposase (IS4 insertion element *orf2*), Polyketide synthase (*cynB*) Transposase (IS4 insertion element *orf1*) Polyketide synthase (*cynA*)

3.3.1.3.1 The completion of the CYN biosynthetic cluster

A further 24.8 kb pertaining to the biosynthetic cluster was sequenced from the CD7 cosmid clone, extending from the 7671 bp sequence previously identified by Fergusson (2003). A final nucleotide sequence of 42 kb was obtained through the molecular techniques presented in this chapter. Overall, an additional 9 ORFs were detected and identified on the cosmid CD7, with predicted functionalities ranging from polyketide synthases (PKS), a chimeric peptide synthetase (PS)/PKS, an amidotransferase, aminotransferase, aminohydrolases, transpoase (IS4 insertion-like element) and sodium (Na) efflux pump. These will be discussed in greater detail in Chapter 4. A completed ORF for the *pks* gene, *cynJ* was not possible using cosmid CD7, however whilst the sequencing work presented in this chapter was being finalised, another research laboratory published the CYN biosynthetic cluster, also isolated from *C. raciborskii* AWT205 (Mihali et al. 2008). Mihali's et al. (2008) work completed the biosynthetic cluster, and it is their sequence that is used in this work, to finalise ORFs for the *pks* genes *cynJ* (*cyrF*) and *cynK* (*cyrD*). The structural organisation of the CYN biosynthetic cluster is represented by Figure 3.11.

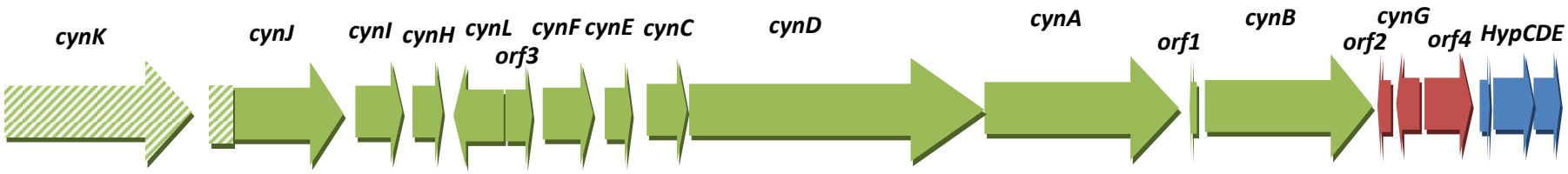


Figure 3.11 Gene organisation of a 44.2 kb region sequenced from *C. raciborskii* AWT205

Sequence obtained from the cosmid CD7 is shown in green, sequence obtained from *C.raciborskii* AWT205 CYN gene cluster sequence as found by Mihali et al 2008 is shown in hatched green, sequence extended by inverse PCR is shown in red, sequence extended by adapted mediated walking shown in blue.

Whilst this was not an ideal scenario, this did not distract from the principal aim of this chapter, which was to permit full consideration of the molecular architecture involved in CYN biogenesis, with the priority toward identifying the most robust molecular markers, to implicitly determine CYN toxicity. Further to this, Mihali's work complements the findings within this chapter and demonstrates the stability of the sequence within the cyanobacterial genome. The original isolate pertaining to *C. raciborskii* AWT205, was first isolated from an ornamental pond some 13 years ago (Hawkins et al. 1997). It has been a part of the AWQC collection for 12 years, where it undergoes routine passage on a tri-monthly basis. Of the comparable sequences, only 2 single nucleotide polymorphisms (SNPs) were detected; one at 3356 bp (with respect to sequence presented in this thesis) the other at 30680 bp (with respect to sequence presented in this thesis), both resulting in non-synonymous changes, aspartic acid to histidine and alanine to glycine respectively. The rate of sequence divergence can be calculated by the following formula:

$$K = -\frac{3}{4}\ln\left(1 - \frac{4}{3}\,d\right)$$

Where sequence difference (d) = 2 mismatches/ 35312 total comparable sites. An evolutionary rate (in this instance for non synonymous substitution) is then calculated using:

$$E = K/2T$$

Where T is the amount of time the sequences have had to diverge (13 years). The found evolutionary rate of change between the two *C. raciborskii* strains was found to be 10^{-6} non synonymous base changes per year. Variations in evolutionary rates are reflective of the diversity of a sequence, with evolutionary rates generally being found in the order of 10^{-10} to 10^{-7} bp substitutions per year, exceptions are observed in highly evolving species, for example viral sequences that are up to three orders of magnitude faster (Avise 1994)

The G+C content of the nucleotide sequence of the genes involved in the biosynthesis including those adjacent to the CYN cluster was 45.9%. This is comparable to total G +C content of other filamentous cyanobacteria, *Anabaena* (*Nostoc*) sp. (PCC7120, 41.3% (Kaneko et al. 2001) and 35.9 to 56.4 % for 50 *Anabaena* strains biochemically determined

by (Li and Watanabe 2002). The sequence obtained during the course of this work is presented in Appendix 1.

3.3.1.3.2 Implicating genes in CYN biosynthesis

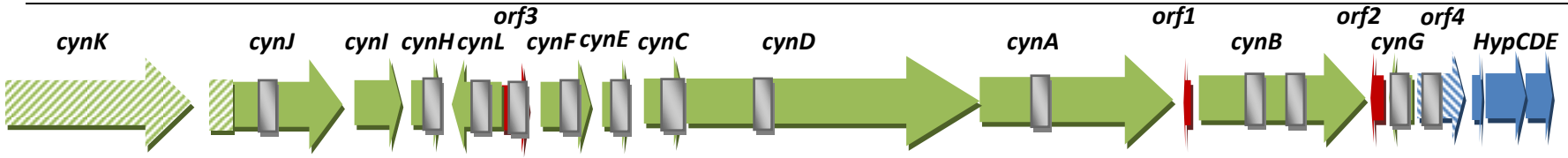
Overall, a total of 19 ORFs were identified, with 13 ORFs surveyed by gene-directed PCR, to assist in the determination of their putative involvement in the CYN biosynthetic cluster. This was performed firstly, as an alternative to transposon mutagenesis, as filamentous cyanobacteria are notoriously difficult to transform (Cai and Wolk 1990; Cohen et al. 1998) and *C. raciborskii* has been recalcitrant to transformation in previous attempts (Fergusson 2003). The perceived difficulties for filamentous cyanobacteria toward transformation, primarily arise as a result of multiple genome equivalents that have been observed in some species, complicating the segregation of mutants from wild type genomes (Cai and Wolk 1990).

As highlighted in chapter 1, CYN is a potent agent that illicits a cytotoxic response in a range of mammalian organs. There is also preliminary evidence suggesting long-term exposure can lead to carcinogenic consequences. For these reasons, the presence of CYN in potable and recreational waters poses significant public health risks; therefore the recognised producer organisms are routinely monitored. Whilst the current molecular assays available for CYN detection (Fergusson and Saint 2003; Rasmussen et al. 2008b) have on a whole been proven to be robust, there has been recent evidence that the established molecular probes may fail to amplify and therefore detect a novel CYN toxigenic species (Preussel et al. 2006). Ultimately, this is not desirable when promoting molecular methods as a beneficial screening/detection tool. Therefore, the survey of the biosynthetic genes by gene-directed PCR, was also performed with the purpose to identify the molecular marker that is inherent for CYN toxicity across a range of cyanobacterial species and strains, using the sequence obtained from *C. raciborskii* AWT205 as a base.

CYN gene-directed PCR

Gene-directed PCR was performed as described in section 3.2.4 with oligonucleotide primers sets listed in Table 3.5. All reactions were performed as a minimum of 3 independent trials. A summary of the results is represented in Figure 3.12 and more details are documented in Appendix 2. The domain targeted within each of the PKS modules was the acyltransferase.

Acytransferases (AT) domains upload the successive acetate unit for incorporation into the evolving polyketide product. This region was selected through the KODON™ repeat and match analysis application, which enables recognition of homologous regions within a sequence. As discussed in chapter 1, the production of many of the secondary metabolites by cyanobacteria, involve PKS and NRPS genes, therefore, to minimalise the possible amplification of unrelated PKS modules an effort to identify possible regions within modules, which may demonstrate an association toward each other and therefore CYN biosynthesis, was undertaken. Additionally, it was also desirable for the homologous regions to contain sufficient differences in the flanking sequence to permit targeting of each individual module. In addition to corroboration of CYN toxicity by gene-directed PCR, toxicity of the majority of samples was established by either HPLCMS/MS or Protein inhibition assays (PI) (Table 3.1). Of the 47 cyanobacterial strains surveyed throughout this work 27 were CYN-producers and 20 were not (Table 3.1, Appendix 2).



Species (number screened)										€				CYN toxicity**
C. raciborskii (14)	+		+	+	+	+	+	+	+	+	+	+	+	+
C. raciborskii (3) AUS	-		-	-	+	-	2- 1+	2- 1+	2- 1+	-	-	-	- +	-
C. raciborskii (7) INT	-		-	-	-	4- 3+	-	4- 3+	6- 1+	6- 1+	-	6- 1+	+ -	-
Aph. ovalisporum (10)	+		-	-	+	+	+	+	+	+	+	+	- -	+
Aph.issatschenkoi (1)	+		+	+	+	+	+	+	+	+	+	+	+ -	-
A.aphanizomenoides (2)	-		+	-	-	-	-	-	-	1- 1+	-	1- 1+	- -	-
Aphanizomenon sp.§ (3)	+					+	+	+	+	+	-	+		+
A.lapponica (1)	-		+	-	-	-	+	+$^{\Delta}$	-	-	-	-	+ -	+
A.bergii (2)	+		+	-	-	+	+	+	+	+	+	+	- -	+
A.bergii (5)	-		-	-	-	-	-	-	-	-	-	-	- -	-
L. wollei (1)†							+				-	+*		+

Figure 3.12 Cylindrospermopsin biosynthetic cluster Gene-directed PCR results

Key:
€ PCR performed using β-ketosynthase primers designed by Fergusson and Saint (2003)
§ Sequences obtained from Dr A. Stüken using regions targeted in this study
Δ *cynC* product amplified successfully using primers described by Kellmann et al (2006)
† PCR performed by Dr B. Sendall, QHSS, using primers from this study
†* PCR product amplified using degenerate in house primers designeb by Dr B. Sendall

In general, the genes surveyed are representative of either structural encoding genes toward the polyketide, alkaloid derived CYN backbone, or tailoring and accessory genes. The role these genes play in CYN biosynthesis will be discussed at greater length in chapter 4, but are briefly outlined in Table 3.6.

Table 3.6 The putative biochemical roles of ORFs identified within the CYN biosynthetic cluster

Gene ID	Closest Similar Protein	Proposed role in CYN biosythesis
orf4	ABC transporter /AAA^+ATPase	Accessory/ /exporter
cynG	Adenylysulfate kinase	Tailoring-involved in the sulfonation of CYN
cynB	PKS	Structural
cynA	PKS	Structural
cynD	Chimeric NRPS/PKS	Structural
cynC	amidotransferase	Structural
cynE	aminotransferase/sulfotransferase	Structural/ tailoring
cynF	aminohydrolase	Structural
orf3	IS4 insertion element	Accessory
cynL	Na^+ efflux pump	Accessory/ regulator
cynH	prolyl 4-hydroxylase	Tailoring-hydroxylating CYN
cynJ	PKS	Structural
cynK	PKS	Structural

Interestingly, the absence or presence of PCR amplicons between species was strongly influenced in respect to the targeted genes putative biosynthetic role. In almost all circumstances, genes that played integral roles in respect to the structural backbone of CYN were readily detected by PCR from all extracts of CYN-producing cyanobacteria (Figure 3.12, Appendix 2). An exception to this was the *Anabaena lapponica*, a Finnish isolate that was first identified as a CYN-producer by Spoof et al (2006). All genomic extracts of *A. lapponica* analysed within this study, failed to amplify products for any of the primer sets that were directed toward the structural genes. Confirmation that the isolate was still a CYN-producer was obtained from three separate chemical analyses from two independent laboratories. Difficulties toward amplification of CYN homologues from *A. lapponica* was also noted by A. Stüken, using primer sets based on the German *Aphanizomenon* sp. 10E6 (personal communication Anke Stüken). A *cynC* homologue from *A. lapponica* was

subsequently amplified and sequenced (sequence similarities 89% *C. raciborskii* AWT205, *cynC* and 89% *Aphanizomenon ovalisporum* ILC-146 *aoaC*), using a primer set described by Kellmann et al (2006). Comparisons between the primer sets designed in this study to target the *cynC* gene, CC1 and CC2 (Table 3.5) and those designed by Kellmann et al (2006) revealed the 5' end of CC1 5bp downstream of Kellmann's CYLATR and the reverse primers priming from different regions of cynC with CC2 located 326 bp downstream of the priming site for CYLATF. A 2.5 KB fragment amplified from the *cynB* gene using new primers AL1 and AL2 (Table 3.4); based on a small fragment previously amplified and sequenced from *A. lapponica* by A.stüken; was obtained from the *A. lapponica* genomic DNA. The 2.5 KB fragment isolated from this study demonstrated high sequence similarity to other identified *cynB* homologues (90% *C. raciborskii* AWT205 and 89% *Aphanizomenon ovalisporum* ILC-146).

Contrary to the observations for the structural genes, genes whose predicted biochemical roles were accessory or tailoring-like functions were detected almost exclusively in extracts from CYN-toxic *C. raciborskii*, but rarely in *Aphanizomenon ovalisporum* or *Anabaena lapponica* (Figure 3.12, Appendix 2). The putative biochemical activities of the ORFs encoding an ABC transporter/AAA+ ATPASE (*orf4*) homologue and a sodium transporter homologue (*cynL*), do not appear to play inherently obvious roles in the CYN biosynthetic pathway. These ORFs were therefore tentatively categorised as accessory genes, as they could putatively be involved in regulation or export of CYN. PCR amplification of the ABC transporter/AAA^+ ATPase ORF from DNA extracts of toxic and non-toxic *C. raciborskii* strains demonstrated that homologous subsequences were present, whether the other *cyn* genes were detected by PCR or not. It is perhaps suggestive that the position of this ORF, being adjacent to the CYN biosynthetic genes may be purely coincidental. Moreover, if there is a found association with CYN production, it may be indicative that this ORF is not implicitly linked to CYN toxicity alone for *C. raciborskii* strains. The sodium transporter was only identified from extracts that were isolated from CYN-toxic *C. raciborskii*. Whilst it is likely this ORF may have some involvement in CYN biosynthesis for *C. raciborskii*, the failure of detection by PCR in other CYN-producing cyanobacteria, may either be indicative of mispriming oligonucleotides due to the presence of polymorphisms or alternative enzymes being employed for similar roles. Intriguingly, ORFs encoding ABC and sodium transporters have also been identified adjacent to, or embedded amongst other cyanobacterial biosynthetic clusters (Tillett et al. 2000a; Rouhiainen et al. 2004; Kellmann et al. 2008b). Furthermore,

disruption of an ORF encoding a putative ABC transporter in a *Microcystis aeruginosa* PCC 7806 resulted in the cessation of microcystin production completely (Pearson et al. 2004)

Two ORFs, whose biochemical activities can be logically assigned to a role in CYN biosynthesis as tailoring activities, a putative adenyl sulphate kinase (*cynG*) and prolyl 4-hydroxylase (*cynH*), were successfully amplified from DNA extracts obtained almost exclusively from both CYN- and non-CYN producing *C. raciborskii*. This was also true upon relaxing the specificity of the PCR annealing (from 55 ºC to 50 ºC) conditions. Three alternative hypotheses may exist to explain this observation (1) it is suggestive of bias toward detection of other cyanobacterial species due to coding differences and possible mis-priming of the primers that were based on the *C. raciborskii* sequence, (2) alternative genes/enzymes involved in tailoring activities of the CYN structure in other cyanobacterial taxa or (3) the genes are not explicably linked to CYN toxicity. Certainly there is some evidence that coding bias may be at play in the detection of gene homologues for CYN biosynthesis, as witnessed with the amplification difficulties toward the *A. lapponica* extracts. Differences in genes coding for tailoring enzymes and gene arrangements are noted between the microcystin synthetase clusters identified from *Plankthothrix agardhii* (Christiansen et al. 2003), *Microcystis aeurginosa* PCC 7806 (Tillett et al. 2000a) and K-139 (Nishizawa et al. 2000) and *Anabaena* 90 (Rouhiainen et al. 2004). Knockout experiments have also begun to delineate roles of genes involved in the biosynthesis of microcystin (Nishizawa et al. 2001; Christiansen et al. 2003; Pearson et al. 2004). Therefore, resolution of these observations in respect to CYN biosynthesis will also lie with the full characterisation of additional CYN biosynthetic cluster from other cyanobacterial species or successful mutagenesis experiments undertaken to elucidate roles of specific genes within the CYN biosynthetic process.

Statistical support for the toxic genotype

In order to establish whether toxicity was independent of genotype, a chi-square contingency analysis was performed. Three genotypes were defined, all positive, all negative or patchy. The taxa were classified in the respective genotype group with consideration only to the presence or absence, of the genes encoding ORFs that were significant with respect to the synthesis of the CYN backbone i.e. PKS (*cynA,cynB,cynJ,cynK*), PS (*cynD*), amidotransferase (*cynC*) and amino-modifying enzymes (*cynE* and *cynF*). All accessory and

tailoring genes were disregarded (*orfs 1,2,3,4,cynL* and *cynG,cynH*) due to possible bias toward *C. raciborskii* strains, evident in the gene-directed PCR. The cyanobacterial species omitted from this analysis was *Lynbgya wollei* as there is an incomplete data set for the majority of the genes analysed. The values obtained are represented in Table 3.7, where expected values are determined by using the corresponding (row total × column total)/grand total. The determined degree of freedom for the given dataset was 2; degree of freedom = (Number of columns-1)(Number of rows-1).

Table 3.7 Chi-square contingency analysis to collaborate CYN toxicity with presence of CYN genes.

			CYN Toxicity		
Genotype	**Positive (observed)**	**Positive (expected)**	**Negative (observed)**	**Negative (expected)**	**Row total**
All positive	27	16.68	1	11.32	28
All Negative	0	5.957	10	4.042	10
Patchy	1	5.361	8	3.638	9
Column total	28	-	19	-	**Grand total 47**

The chi square value was determined using the following formula:

$$c^2 = \sum \frac{(observed - expected)^2}{expected}$$

The obtained chi-square value was 39.3, which when evaluated with the chi-square probability table (Appendix 3), rejects the null hypothesis that genotype is independent of toxicity (P = > 0.001), supporting the hypothesis that the genes identified are synonymous with CYN biosynthesis.

Ideal molecular marker for determining CYN toxigenicity

The majority of primers sets designed to target structural genes identified within the biosynthetic cluster for *C. raciborskii* AWT205 (*cynABCDEFJK*), resulted in the presence of amplification products. There was a strong correlation apparent with respect to PCR amplification and the CYN-toxic phenotype (Figure 3.12 and Appendix 1). However, there were also a low percentage of false positives, with 9 of 20 non-CYN producing strains demonstrating amplicons of expected size, primarily from primers sets targeting *pks* genes. A notable exception with respect to false positives was the *cynE* primers (AMT1 and AMT2 see Table 3.5) targeting a putative aminotransferase/sulfotransferase. Of 47 species surveyed only one (CYP014A), amplified a product from the *cynE* primers, irrespective of its nontoxic phenotype. Crucially, it was also the *cynE* primer set that successfully amplified a product from the genomic extracts obtained from the Finnish isolate, *Anabaena lapponica* 966. Genomic extracts from the same culture material failed to amplify for all PCR reactions initially designed within this study, with exception of primer sets specific for the *cynE* and *cynG* gene regions (Figure 3.12, Appendix 1). The *cynE* targeted primer set was also successful in amplifying a fragment from the CYN-producing Oscillatoriale *Lyngyba wollei*, demonstrating its specificity across cyanobacterial orders as well as genera and species.

Clearly, false negatives are not desirable when developing an assay that is being ultimately designed to assist in the protection of public health. Furthermore, with the prominence of CYN-producing cyanobacteria increasingly becoming a global issue, a universal platform would serve a common purpose. Therefore, the ideal candidate gene for CYN detection would ultimately be one that translates across a range of genera and species, as CYN-producing cyanobacteria have specific geographical significance and progressively more new species are being identified (Stirling and Quilliam 2001; Spoof et al. 2006; Seifert et al. 2007; Yilmaz et al. 2008; Kokocinski et al. 2009). It appears therefore that *cynE* is an attractive toxigenic marker; it was successfully detected within this study, in all cases where CYN toxicity was additionally established by either HPLC or Protein inhibition assays (PI) (all 27 CYN producing species and strains). This included 2 orders within the cyanobacteria, 4 distinct genera and 6 different species, and 11 strains from international sources. Further to this, *cynE* was not readily detected in non-CYN-producing cyanobacteria (1 out of 20 positive). This is suggestive that the role of *cynE* must be relatively constrained within the biosynthetic process and therefore the sequence encoding it must be highly conserved. Genes that demonstrate high levels of polymorphisms can prove to be problematic when optimising molecular probes toward detection, as was demonstrated for the adenylation domain within

the *mcyB* gene in microcystin (Kurmayer et al. 2002), leading to an overestimation of toxin producing species or as observed for *mcyA* (Mbedi et al. 2005), failing to detect the presence of more than one genotype of microcystin-producer.

3.3.1.3.3 Natural deletion mutants

Observations of naturally occurring deletion mutants within populations of microcystin-producing species are not atypical (Kurmayer et al. 2004; Mbedi et al. 2005; Christiansen et al. 2006; Fewer et al. 2008; Tooming-Klunderud et al. 2008). These mutations however, have not always led to the disruption of microcystin production, but have contributed to the variations in chemical phenotype of microcystin-producing *Planktothrix* (Fewer et al. 2008). Kellman et al (2006) acknowledged the identification of a possible natural deletion mutant in a non-CYN-producing *C. raciborskii,* when investigating the phylogeny of CYN-producing cyanobacteria.

From surveying over 40 cyanobacterial species with CYN biosynthetic gene-directed PCR, a small subset of non-CYN-producing cyanobacteria amplified PCR products within this study (appendix 2). To ascertain that the amplification was not an artefact due to mispriming, long template PCR was used to demonstrate that deletion mutants exist within natural CYN-producing cyanobacterial populations. Long template PCR was performed on genomic extracts of CYP018A, as outlined in section 3.2.5, with the positive control fragments amplified from *C. raciborskii* AWT205 and CYN-producing CYP032A. As the CYN producing strains were predicted to contain all the ORFs required in the biosynthesis of CYN, the amplified fragment spanning *cynD* AT and the putative ABC transporter was expected, and found to be approximately 20 kb (Figure 3.13C). The putative deletion mutant CYP018A however, amplified a product that was approximately 6 kb (Figure 3.13B). The truncated product is surmised to represent a deletion occurring within the bounds of the *cynD* AT region and *orf4* putative ABC transporter/AAA^+ ATPase. The 3' and 5' ends of the truncated product were sequence to confirm they were partial genes.

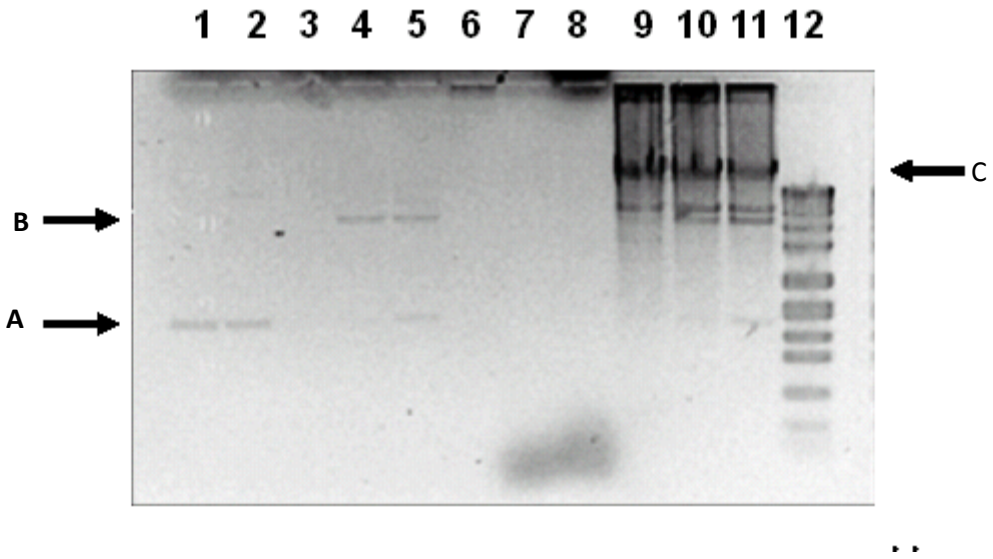


Figure 3.13 Negative image of long-templated PCR to verify CYN gene deletion mutant *C. raciborkii* CYP 018A

Lanes 1 and 2 CYP 014A non-CYN-producing *C.raciborskii* revealing non-specific amplification **(A)**; Lane 3 no template control; Lanes 4 and 5 CYP 018A non-CYN-producing *C.raciborskii* deletion mutant and ~ 6 kb **(B)** amplifiction product; Lane 6 *Microcystis* NIES 102 negative control; Lane 7 and 8 APH35C CYN-producing *Aph. ovalisporum;* Lanes 9 and 10 CYP 032A CYN-producing *C.raciborskii* and ~20 kb amplification product **(C)**; Lane 11 *C.raciborskii* positive control;

Lane 12 Molecular weight marker Spp1/*Eco*R1
(bp):8557,7427,6106,4899,3639,2799,1953,1882,1515,1412,1164,992,710,492

Sequence similarity between the *cynD* and *orf4* PCR subsequence with respect to *C. raciborskii* AWT205 was between 98%-100%, respectively. The sequence obtained from the *cynD* homologue was deposited in Genbank under accession EU216431 (Rasmussen et al. 2008b). Possible mechanistic explanations for the disruption of CYN biosynthesis in strain CYP018A may be connected to transposase activity. There are ORFs identified within the C. *raciborskii* AWT205 biosynthetic CYN cluster, encoding putative IS4 insertion elements (*orf2* and *orf1*) within the corresponding region where the deletion occurred for the strain CYP018A. This insertion element appears to be ubiquitous amongst all strains of non-CYN and CYN-producing *C. raciborskii* analysed thus far. Inactivation of microcystin production in natural populations of *Plankothrix* has been linked to transposons (Christiansen et al. 2006).

These findings contradict Stucken et al. (2009) who nominated that the CYN biosynthetic cluster was transferred into the genome as a pathogenicity island, due to the lack of natural deletion mutants evident in the strains observed in that study. However, this study surveyed a considerably larger dataset in respect to both strains and genes. There is latent support for the pathogenicity island-like theory, through two independent observations during the work presented here. A previously identified CYN-producing *A.bergii* ANA283A, isolated in 1992 from Glenbawn Dam, NSW, Australia, demonstrated absolute and rapid loss of CYN toxicity with corresponding absence of amplifiable gene targets after 14 years in culture (Appendix 2). Amplification of PCR products directed to CYN biogenesis was also rapidly lost from a New Zealand *Aph. issatschenkoi* CAWBGO2that when first isolated seemingly contained the full suite of CYN genes despite CYN never being detected from any of the cultural extracts (Wood et al. 2007b). Whether this could be attributed to the complete loss of the biosynthetic cluster from the genome or, preferential culturing toward a strain that was a non-CYN producer is unclear. However, the potential that partial and inactive CYN genes may still exist in the genome and failed detection as they were positioned outside of the bounds of the PCR-targeted regions, cannot be excluded. Furthermore, there has been no conclusive evidence supporting transferral mechanisms within filamentous cyanobacterial genomes, which can facilitate the movement of large genomic fragments. To date, a type IV Pilus system, is being investigated, to ascertain its potential role in mediating DNA uptake in naturally transformable *Microcystis aeruginosa* (Nakasugi et al. 2007).

3.4 Summary

1. A gene cluster believed to play a significant role in the biosynthesis of the potent cyanotoxin cylindrospermopsin has been sequenced from a *Cylindrospermopsis raciborskii*, strain AWT205. Overall a nucleotide sequence of 42 kb was determined during this study by a suite of molecular methods. Sequence analysis of a final 43 kb region reveals 15 Open Reading Frames (*cynA-cynL*) encoding a family of amino modifying enzymes, tailoring genes, a chimeric polyketide synthase/ non ribosomal peptide synthetase and successive polyketide synthases. This extends from homologous genes identified earlier in *Aphanizomenon ovalisporum* ILC-146.
2. Rigorous screening of over 40 species and strains with gene-directed PCR was undertaken to implicate ORFs in CYN biogenesis. Utilising oligonucleotides designed

from ORFs indentified within the putative *C. raciborskii* AWT205 CYN biosynthetic cluster, revealed strong correlation between CYN toxicity and the amplification of targeted genes. Chi square contingency analysis of the PCR results offers strong statistical support toward CYN toxicity not being independent of genotype ($P > 0.01$). Furthermore, detection of natural deletion events within the CYN biosynthetic molecular framework with consequential disruption of CYN production, for strain CYP018A along with the apparent total loss of amplifiable toxin related genes from ANA283A, offers additional support to the biochemical role of the characterised gene cluster presented in this study.

3. The work presented in this chapter suggests future directions toward the detection of CYN toxigenic cyanobacterial species should target the *cynE* gene, as it appears likely to be the molecular marker least likely to generate false positives. More importantly, *cynE* appears to be unique and specific enough across a range of cyanobacterial species, reducing the occurrence of false negatives. This suggests that future technologies that may be developed utilising this molecular probe will provide increased assurance for water quality managers that all CYN–producing organisms are rapidly detected, affording enhanced protection for the consumer.

4. The putative functional role of *cynE* and other ORFs identified within this study will be outlined at greater length in the following chapter and the biochemical pathway in respect to CYN biosynthesis discussed.

4 BIOINFORMATICS AND PUTATIVE BIOSYNTHETIC PATHWAY OF THE CYLINDROSPERMOPSIN GENE CLUSTER

4.1 Introduction

Cylindrospermopsin (Figure 4.1) is of polyketide origin, as indicated via precursor incorporation experiments investigating the structural backbone of cylindrospermopsin. The feeding experiments revealed five contiguous acetate units and a novel enzymatic pathway leading to the incorporation of glycine into a guandinoacetic acid starter unit. Unambiguous resolution of the origin of the uracil moiety, however, was not achieved (Burgoyne et al. 2000). Polyketide synthase (PKS) involvement was later confirmed through genetic studies (Schembri et al. 2001; Shalev-Alon et al. 2002). Genetic studies have also implicated the involvement of a nonribosomal peptide synthetase (NRPS) in CYN biosynthesis (Schembri et al. 2001; Shalev-Alon et al. 2002).

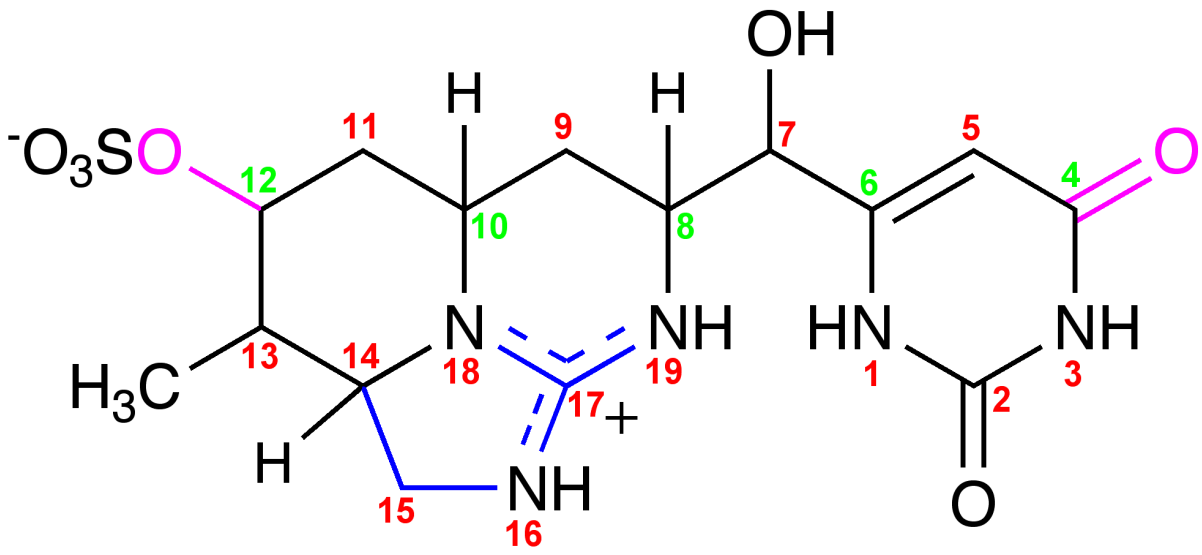


Figure 4.1 Structure of cylindrospermopsin with assigned designation of structural atoms (red)

Green numbering pertains to C1 of incorporated acetate units; pink indicates acetate oxygens retained postmodification,blue indicates the position of the guandinoacetic acid starter unit.

Polyketides are a class of natural compounds that are derived from a series of successive claisen-like, decarboxylative condensation reactions between acyl-CoA subunits by PKSs (Figure 1.2). As discussed in Chapter 1, there are three classical paradigms for bacterial PKSs, Type I (modular), Type II (iterative, non-modular) and Type III (iterative, acyl carrier protein absent) (Shen 2003). Analysis of sequencing data obtained through work presented in Chapter 3, indicated PKS Type I modules being involved in CYN biosynthesis. Type I PKSs are characterised by each successive chain extension being processed by a large modular, multifunctional polypeptide, which house distinct, non-iterative enzymatic domains required for one round of elongation and β-keto processing (Figure 1.2 and Figure 1.3).

Nonribosomal peptide synthethases are large enzymatic complexes that catalyse the synthesis of biologically active peptides, from the condensation of amino, hydroxy, carboxy acid monomers. Whilst viewed as structurally and mechanistically distinct from PKSs, there are also remarkable similarities with respect to organisation and basic functions. The NRPS are also large modular, multifunctional polypeptides, which are organised as successive groups responsible for discrete stages of polypeptide elongation. Within each module exists domains that mediate functional group modifications of the extending polypeptide chain (Cane and Walsh 1999). The adenylation domain recognises and activates the cognate amino acid substrate. Comparisons of docking residues between crystal structures of phenylalanine-activating NRPS adenylation domains, permitted identification of critical residues and prediction of substrate specificity of uncharacterised adenylation domains (Stachelhaus et al. 1999; Challis et al. 2000). Chimeric PKS/NRPS are known and feature in several characterised cyanobacterial secondary metabolite biosynthetic clusters, including microcystin (Tillett et al. 2000a; Christiansen et al. 2003; Rouhiainen et al. 2004), barbamide (Chang et al. 2002) and jamaicamides (Edwards et al. 2004).

Sequencing of the putative cylindrospermopsin gene cluster, presented in Chapter 3, has provided the basic genetic code to explore the proposed CYN biosynthetic pathway. The ORF analyses of the nucleotide sequence revealed a 44.2 kb region with 18 open reading frames. Analysis of this region, coupled with the insights from the radio feeding experiments by Burgoyne et al (2000) permits a sequential, reconstruction toward the biosynthesis of CYN. The findings are presented in this Chapter.

4.2 Specific Materials and Methods

The identification of open reading frames and their putative functions were determined as outlined in Chapter 2 section 2.7.

4.3 Results and Discussion

A 44.2 kb genic region was sequenced, of which 43kb is predicted to fulfil the biosynthetic requirements to manufacture CYN (Figure 4.2). The gene organisation is further detailed with the domains within the PKS and NRPS modules, which are responsible for the chain or side chain modifying activities. Below, the individual ORFs are discussed with respect to bioinformatic information they contain and the interpretation of this data is utilised to relate each ORF with their perceived role in the biogenesis of CYN.

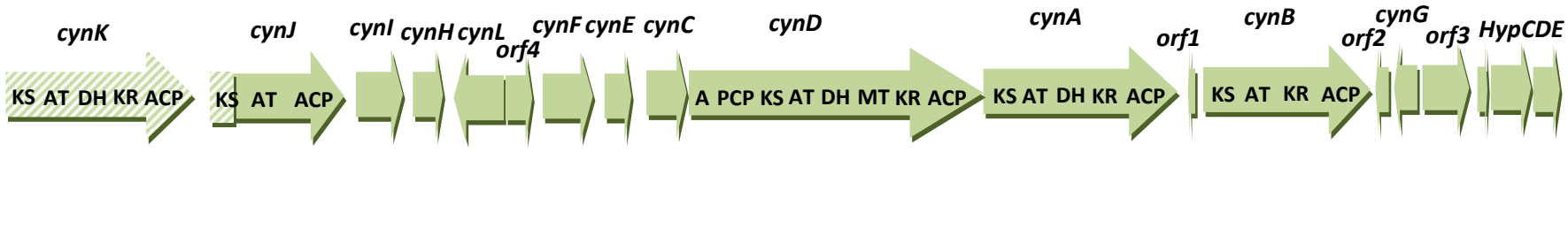


Figure 4.2 Gene organisation of the CYN biosynthetic cluster, with identified domains within the polyketide synthases and nonribsomal peptide synthetase.

Solid green ORFs represents genes identified during this work. Hatched green ORFs represents sequence obtained from Mihali et al (2008). A NRPS adenylation domain; AT PKS acyltransferase; ACP PKS acyl carrier protein; DH PKS dehydratase; KR PKS ketoreductase; KS PKS ketosynthase; MT PKS methyltransferase; PCP NRPS peptidyl carrier protein. Locations of domain are not to scale.

4.3.1 *cynK*-Polyketide synthase (PKS)

As indicated in chapter 3, *cynK* gene was identified and characterised by Mihali et al (2008) and is annotated as *cyrD* within that work. It has been included in this work to allow the predicted biosynthetic pathway for CYN to be fully examined. To permit continuity with respect to the nomenclature of genes identified within this work, was termed *cynK*. The 1877 amino acid (aa) ORF, encodes a 206 435.6 Daltons (Da) polypeptide, with analysis of the predicted primary amino acid sequence revealing core PKS I domains β-ketoacyl synthase (KS), acyl carrier protein (ACP), and acyltransferase (AT), in addition to accessory domains dehydratase (DH) and a β-ketoacyl reductase (KR). Conserved sequence motifs were identified within the AT domain of *cynK* indicating its specificity toward malonyl-CoA (Haydock et al. 1995), (Figure 4.6). Identification of the putative substrates for all AT domains were examined to confirm precursor feeding studies undertaken by Burgoyne et al (2000), which demonstrated the uptake of five acetate units into the CYN structural backbone.

4.3.2 *cynJ*-Polyketide synthase (PKS)

The AUG start codon for the *cynJ* is located 55 bp from the UAG stop codon for *cynK*. The 1358 aa ORF encodes a 148 986.5 Da polypeptide revealing high similarity to other *pks* genes (*Nostoc* sp. ATCC 53789, *CrpB*, 53% identity and 68% similarity, accession no. ABM21570; *Lyngbya majuscula JamK*, 52% identity, 67% similarity, accession no. AAS98782). Analysis of the predicted primary amino acid sequence of the second PKS module revealed core PKS I domains KS, ACP, AT, with no additional accessory domains. Conserved sequence motifs were identified within the AT domain of *cynJ*, indicating its specificity toward malonyl-CoA (Haydock et al. 1995), (Figure 4.6). This is consistent with the precursor studies undertaken by Burgoyne et al (2000), which demonstrated the uptake of five acetate units into the CYN structural backbone.

4.3.3 *cynI*-amidohydrolase

Located 47 bp downstream from the UAA stop codon of *cynJ* is the AUG start codon for a 479 aa (1443 bp) ORF encoding a 51 470 Da polypeptide with high similarity to a class of enzymes of the amidohydrolase superfamily (5-methylthioadenosine/S-

adenosylhomocysteine deaminase, *Clostridiales* bacterium 1_7_47_FAA; 40% identity, 60% similarities). The amidohydrolase superfamily consists of a diverse group of mono or binuclear metalloenzymes, that catalyse the hydrolysis of amide and ester functional groups for a wide range of substrates (Seibert and Raushel 2005). To date there are approximately 20 distinctive identified members of the amidohydrolase superfamily, of which16 have been characterised (Seibert and Raushel 2005). One amidohydrolase member that plays a crucial role in pyrimidine biosynthesis is dihydrooratse (Figure 4.3) (Thoden et al. 2001). It is likely that *cynI* ORF is homologous to dihydrooratse (DHO) and is implicated in the formation of the uracil moiety in CYN. In bacteria, DHO appear monofunctional and homodimeric. Supporting the premise that *cynI* encodes a DHO-like enzyme, an identical ORF, *cynF*, is detected within the putative CYN biosynthetic cluster.

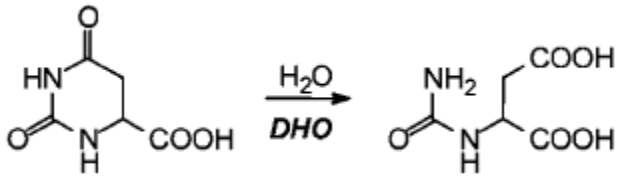


Figure 4.3 Reaction scheme for amide bond hydrolysis catalysed by dihydrooratase (DHO)

4.3.4 *cynH*-Prolyl 4-hydroxylase

The AUG start codon for the fourth ORF (*cynH*), is located 53bp downstream of the UAG stop codon of *cynI*. This 830 bp ORF encodes a 32 624 Da polypeptide (268 aa) demonstrating highest similarity to a hypothetical protein isolated from *Pseudomonas fluorescens* Pf0-1(28% identity, 48% similarity; accession no. ABA75925). Whilst the role of this ORF is not immediately obvious, iterated Psi-Blast searches reveal similarities to numerous prolyl 4-hydroxylase (P4H) alpha subunits, which belong to the 2-oxoglutarate dioxygenase (Fe (II)–dependant oxygenase) superfamily. The roles that these particular enzymes play within eurkaryotic systems are relatively well defined and appear dependent on whether they are located in the cytoplasm or in the lumen of the endoplasmic reticulum (Myllyharju 2003; Koski et al. 2009). The former undertakes more regulatory like roles, whilst the latter plays an important structural role and is actively involved in the hydroxylation of collagen in vertebrates and cell walls of plants (Myllyharju 2003; Koski et al. 2009). Therefore, it is proposed that *cynH* may undertake a comparable role to that of the

structural P4H in *C. raciborskii*, and catalyses the hydroxylation of C7 of CYN. There was a strong correlation with PCR detection of this ORF and CYN-toxic phenotype in all *C. raciborskii* isolates screened in this study (14 out of 14). However, apart from *A. lapponica*, *cynH* was not amplified in the other CYN-producing species (Figure 3. 12, Appendix 2). This is either evidence of sequence divergence between species or alternative enzymes are used to fulfil a similar role in other cyanobacteria.

4.3.5 *cynL*-sodium-dependent multidrug extrusion pump

Transcribed in the opposite direction to the majority of the ORFs within the CYN biosynthetic cluster, is a 466 aa ORF encoding a predicted 50 574.5 Da polypeptide. The ORF reveals 45% identity (66% similarity) to a sodium dependent (Na^+) multidrug and toxic compound extrusion (MATE) pump from *Nostoc punctiforme* PCC 73102 (accession no. YP_001869940). Proteins belonging to this family are widely distributed throughout eurkaryote, archea and eubacterial genomes and provide the "molecular machinery" to rid cells of toxic metabolites (Brown et al. 1999; Putman et al. 2000). The extrusion of typically, neutral or cationic substrates, is coupled with a proton or sodium exchange (Putman et al. 2000). Similar ORFs have been identified adjacent saxitoxin biosynthetic clusters in cyanobacteria and have been proposed to be putatively involved in the exportation of saxitoxin from the cyanobacterial cells (Kellmann et al. 2008a; Mihali et al. 2009). It is anticipated that *cynL* may play a similar role for CYN-producing *C. raciborskii*. Gene-directed PCR targeting *cynL*, was readily amplified in all CYN-producing *C. raciborskii* but was not amplified in any of the non-CYN producing *C. raciborskii*. Therefore, there appears to be a strong correlation toward detection of this ORF and the toxic phenotype. However, PCR amplification of *cynL* homologues was not evident from CYN-producing *Aphanizomenon* (with exception of *Aph. issatschenkoi* CAWBGO2) and *Anabaena* sp. (Figure 3.12, Appendix 2). This may be indicative of sequence divergence for homologous ORF between the species, resulting in decreased efficiencies in PCR amplification, or alternatively it may be indicative of alternate proteins fulfilling similar roles in the other CYN-producing species. Support for the latter conclusion is perhaps evident through comparisons between the four characterised saxitoxin biosynthetic clusters. The MATE ORF homologue, *sxtF* was only identified for the *C. raciborskii* T3; however *A. circinalis* AWQC131C, *Aphanizomenon* sp. NH5 and *Lygnbya wollei* alternatively contained a gene

annotated as *sxtPER* which potentially fulfils a similar role, but did not belong to the MATE family of transporters (Mihali et al. 2009).

4.3.6 *orf 1*, *orf 2* and *orf 3* – IS4 insertion elements

Scattered amongst the "active" CYN biosynthetic cluster, are three interrupted ORFs revealing high similarity to other transposoase, IS4 protein family. Fergusson (2003) identified the first of these ORFs, *orf1*, which is located between *cynA* and *cynB* and transcribed in the opposite direction. The second 354 bp ORF (*orf2*), encodes a polypeptide of 106 aa which reveals 68% identity (77% similarity) to an ORF putatively assigned as a IS4 transposase, in *Cyanothece* sp. PCC 7425 (accession no. YP002478395). It is located downstream of *cynB* and transcribed in the opposite direction. The third 786 bp ORF (*orf3*) encodes a 250 aa polypeptide and also reveals highest identity (70%, similarity 80%) to the *Cyanothece* sp PCC 7425 ORF. It is located downstream from *cynL* and is transcribed in the direction for the majority of CYN genes. Notably, the convergently transcribed *orf2* and *orf3* contain a terminal inverted repeat of 222 bp. Similar insertion sequences have been sequenced adjacent to other secondary metabolite biosynthetic gene clusters (Tillett et al. 2000a; Moffitt and Neilan 2004; Rouhiainen et al. 2004; Kellmann et al. 2008b; Mihali et al. 2009). The presence of transoposase ORFs adjacent to biosynthetic clusters has been nominated as the mechanism responsible for the transfer of these genomic fragments within cyanobacterial lineages (Moffitt and Neilan 2004; Kellmann et al. 2008a). However to date there is no direct evidence available to support this hypothesis.

Gene-directed PCR toward *orf3*, revealed successful amplification only in *C. raciborskii* strains regardless of toxicity. Analysis of microcystin producers *Anabaena* sp. 90 (accession no. AY212249) and *Microcystis aeruginosa* PCC 7806 (accession no. AF183408), revealed no significant sequence similarity between the transposases identified in the respective microcystin biosynthetic cluster. This was also apparent for transposases identified and characterised in saxitoxin producer's *C. raciborskii* T3 (accession no. DQ 787200) and *Anabaena circinalis* AWQC131C (accession no. DQ787201). Therefore the results observed within the CYN biosynthetic cluster are also suggestive of inherent sequence difference within transposases between species. Assuming transpoases are also identified adjacent the CYN biosynthetic cluster for species other than *C. raciborskii*.

4.3.7 *cynF* amidohyrdolase

The AUG start codon for *cynF* is located 51 bp downstream from the *orf3* UAA stop codon. The 477 aa (1473 bp) ORF encodes a 51 695.8 Da polypeptide that also demonstrates high similarity to the amidohydrolase superfamily. This ORF demonstrates an 80% identitiy (89% similarity) with its homodimeric enzymatic partner, *cynI*. Due to the monofuctionality of DHO enzymes it appears that they may make attractive markers, particularly if these *cynI* and *cynF* represent a novel biosynthetic pathway toward the assemblage of pyrimidine derivatives within CYN biosynthesis. Gene-directed PCR targeting this ORF demonstrated great specificity toward CYN-producing cyanobacteria, generating low false positives (3 out of 20 non CYN-producing cyanobacteria). Amplifications were only observed in non-CYN-producing *C. raciborskii* isolates from international waterbodies (Figure 3.12, Appendix 2). However, the homologous genes from the CYN-producing *A. lapponica* failed to be detected by PCR, suggestive that there are inherent sequence differences in *A. lapponica* that led to mispriming and subsequent failure in PCR amplification.

4.3.8 *cynE*-putative aminotransferase/sulfotransferase

Located 558 bp downstream of the UAG stop codon of *cynF* is the AUG start codon of the *cynE* ORF. This 816 bp ORF encodes a 30 307.7 Da polypeptide of 260 aa with a 42% identity and 57% similarity to a hypothetical protein identified in *Lyngbya* sp PCC 8106 (accession no. ZP_01619244). Supplementary iterated Psi-Blast searches reveal that the ORF displays moderate similarity to numerous enzymes with either putative branched-chain amino acid aminotransferases (31% identity and 50% similarity to a 246 aa ORF identified in *Congregibacter litoralis* KT71, accession no. ZP_01104458) or putative sulfotransferase (30% identity and 47% similarity to a 276 aa ORF identified in a saxitoxin-producing *C. raciborskii*, accession no.ABI75104) functionalities, from a variety of species.

It is dubious whether the *cynE* biochemical role could be that of a branched-chain amino acid transeferases (BCAT). It appears these particular transferases are specific toward the metabolic transamination of branched-chain amino acid precursors for, L-leucine, L-isoleucine, L-valine and typically use glutamic acid as the amino donor (Hutson 2001).They are also the first step toward the catabolism of these same branched-chain amino acids, catalysing the formation of α-keto-acid derivatives. However, evidence suggests bacterial

BCATs display acceptance to a broader range of substrates then mammalian homologues (Cheng et al. 1999; Yvon et al. 2000).

A signature sequence, based on the 3D structure of a BCAT first isolated from *E. coli* (Okada et al. 1997) and later supported from a homologue isolated from a sheep (Okada et al. 1997; Bonfils et al. 2000), has been determined which plays a pivotal role in coordinating the enzyme cofactor for this class of enzymes, pyridoxal 5'-phosphate (PLP). The sequence, encoding for 9 amino residues:

(EX(S/T/A/G/C/I)X(2)N(L/I/V/M/F/A/C)(F/Y)X(6,12)(L/I/V/M/F)XTX(6,8)(L/I/V/M)X(K/R)),

is located toward the C-terminus, however, none of the amino acid residues appeared conserved in similar relative positions for the *cynE* ORF. Nevertheless, considering the proximity of *cynE* to *cynC*; an ORF encoding an enzyme that is involved in the catalytic formation of guanidinoacetic acid (Figure 4.4); it cannot be definitively excluded as not being involved in animating a novel precursor of the guanidinoacetic acid starter unit. Burgoyne et al (2000) were unable to determine the amidino donor for the starter unit during radio-labelled incorporation studies, which is typically found to be L-arginine (Bedekar et al. 1998).

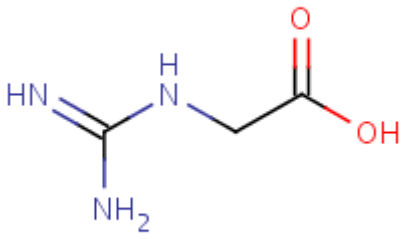


Figure 4.4 Structure of guanidinoacetic acid.

Red indicates the gylcine derivatives and blue is indicative of the amidino nitrogens

To date, there have only been assumptions made in respect to the L-arginine catabolic pathways in cyanobacteria (Kellmann et al. 2006; Schriek et al. 2007). Direct biochemical evidence is required to resolve the anabolic pathway of guanidinoacetic acid and support L-arginine as the donor molecule in CYN-producing cyanobacteria.

Alternatively, *cynE* could encode a sulfotransefase, a large family of enzymes that coordinate the transfer of a sulfuryl (SO_3) group from a donor molecule typically a, 3'-phosphoadenosine 5'-phosphosulfate (PAPS), to a range of endogenous or exogenous biomolecules (Chapman et al. 2004). The majority of the sulfotranserases studied have been isolated from mammals, with relatively few examples in prokaryotes (Mougous et al. 2002). There are effectively two recognised classes of sulfotransferases, cytosolic or membrane bound, which appear to share very little sequence similarity. However, (Kakuta et al. 1998) identified conserved amino acid motifs common to both, corresponding to the N-terminal region which is believed to interact with the 5' phosphate group of PAPS (5' PSB-loop: PK(T/S)GTT(W/A)L) and C-terminal motif coordinating the 3' PAPS phosphate (3'PB: I(T/Y)(V/I/L)LRNP(A/K)D(R/V)(L/A)(V/Y)SYY(Y/Q)). The 3' PB motif is also potentially involved in orientating the PAPS correctly and plays an influential role in directing the specificity of the binding substrate. Four of eight amino acid residues were identified in similar relative positions determined for the 5'PSB-loop motif (Figure 4.5). However there was no obvious 3'PB motif which is perhaps reflective of this regions putative role in substrate specificity.

```
C.raciborskii_A2        RIAMWSVPRSLGTVLLQAWSSRP
T.erythraeum            PIFIFGCPRSGTTLLQSILATHP
C.raciborskii_T3        HIAMWACPRSRSTVITRAFENLD
L.wollei                HIAMWACPRSRSTVITRAFENLD
A.circinalis            HIAMWACPRSRSTVIARAFENLD
mouse                   DLLILTYPKSGTNWLNEIVCLIQ
plant                   DVFLASYPKSGTTWLKALAFAII
Rhizobium               PFAILAMPRTGTHYLEVLLNEHP
Human_hydroxysteroid    DVIILTYPKSGTNWLAEILCLMH
Human_tyrosylprotein    LIFVGGVPRSGTTLMRAMLDAHP
Human_carbohydrate      HVLVLSSWRSGSSFLGQLFSQHP
Human_aryl              DLLISTYPKSGTTWVSQILDMIY
                          . :    ::     :
```

Figure 4.5 Sequence alignment of 5' PSB-loop motifs from sulfotransferases

PSB-loop motifs are boxed. Sequene comparisons between *C. raciborskii AWT205* (*C. raciborskii* _A2*) CynE, Trichodesmium erythraeum* IMS101 sulfotranseferase (accession number YP_723602**)**, *C. raciborskiiT3* SxtN (accession number *ABI75104*), *L.wollei* SxtN1(accession number ACG63828**)**, *A.circinalis* AWQC131C SxtN (accession number ABI75135), *Mus musculus* phenol sulfotransferase (accession number AAC17740),plant flavonol *Flaveria chloraefolia* (accession number AAA33343), *Rhizobium* sp. N33 NodH(accession number U53327), *Homo sapien* hydroxysteroid sulfotransefarse (accession number Q06520) *Homo sapien* tyrosyl protein sulfotransferase (accession number AAH17509), *Homo sapien* carbohydrate sulfotransferase (accession number AAH22567), *Homo sapien* aryl protein sulfotransferase (accession number GO1843)

It is strongly suspected that *cynE* has a crucial and unique biochemical role in CYN biogenesis. Gene-directed PCR toward this ORF, unambiguously amplified products of expected size in all CYN-producing species investigated during this work. This suggests its role is highly conserved. It appears likely that it is implicated in the sulphonation of the CYN structure more so then a putative aminotransferase-like role. Support for this supposition lies with the identification of a motif resembling one that is common for other sulfotransferases.

4.3.9 *cynC*-amidinotransefase

The AUG codon of the ninth ORF (*cynC*) is 439 bp downstream of the UAG stop codon of *cynE*. The ORF encodes a 392 aa polypeptide that displays high similarity to an ORF (*aoaA*) isolated from a CYN-producing *Aphanizomenon ovalisporum* (95% identity and 97% similarity, accession no. AF395828) and a homologue previously isolated by Kellmann et al (2006), also from *C. raciborskii* AWT205 (100% identity and 100% similarity, accession no. AAX81900). The homologue isolated from the *Aph. ovalisporum* was identified as encoding a putative amidinotransefase (Shalev-Alon et al 2002), an enzyme which is responsible for the catalytic formation of guanidinoacetic acid in other biosynthetic pathways. Modelling of

both the *aoaA* and *aoaB* homologue in *C. raciborskii* supported the guanidinoacetic acid as the substrate for these ORFs (Kellmann et al. 2006). Burgoyne et al (2000) had predicted earlier that the anticipated starter unit for the CYN was guanidinoacetic acid from precursor incorporation experiments. However Burgoyne et al (2000) could not determine the putative donor of the amidino group, though it appears that arginine is the preferred donor in similar systems (Bedekar et al. 1998). The unique biochemical role that *cynC* is anticipated to play in CYN biogenesis, also lends itself to being an attractive toxigenic marker. However, there were 5 out of 20 non CYN-producing cyanobacteria positively amplifying a PCR product of expected size. This may be suggestive of either remnants of the CYN gene cluster remaining in the genomes post a deletion event, or an alternative biosynthetic pathway exists within the genome that utlises a gene with high similarity to *cynC*. These results may indicate that *cynC* is not the most discrete marker for CYN toxicity.

4.3.10 *cynD*-Chimeric PKS and nonribosomal peptide synthatase (NRPS)

The 10th and largest ORF (8826bp) for the CYN biosynthetic cluster, encodes a 322 468 Da polypeptide (2918 aa), demonstrating high similarity to other proteins of chimeric PKS and NRPS functionality (*Aph. ovalisporum*, CYN homologue, *aoaB*, 97% identity, 97% similarity, accession no. AAM33470; *Stigmatella aurantiaca* DW4/3-1, 39% identity, 55% similarity, accession no. CAQ34918). The *cynD* ORF overlaps with the UAG stop codon of *cynC* and it appears that the translation for this ORF is initiated from the second AUG which is located 14 bp from the *cynC* stop codon. The predicted amino acid sequence reveals at the N-terminus region a NRPS module composed of an adenylation domain and a peptidyl carrier protein (PCP). The adenylation domain was analysed for substrate specificity by Kellmann et al (2006) using an algorithm developed by (Challis et al. 2000), however there is no precedence in the database for a guanidinoacetic acid and only tentative conclusions about the binding interactions toward the guanidinoacetic acid substrate could be drawn. The NRPS is fused to a PKS component. The PKS module consists of a KR, AT, DH, *S*-adenosylmethionine (SAM) dependent methyltransferase (MT), KR and ACP domains, which are responsible for the condensation of the first acetate unit and β-keto processing of the guanidinoacetic acid starter unit.

4.3.11 *cynA*-Polyketide synthetase (PKS)

The AUG start codon for the 11th ORF *cynA* is located 18 bp from the UAG stop codon for *cynD*. This 5685bp ORF encodes a 210 727 Da polypeptide which demonstrates high similarity to other polypeptides identified as type I PKS. It is the fourth PKS module of a predicted five, and includes the core domains KS, ACP, and AT, along with additional accessory domains, KR and DH. Based on conserved sequence motifs (Haydock et al. 1995), the AT domain of *cynA* appears to accept malonyl-CoA (Figure 4.6).

4.3.12 *cynB*-Polyketide synthetase (PKS)

The AUG start codon for the 11th ORF *cynA* is located 18 bp from the UAG stop codon for *cynD*. This 5685bp ORF encodes a 210 727 Da polypeptide which demonstrates high similarity to other polypeptides identified as type I PKS. It is the fourth PKS module of a predicted five, and includes the core domains KS, ACP, and AT, along with additional accessory domains, KR and DH. Based on conserved sequence motifs (Haydock et al. 1995), the AT domain of *cynA* appears to accept malonyl-CoA (Figure 4.6).

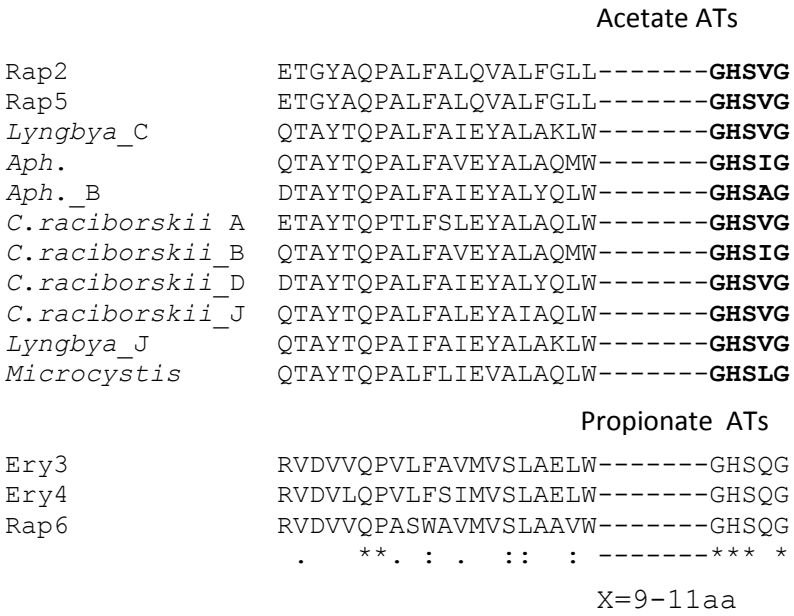


Figure 4.6 Alignment of sequences for motifs found in AT domains in PKS type I (Haydock et al 1995)

Rap 2, 5 and 6 represent different AT domains for rapamycin biosynthesis, *Streptomyces hygroscopicus* (accession number CAA60462), *Lyngbya majuscula* CurA for curacin A biosynthesis (accession number AAT 70096), *Aph. ovalisporum*, AoaC and AoaB, CYN biosynthesis (accession number FJ234859), *C. raciborskii* AWT205 *CynA-CynJ* (this work), *Lyngbya majuscula* JamL for jamaicamides biosynthesis (accession number AAS98783), *Microcystis aeruginosa* PCC7806, McyG (accession number CAO90231), Ery 3 and 4 represent *Saccharopolyspora erythraea*, 6-deoxyerythronolide B synthase (DEBS) from module EryA (accession number M63676)

4.3.13 *cynG*-adenylylsulfate kynase

Analysis of a region downstream of *cynB* reveals an ORF that encodes a 200 aa polypeptide, transcribed in the opposite direction to the majority of the gene cluster. The ORF demonstrates high similarity to adenosine 5'-phosphosulfate kinases (APS), alternatively referred to as adenylylsulfate kinase (ASK). The ASK kinases are involved in the second and ultimate step toward the conversion of inorganic sulphate into a biologically active form (MacRae et al. 2000). The biosynthesis of 3'-phosphoadenosine-5'-phosphosulfate (PAPS) is represented below:

$$MgATP + SO_4^{2-} \xrightarrow{\text{ATP sulfurvlase}} APS + PPi \quad (1)$$

$$MgAPS + ATP \xrightarrow{\text{ASK kinase}} PAPS + ADP \quad (2)$$

The transfer of the sulfuryl group (SO_3) from PAPS to an amino or hydroxyl substrate is mediated by sulfotransferases. The characteristic C-terminus purine-binding walker A (GLSASGKS) and N-terminus PAPS 3' phosphoadenosine 5' phosphosulfate motifs (KAREGVIKEFT) are readily identifiable in the *C. raciborskii* ASK ORF (Figure 4.7) (Satishchandran et al. 1992). The ASK kinases are found to be highly conserved across all species at an amino acid level (Chapman et al. 2004). Gene-directed PCR detected the ASK ORF only in CYN-producing *C. raciborskii*, with exception of internationally sourced non-CYN–producing *C. raciborskii*, which also amplified a product of expected size (Figure 3.12, Appendix 2). This is probably indicative of coding differences between the CYN-producing species. Very low nucleotide sequence similarity was found to exist between the *ask* homologues for the saxitoxin-producing *C. raciborskii* T3 (accession no. DQ787200) and *Anabaena circinalis* AWQC131C (accession no. DQ787201) (data not shown). Alternatively there are identified sulfotransferases that utilise alternative sources of activated sulphur (Chapman et al. 2004).

```
C.raciborskii        FTGLSGSGKTLD-------YAKARRGEIESFTGI
Proteobacterium      LTGLSGAGKSLD-------YKKARKGEISEFTGV
S.hominis            FTGLSGSGKSLD-------YEKARSGEIKEFTGI
S.warneri            FTGLSGSGKSLD-------YKKARTGEIPEFTGI
Synechoccocus_9605   FTGLSGSGKSLD-------YKARAGQIKEFTGI
Synechoccocus_9902   FTGLSGSGKSLD-------YAKARSGAIKEFTGI
Synechococcus_9916   FTGLSGAGKSLD-------YAKARAGEIKEFTGI
C.raciborskii_T3     FTGLSGAGKTLD-------YLKARSGEIDGFTGI
                     :*****:**:**-------* *** * *  ***:
                                (X=104)
```

Figure 4.7 Sequence comparisons of adenylsulfate kinase

C.raciborskii AWT205 *CynG*, α-proteobacterium BAL199 NodQ(accession number ZP_02187262), *Staphylococcus hominis* SK119 (accession numberZP_04060185), *Staphylococcus warneri* L37603 (accession number ZP_04678803*), Synechococcus* sp. CC9605 (accession number YP_382707), *Synechococcus* sp. CC9902 (accession number YP_378101), *Synechococcus* sp. RS9916 (accession number ZP_01471235), *C.raciborskii* T3 SxtO (accession number ABI75115).

4.3.14 *orf4*-AAA + ATPases

Analysis of a region downstream of the *cynG*, revealed an ORF encoding a 457 amino acid (aa) polypeptide, whose function is not clearly assignable. The ORF demonstrates highest similarity to a 581 aa ORF encoded by *Microcystis aeruginosa* PCC7806 (63% sequence identity, 74% similarity). The *Microcystis* polypeptide also has no function assigned, with the exception of a conserved domain within the N-terminal, demonstrating homology to P-loop containing nucleoside triphosphate hydrolases (Genbank accession CAO86728). The P-loop NTPases superfamily encompasses enzymes that include , nucleotide kinases, ATP- binding cassettes (ABC transporters) and AAA^+ ATPases amongst others. These enzymes are involved in chaperone like functions that mediate protein to protein interactions and the assembly and disassembly of multiprotein complexes (Neuwald et al. 1999). A gene, *mcyH*, belonging to this superfamily and encoding a ABC transporter, has similarly been identified adjacent microcystin biosynthetic clusters (Tillett et al. 2000a; Christiansen et al. 2003; Rouhiainen et al. 2004). Disruption of *mcyH* in *Microcystis aeruginosa* PCC 7806 resulted in the subsequent loss of toxicity and supported the genes implicit involvement in the biosynthetic pathway (Pearson et al. 2004). Likewise, *orf4* may also function in the capacity of a regulator or exporter of cylindrospermopsin and be intimately involved in the biosynthetic pathway. However, PCR screening to determine the presence of this ORF in CYN and non-CYN producing species and strains (Figure 3.12, Appendix 2) revealed it as specific to *C. raciborskii*, regardless of the presence of the other *cyn* genes. Whilst the absence of PCR amplification in all CYN-producing cyanobacterial can be representative of

coding bias as discussed in chapter 3 (section 3.3.1.3), the supposition that this ORF is complicit with CYN biogenesis is not clearly definable and therefore was not assigned a *cyn* prefix to denote its involvement.

4.3.15 Hydrogenase expression cluster *hypC-E*

The 3' terminus of the putative CYN biosynthetic cluster is flanked by three ORFs encoding hydrogenase expression genes (*hypC*, *hypD*, *hypE*). The Hyp genes (encoded by *hyp ABCDEF*) are involved in the maturation for uptake and bidirectional [NiFe] hydrogenases in cyanobacteria (Tamagnini et al. 2007). Uptake hydrogenases catalyse the uptake of H_2 that is produced during the reduction of N_2 to NH_3 by nitrogenase(s).The maturation process is complex and involves several proteins to coordinate the incorporation of the metal ions (Ni, Fe) and CO and CN ligands into the active centre of the [NiFe] hydrogenases (encoded by *hup*SL), the coordinating of FeS clusters and the cleavage of the complex from the C-terminus of the larger subunit (*hup*L).

Previous investigations into hyp gene homolgues in *E. coli* have identified conserved motifs that are possessed by corresponding proteins across different organisms and cyanobacterial homologues (Leitao et al. 2006). The first of the *C. raciborskii* AWT205 ORFs, encodes a 93 aa polypeptide which contains a *HypC* motif MC (L/I/V) (G/A) (L/I/V) P, at the N-terminus (MCLGIP). The second ORF encodes a 376 aa polypeptide which reveals 5 cysteine residues and a CPVC motifthat are characteristic of *hypD*. These cysteine residues play integral roles in coordination of the Fe ion. The third ORF encodes 249 aa polypeptide and appears to be a homologue of *hypE*. The ORF for *hypE* is incomplete, determined by both the absence of the characteristic C-terminus motif, PRIC and lack of stop codon. Comparisons with other cyanobacterial *hypE* homologues submitted in Genbank, indicate that the ORF is approximately 139 residues short. All the *hyp* genes exist as one operon and are transcribed in the same direction of *orf4* and the CYN biosynthetic genes (with one exception, *cynG*).

In cyanobacteria, *hyp* genes are frequently clustered together and located in the vicinity of uptake hydrogenase structural genes (*hup*SL), though there are exceptions where the *hyp ABCDEF* genes are scattered throughout the genome (Leitao et al. 2006; Tamagnini et al. 2007). The order of *hyp* genes in other N2-fixing cyanobacteria is generally found to be *hyp FCDEAB*. It appears that this order is disrupted for *C. raciborskii* AWT205, because an ORF

encoding a *hypF* homologue was not identified upstream of *hypC*, however *orf3* was. Interestingly, ORFs with similar putative functional roles as that predicted for *orf3*; involving assembly and coordination of multi-protein complexes; have been identified adjacent to or scattered amongst *hyp* gene clusters in other cyanobacteria genomes (Leitao et al. 2006). This is perhaps suggestive that *orf3* may potentially not be involved in CYN biosynthesis, and is perhaps more closely associated with the *hyp* gene cluster. This premise maybe further supported by the presence of *orf3* in the majority of both CYN and non-CYN producing *C. raciborskii* through the detection of amplification products using gene-directed PCR (Figure 3.12, Appendix 2).

4.4 The proposed biosynthetic pathway

The PKS enzymatic architecture for the CYN biosynthetic pathway appears modular and should obey the classic modular PKS I co-linearity rule: that active sites responsible for one round of condensation and β-keto processing are housed within a multifunctional polypeptide and are transcribed in an order representative of their inclusion in the resultant protein (Donadio et al 1991). Whilst for *C. raciborskii*, the progression of the PKS and tailoring ORFs involved in CYN biogenesis appears to be transcribed in a co-linear manner they however, are not necessarily reflective of the temporal order of the biochemical events. Frequently, the organisation of the genes is found in the same order as the expected incorporation of their resultant protein into the metabolite (Cortes et al. 1990; Kuhstoss et al. 1996), however this is not a requirement (Ikeda et al. 1999). This is further exemplified by the rearrangements and divergent translation already apparent in the partially elucidated *Aph. ovalisporum* CYN biosynthetic gene order (Shalev-Alon et al. 2002). To assist in the determination of which ORF is responsible for the extension and processing of each successive acetate unit, the precursor studies undertaken by Burgoyne et al (2000) are a useful starting point. Figure 4.1 represents the CYN structure with numbering and highlighted features derived from the precursor's studies.

The first step toward the synthesis of the CYN backbone is the formation of the guanidinoacetic acid starter unit. Precursor studies undertaken by Burgoyne et al (2000) demonstrated the incorporation of labelled guanidino into N16 and C17, with N16 derived

from an intact labelled glycine, which also corresponded to C14 and C15. An amidinotransferase encoded by *cynC*, is expected to catalyse the synthesis of guanidinoacetic acid through the transamidation of glycine by an as yet unequivocally established donor; though it appears likely to be arginine (Bedekar et al. 1998; Kellmann et al. 2006). The subsequent biosynthetic steps involves the successive condensation of 5 acetate units to the guanidinoacetic acid starter unit, with is assembled by the identified modular PKS type I ORFs *CynD*, *CynA*, *CynB*, *CynJ* and *CynK*. Protein sequence analysis of the acyl transferase domains (AT) for the PKS modules, indicated the loading of malonyl CoA as the extender unit (Figure 4.6), as predicted by Burgoyne et al (2000).

The chimeric *CynD* is the only ORF that has the ability to recruit the guanidinoacetic acid starter unit. The guanidinoacetic acid is uploaded by the NRPS adenylation domain of *cynD* where it is covalently attached and activated as a thioester. The activated guanidinoacetic acid is transferred by the swinging arm of the peptidyl carrier protein (PCP) to the KS domain of the *cynD* PKS component. The KS domain mediates the condensation by decarboxylative acylation of the malonyl donor unit; which is primed by the AT domain; with the activated guanidinoacetic acid. There are three auxillary domains in *cynD*, two of these, dehydratase (DH) and a ketoreductase (KR) leads to the full reduction of the C14 keto group to an alkene bond between the acetate derived C13 and the glycine C14. The third auxillary domain is a methyltransferase (MT) which is responsible for the methylation of C13 from the C_1 pool via *S*-adenosylmethionine, as predicated by the radiofeeding studies (Burgoyne et al 2000).

The elongation of the growing polyketide chain with the second acetate unit is performed by an ORF that catalyses only one additional auxillary step. The precursor studies indicated that C12 retained the oxygen from the acetate therefore; it does not undergo full reduction, being reduced only to a hydroxyl. Only *cynB* contains a KR domain without any additional auxillary domains, so it is predicated to be the enzyme that receives the *cynD* product. The C12 requires additional tailoring outside of the PKS activity and undergoes sulphonation, putatively by *cynE* with *cynG* providing the sulphur donor group through the biosynthesis of PAPS. The sulphonation of C12 is likely to be performed when the formation of the structural backbone is complete.

The ORF responsible for the decarboxylative condensation of the third malony-CoA unit also leads to the full reduction of the C10, so the module must contains both a KR and a DH. This

results in an alkene bond between C10 and C9. This is further supported by the precursor studies indicating that C9 retained only one deuterium. The PKS downstream of *cynD*, *cynA* contains the mandatory domains KS, AT and ACP but also the auxillary domains, DH and KR. However, *cynK* may also fulfil this biochemical requirement, as it too houses the domains that facilitate the dehydration of the hydroxyl to an alkene, and a homologous gene in the German CYN-producing *Aphanizomenon* sp. has a start codon overlapping with the stop codon in the equivalent *cynB* homologue, suggesting they maybe concerted (A.Stüken, personal communication). Therefore *cynA* and *cynK* are responsible for the next two elongation steps and catalyse the full reduction of the keto groups, forming alkene bonds between C9 and C10 and also C8 and C7. However it is difficult to determine what is the actual temporal order of the enzymes. The precursor studies indicate that the hydroxyl group on the C7 is not of acetate origin (Burgoyne et al 2000), therefore the C7 is hydroxylated by *cynH* which encodes Prolyl 4-hydroxylase.

The final PKS is a minimal module, which participates in the uploading of the 5th acetate unit, but as it only contains the mandatory domains (KA, AT, ACP), the C4 and C6 keto groups remain unreduced. The unreduced keto groups are the reactive sites toward the formation of the uracil ring that is catalysed by the homodimeric polypeptides encoded by *cynF* and *cynI*.

The cyclisation of the guanioacetic group with the polyketide backbone requires three enzymatically controlled dehydration steps, in a laboratory similar reactions require rigorous conditions (Murphy and Thomas 2001). Two of the PKS genes (*cynA* and *cynK*) along with the PKS module contained within the intergrated NRPS/PKS *cynD*, house DH enzymatic domains, whose activities lead to the dehydration of the β-keto carbonyl. This processing results in the formation of a reactive alkene bond, which presents as an attractive nucleophilic target for the nitrogens of the amidino group. Whether cyclisation of the overall structure is spontaneous or activated by an as yet unidentified specialised enzymatic domain, is a matter of conjecture. There was no obvious intrinsic, integrated thioesterase domain (TE) evident within the CYN biosynthetic cluster, which is usually observed at the C-terminus of the final PKS extension module in similar systems (Ikeda et al. 1999). Certainly, the modelling studies additionally undertaken by Burgoyne et al (2000) have indicated that a conformation of the nascent intermediate is one of minimum energy, possibly leading to spontaneous cyclisation and release. Cyclisation may occur either sequentially throughout the processing or post

production of the compound. Furthermore, there is experimental evidence demonstrating TE domains for enzymatic product release and cyclisation, are not always an essential requirement, though the majority of these spontaneously released products appear to be relatively simple triketide intermediates (Bindseil and Zeeck 1993; Kao et al. 1995). The majority of polyketide synthase complexes appear to undergo enzyme-catalysed cyclisation, termination and release (Chang et al. 2002; Christiansen et al. 2003). Additionally, TE domains have also been recognised as important editors for recognising aberrant polyketide products and can be important in dictating metabolic rate (Cortes et al. 1995; Butler et al. 1999). Therefore whilst a corresponding TE domain cannot be identified in the CYN gene cluster, it cannot be excluded that a TE domain is potentially recruited elsewhere from within the chromosome. Alternatively there may be an as yet unidentified discrete TE within the cluster that fulfils this role for CYN. However, to date, discrete TE domains have only been identified within PKS type I polyether biosynthetic clusters (Harvey et al. 2006; Liu et al. 2006).

The putative biosynthetic pathway for the production of CYN is presented in Figure 4.8. Overall it is in general agreement with the findings of Mihali et al. (2008). However, there is some reservation with confidently assigning the biochemical role of *cynE* and the consideration of *orf4* as having any active involvement in the overall biosynthetic pathway.
To facilitate visualisation of the sequential construction of CYN, the cyclisation steps are detailed as successive. However, this representation is not conclusive of the actual temporal biochemical events leading to the final product. This may only be possible after future mutagenesis experiments.

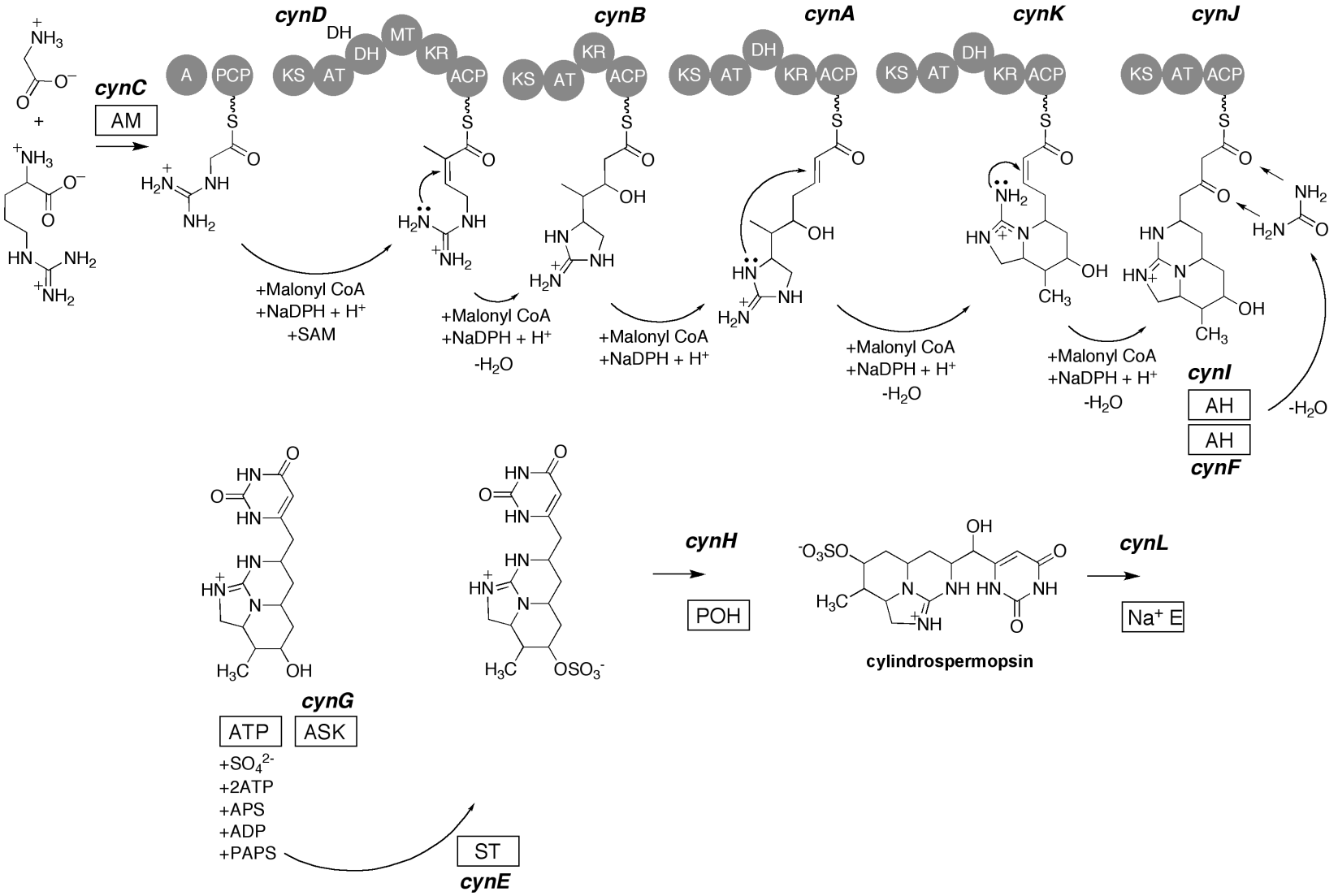


Figure 4.8 The proposed biosynthetic pathway for the production of cylindrospermopsin

4.5 Summary

1. The genetic architecture and domains are organized in a manner that supports the perceived co-linear assignment of the final molecular structure. However, not all ORFs identified within the cluster can be unequivocally assigned a biochemical role.

2. Rigorous PCR screening of genetic determinants undertaken in both CYN toxigenic and non-CYN toxigenic cyanobacteria (Chapter 3), coupled with the gene organization being concomitant with the structural architecture of the compound (Chapter 4), presents a strong case in implicating these genes. Of the 18 ORFs identified, 12 appeared to play a functional role toward the biosynthesis of CYN in *C. raciborskii*, and have been annotated *cynA-cynL* accordingly.

3. The nucleotide sequence presented in this thesis provides the basic information to begin to unravel the regulatory and functional utility of cylindrospermopsin. This information will better facilitate the development of rapid detection strategies toward both the toxin and identification of its producer organisms. Moreover, water quality managers will be armed with a deeper understanding of why and when cyanobacteria are likely to produce CYN, which will facilitate the development of pre-emptive and cost-effective management strategies in the future.

4. The molecular information identified through components of this work will be exploited in the next chapter to obtain insights into the mechanisms of a CYN-producing cyanobacterial bloom, which occurred in the River Murray over the summer season, 2008.

5 INVESTIGATION OF A CYLINDROSPERMOPSIN PRODUCING *CYLINDROSPERMOPSIS RACIBORSKII* BLOOM IN THE LOWER RIVER MURRAY

5.1 Introduction

The previous research chapters explored, at a molecular level, the characterisation of a 43 kb gene cluster, whose molecular architecture was amenable to the anticipated synthetic pathway for the cyanotoxin cylindrospermopsin (Burgoyne et al. 2000). During the course of this work an opportunity arose to investigate a cylindrospermopsin-producing *Cylindrospermopsi*s *raciborskii* bloom, which was developing in the River Murray. This opportunity permitted the practical application of the molecular knowledge acquired in Chapter 3, through the implementation of a real-time PCR assay developed within our laboratory (Rasmussen et al. 2008b).

Real-time PCR enables simultaneous monitoring of exponential amplification of more than one specific gene target. *In situ* monitoring capabilities are possible through the observation of increased fluorescence, resulting from the 5' exonuclease activity releasing a template-specific fluorescent probe, bound to an internal region of the PCR product. Furthermore, quantification is achievable when end products are analysed in conjunction with appropriate DNA standards (Saunders 2004a). The fundamentals of real-time PCR were discussed in greater depth within chapter 1 (section 1.6.2).

Current cyanobacterial orientated research, employing quantitative real-time PCR (qPCR), has been primarily used to investigate the dynamics of microcystin-producing cyanobacterial blooms (Vaitomaa et al. 2003; Briand et al. 2008; Ha et al. 2009). This is reflective of the worldwide prevalence of microcystin and the microcystin-producing cyanobacteria, in addition to the better characterisation of biosynthetic pathways. However, to date there is limited information regarding bloom-forming cylindrospermopsin-producing cyanobacteria at the molecular level.

It is desirable for water quality managers to be better informed about cyanobacterial issues that exist within water bodies, so they are able to appropriately manage bloom incidents. Currently, management of blooms comprise of reactionary, more than proactive responses.

Strategies often employ detection methods that are time expensive and laborious, whilst remedies may involve the application of heavy metal algicides (Chorus and Bartram 1999). Algicides lead to the lysis of cyanobacterial cells and subsequent release of intracellular toxins into the water column. This can result in delays to the water being restored to the reticulated supply, and furthermore the use of heavy metals holds environmental consequences. However, the use of molecular techniques will enable development of assays that are adaptable to platforms that are portable, time efficient and permit sensitive, unambiguous detection that may facilitate species-specific and time crucial resolutions. Integrative research encompassing the molecular and physical dynamics, in addition to the composition of cyanobacterial populations, will further support pre-emptive and active decision-making in regards to water restoration strategies (Dokulil and Teubner 2000).

This chapter's principal aim is to extend upon current knowledge regarding the molecular dynamics within a bloom of cylindrospermopsin-producing *C. raciborskii*. This is achieved through the application of a duplex qPCR established by Rasmussen *et al.* (2008), targeting cylindrospermopsin-specific *pks* and *C. raciborskii*-specific *rpo*C1 gene sequences. These genes are known to exist as single copies within a genome; therefore any departure from a 1:1 ratio between the two along with comparisons to observed *C. raciborskii* cell numbers, permits the construction of a bloom profile, with respect to the proportion of CYN-producing *C. raciborskii* over its time course and the homogeneity of the bloom, are all CYN-producers *C raciborskii*?

In addition to the duplex PCR, a brief review of the current environmental conditions of the River Murray was undertaken to further contribute to the understanding of the unusual dominance of *C. raciborskii* within the river system. Records dating back ten years, do not record significant cell counts (> 500 cells/mL) of *C. raciborskii* along this stretch of the river, with exception of a small bloom of 5000 cells/mL that occurred over a one week period in late summer 2007. The cyanobacterial biodiversity of the bloom was investigated using denaturing gradient gel electrophoresis analysis of a cyanobacterial-specific 16S rRNA gene sequence.

5.2 Specific Materials and Methods

5.2.1 Location and sampling details

Weekly water samples, collected by the SA Water Corporation for monitoring purposes, were taken from numerous sites, along a 500 km section of the River Murray, within South Australia (SA). Sub samples, from those collected for routine monitoring of surface waters, were decanted and analysed within this work. The primary focus for work undertaken in this study was from the water samples and data collected at the sampling site adjacent to the township of Swan Reach which was the origin of the primary bloom (Figure 5.1).

Cylindrospermopsis raciborskii and *Pseudanabaena* sp. cell counts in addition to water chemistry parameters recorded for the River Murray from surface waters, were obtained through the courtesy of Dr Jeremy Lucas (Water Quality and Environment Group of SA Water Corporation). Information regarding flow rates within the River Murray was obtained from records publically available from the Murray Darling Basin Commission website: http://www.mdbc.gov.au/rmw/river_information_centre

5.2.2 Isolation of *C. raciborskii*. CYP 043

It was desirable to obtain a mono-cyanobacterial culture of *C. raciborskii* from bloom material for research purposes. A water sample collected from the River Murray on the 11th of March 2008 at Swan Reach was placed in a 50 cm^2 Greiner Bio-one Tissue Culture Flask with ventilation (Frickenhausen, Germany), containing 30 mL of ASM-1 medium and incubated at 25 °C under photon irradiance of 20 $\mu Mm^{-1}s^{-1}$ provided by cool white fluorescence bulbs (12 hr light: dark cycle). Periodically, the flask was examined over a two week period under an Olympus CK2 dissecting microscope until sufficient cell densities were observed. Individual filaments of *Cylindrospemopsis raciborskii* were then isolated using a hand drawn micropipette, washed three times in ASM-1 and placed in individual wells within a 24 well culture plate (Sarstedt®, Numbrecht-Rommelsdorf, Germany). Following the dense growth of mono-cyanobacterial *C. raciborskii* colonies, the culture was transferred to and maintained in 50 cm^2 Greiner Bio-one Tissue Culture Flask with ventilation, containing 30 mL of ASM-1 medium. The culture was incubated upright at 20 °C under photon irradiance



Figure 5.1 Location of River Murray sampling sites used in this study. Lock 9 (not shown) is approximately 450 km further upstream from Morgan.

of 20 $\mu Mm^{-1}s^{-1}$ provided by cool white fluorescence bulbs (12 hr light: dark cycle). The culture was registered in the AWQC culture collection as CYP 043.

5.2.3 Toxin Analysis

Analysis and quantification of CYN content for samples used within this work, were determined using methods described in General Material and Methods, Chapter 2, section 2.3.2.

5.2.4 Quantitative real time PCR (qPCR)

A duplex, real-time PCR was used to detect and determine the proportion of cylindrospermopsin-specific (*pks*) and *C. raciborskii*-specific (*rpo*C1) DNA sequences within genomic extracts obtained from River Murray environmental water samples, in accordance to methods established by (Rasmussen et al. 2008b). All reactions were performed on a Corbett Rotor-Gene 6000 HRM (Corbett Research, Sydney Australia). Each 25 µL reaction contained 0.25 µM each of cyl2, cyl4, m4 and k18 primers (Geneworks, Table 5.1), 1 × PCR buffer (Invitrogen), 3 mM of $MgCl_2$, 1 U of platinum *Taq* (Invitrogen), 200 µM of dNTPs, 0.1 µM of 5' FAM-labelled PKS and 0.1 µM of 5' JOE-labelled RPOC1 Taqman probes (Biosearch, Novato, CA, USA, Table 5.1), and 5 µL of genomic DNA that was diluted 1:50 with nuclease free water.

Table 5.1 Oligonucleotide primers and fluorescent probes used in real-time qPCR

Primer	Sequence (5'-3')	Gene target	Product (bp)	Reference
cyll2	GGCATTCCTAGTTATATTGCCATACTA	*C. raciborskii* *rpo*C1	308	Wilson et al. 2000
cyll4	GCCCGTTTTTGTCCCTTTGCTGC			
RPOC1 –JOE Probe	ROX TCCTGGTAATGCTGACACACTCG BHQ2	*C. raciborskii* *rpo*C1		Rasmussen et al. 2008b
k18	CCTCGCACATAGCCATTTGC	Cylindrospermopsin *cynB, pks, ketosynthase*	422	(Fergusson and Saint 2003)
m4	GAAGCTCTGGAATCCGGTAA			Schembri et al. 2001
PKS-FAM probe	FAM CGGCAGCAACACTCACATCAGT BHQ1	Cylindrospermopsin *cynB, pks, ketosynthase*		Rasmussen et al. 2008b

5.2.4.1 Generation of a DNA standard curve for *pks* and *rpo*C1 genes

External standards for quantification of *rpo*C1 and *pks* gene copy numbers were prepared using genomic DNA obtained from *C. raciborskii* CYP043, isolated from Swan Reach environmental waters during this study. The concentrations of the *rpo*C1 and *pks* amplicons were determined using the Agilent bioanalyser, post purification (Chapter 2 section 2.5.6).

The concentration of *pks* and *rpo*C1 purified amplicons were 77.37 and 98.25 ng/μL respectively, which is equivalent to 1.88×10^{11} and 2.36×10^{11} copies / μL. A standard curve was prepared using 10-fold serial dilutions (in nuclease free water) to cover the range of 1.88×10^{2} to 1.88×10^{9} copies/reaction of *pks* and 2.36×10^{2} to 2.36×10^{9} copies/reaction of *rpo*C1 and run in parallel with samples. All reactions were performed in triplicate. Test samples were quantified against the generated standard curve.

5.2.4.2 Incorporation of spike control to test for PCR inhibition

Inhibitory elements, carried through the genomic DNA extraction process from environmental samples, can impede the amplification efficiency of PCR reactions. Therefore, it is desirable to include a spiked control that is run in parallel to determine if inhibitory compounds are present and therefore affect accurate quantification (Saunders 2004).

In a separate reaction to those described in section 5.2.3.1, PCR was performed for inhibition by determining the effect of the template DNA preparation in a *pks* spiked mastermix. A similar mastermix was prepared as above, except that cyl12, cyl14 primers, and RPOC1 JOE probe were omitted, and purified *pks* template was added to give an approximate concentration of 10^{3} copies per reaction. The Cycle threshold (Ct) obtained from the reaction with *pks* template DNA added was compared to Ct obtained for the reaction without any included template DNA. Any significant difference in Ct value would be attributed to PCR inhibition.

5.2.5 Universal cyanobacteria hemi-mested 16S *r*RNA gene-directed PCR

PCR reactions were performed utilising cyanobacterial-specific 16S rRNA gene universal primer set 338F-GC/ 780R(a)/ 780R(b) as modified by (McGregor and Rasmussen 2008)

from primers originally described in (Nübel et al. 1997). The 40 bp GC clamp attached to the forward primer is underlined. An amplification product of 422 bp was expected.

338F-GC, 5'- <u>CGC CCG CCG CGC CCC GCG CCC GTC CCG CCG CCC CCG CCC</u> TCC YAC CGG GAG GCA GCA G-3' (Y = C or T)

780R (a), 5'-GAC TAC TGG GGT ATC TAA TCC CAT T-3'

780R (b), 5'-GAC TAC AGG GGT ATC TAA TCC CTT T-3'

PCR amplifications were performed as 100 µL reactions containing 0.5 µM of the forward primer and a 0.5 µM equimolar mixture of the reverse primers (Geneworks, Adelaide), 1 × PCR buffer, 2.5 mM $MgCl_2$, 2 U of Platinum *taq,* 200 µM of deoxynucleoside triphosphate, 200 ng/µL BSA and approximately 5-15 ng of genomic template. A touchdown protocol was implemented to minimise non-specific amplification. Thermal cycling conditions for the PCR were 94 °C for 5 min; then 11 cycles of 94 °C for 1 min, 65 °C for 1 min and 72 °C for 1 min, where the annealing temperature was decreased by 0.5 °C each cycle; followed by a further 32 cycles of 94 °C for 1 min, 55 °C for 1 min and 72 °C for 1 min and a final extension of 72 °C for min, 4 °C hold. All reactions were performed on a Perkin Elmer GeneAmp 2400 PCR thermal cycler (Perkin Elmer Life and Analytical Sciences Inc., Boston, MA, USA). Products were visualised on 1.0 % agarose gel (Chapter 2 section 2.5.5) prior to loading onto polyacrylamide gels in preparation for Denaturing Gradient Gel Electrophoresis (DGGE).

5.2.6 Denaturing gradient gel electrophoresis

Denaturing Gradient Gel Electrophoresis (DGGE) was performed using the Bio-Rad D-GENE Denaturing Gel Electrophoresis System (Bio-Rad, Hercules, CA, USA). Cyanobacterial specific 16S rRNA gene PCR fragments, generated from the universal PCR reaction described in chapter 5 section 5.2.4, were concentrated to 20 µL using the Montage PCR filter columns (Millipore Corporation, USA), combined with 15 µL of 6 x DGGE loading buffer (2.0 g ficoll, 200 µL of 250 mM EDTA (pH 8.0), 0.5 mL of bromophenol blue, MQ H_2O to 10 mL) and loaded onto 8 % (w/v) polyacrylamide gel in 1 × TAE buffer (4.84 g/L Tris Base, 1.1 mL glacial acetic acid, 2 mL of 0.5 M EDTA (pH 8.0)). Polyacrylamide gels were made with a denaturing gradient ranging from 40 % to 70 % (100

% denaturant contains 7 M urea and 40 % formamide). Electrophoresis was performed at 56 °C with constant voltage of 60 V over 16 hrs. Upon completion of electrophoresis, gels were post stained with 1 × SYBR Green I solution (Molecular Probes, Invitrogen Corporation, Carlsbad, CA, USA) in 1 × TAE for 2 hr. The result was visualised on a DR-45M Dark Reader transilluminator (Clare chemical Research Inc., Dolores, Co, USA) and image captured using a Kodak DC290 digital camera fitted with a 49 mm lens. Documentation and further image analysis was performed using Kodak ID version 3.5 image analysis Software (Eastman Kodak Company, Rochester, NY, USA). Bands of interest for cloning and sequencing purposes where excised and DNA eluted in TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA) by incubation at 100 °C for 10 min.

To confirm that cloning reactions successfully incorporated DNA from the band of interest, all cloning reactions, were reamplified using the original PCR primers described in section 5.2.4. The amplification products were then applied to DGGE and their relative mobility was assessed alongside the corresponding environment sample profile.

5.3 Results and Discussion

5.3.1 *C. raciborskii* isolate, CYP043

A *C. raciborskii* strain from the Swan Reach bloom was successfully isolated during the course of this work. The strain was listed in the AWQC collection as CYP043, and taxonomically classified by both morphological characteristics and 16S rRNA gene sequencing. CYP043 was revealed to be a CYN producer by both molecular probes developed through the work discussed in chapter 3, and through physico-chemical methods by two independent laboratories. Interestingly, the isolate was of coiled morphology (Figure5.2), which redefines the observations of (Wiedner et al. 2007), who reported that only straight filament phenotypes of *C. raciborskii* were observed in temperate regions.



Figure 5.2 River Murray isolate, *Cylindrospermopsis raciborskii*, CYN43.

Photomicrograph courtesy of Peter Hobson.

5.3.2 Quantitative real time PCR

Quantitative real-time PCR was performed on 12 environmental samples collected from the River Murray, at the Swan Reach sampling point, over the time course of the *C. raciborskii* bloom from January to April 2008. The limit of detection for this assay had previously been found to be 1000 cells/mL (Rasmussen et al. 2008b). Extracted genomic material, from samples that demonstrated a *C. raciborskii* cell count of > 1000 cells/mL, were analysed by the duplex qPCR assay developed by Rasmussen et al (2008). To permit quantification of the assay it was necessary to generate a standard curve.

5.3.2.1 Standard curve

A dilution series of *pks* and *rpo*C1 gene targets was prepared, ranging from 10^2 to 10^9 copies/reaction, and analysed in triplicate as described in further detail in section 5.2.3. The resultant standard curves are presented below in Figure 5.3 and Figure 5.4. Linear regression over the range 10^2 to 10^9, presented strong statistical support for both *pks* and *rpo*C1 gene targets, with calculated R^2 values of 0.995, 0.999, respectively.

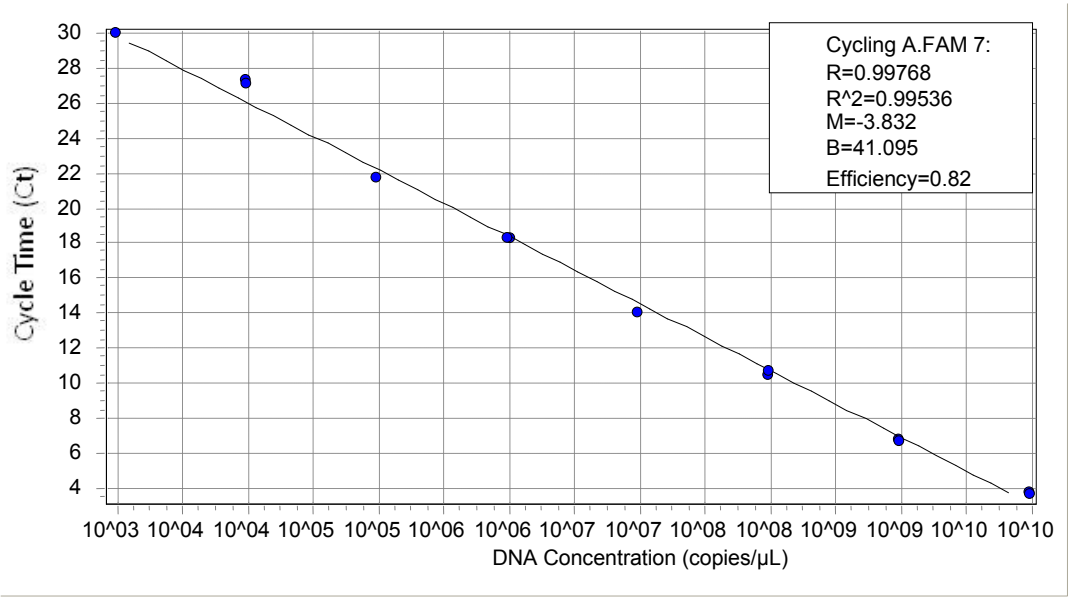


Figure 5.3 Quantitative real-time standard curve for quantification of *pks* gene target copy numbers.

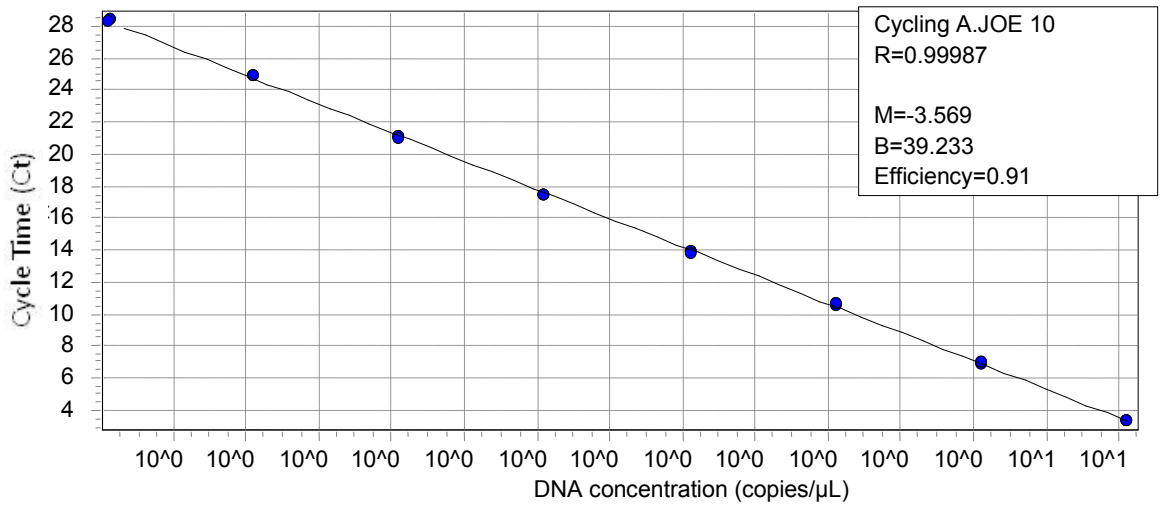


Figure 5.4 Quantitative real-time PCR standard curve for quantification of *rpo*C1 gene target copy numbers

5.3.2.1.1 Comparisons between *pks* and *rpo*C1 gene copies during the *C. raciborskii* bloom.

The quantification and correlation between cylindrospermopsin (*pks*) and *Cylindrospemopsis* (*rpo*C1) specific genes were obtained through qPCR and related to their respective *C. raciborskii* cell counts. The results are represented in Table 5.2 and as Figure 5.5. Out of the 12 samples analysed, 6 successfully amplified both gene targets, whilst one only amplified the *rpo*C1 gene target. The other remaining 5 samples, did not successfully amplify either target. Environmental samples collected at Swan Reach on the 07th, 12th of Feb and 25th of

March 2008, did not amplify either *rpo*C1 or *pks* genomic targets, despite being all over the limit of detection for the assay (1000 cells/mL). A similar result was also evident for previous dilution reactions performed to optimise the assay, for these same time points (data not shown). The reasons for this are unclear; as there was no indication of inhibition in the spiked reactions as evident by the cycle threshold (Ct) being closely comparable in both spiked and unspiked reactions (data not shown). However, the inherent issues involved with extracting genomic material from environmental samples, and the long-term storage, may have contributed to a loss of signature for these samples.

Table 5.2 Quantitative real-time PCR data for the duplex CYN and *C. raciborskii-specific* assay (after correction for 50-fold concentration).

Sampling Date	Duplex *Taq*man probe qPCR (copies/mL)*		*C. raciborskii* cell density (cells/mL)
	pks	***C.raciborskii rpoC1***	
06/01/08	2610	13000	6000
15/01/08	59600	80700	45100
23/01/08	87700	185000	43500
30/01/08	5730	17300	9050
07/02/08	n/d	n/d	27800
12/02/08	n/d	n/d	12100
20/02/08	1320	14100	1440
26/02/08	n/d	n/d	2510
05/03/08	6.14	1070	2410
11/03/08	n/d	1030	3740
25/03/08	n/d	n/d	1340
24/04/08	n/d	n/d	1330

*Copies /mL represented as an average of triplicate reactions
n/d not detected in qPCR

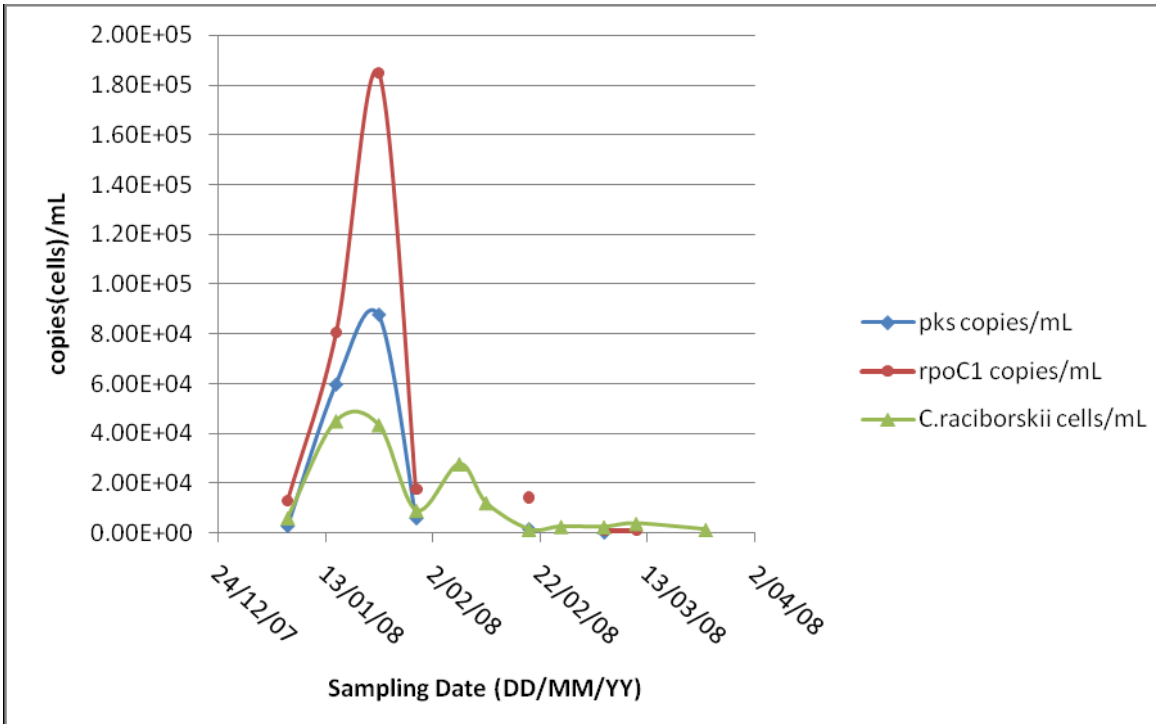


Figure 5.5 Graphical representation of qPCR for the cylindrospermopsin *pks* and *C. raciborskii*-specific *rpo*C1 genes and corresponding cell counts of *C. raciborskii*

Estimates of the cylindrospermopsin, *pks* and *Cylindrospermopsis*-specific, *rpo*C1 gene copies showed trends that reflected those of the cell counts (Figure 5.5). The gene targets used within the assay were selected based on evidence that they exist within the genome as single copies (Schembri et al. 2001; Rasmussen et al. 2008b). However, comparisons between the *pks* gene copy counts and cell counts showed little correlation on a 1:1 ratio; this is also predominantly evident for comparisons between the *rpo*C1 copies and cell counts and also the relationship between *pks* and *rpo*C1 gene copies (Table 5.2 and Figure5.5).

The number of *pks* gene copy numbers ranged from 2 to 392 fold lower than cell counts. Only one sampling timepoint was taken at the bloom maxima where a 2 fold increase in *pks* copies to the corresponding cell count was revealed. Conversely, *rpo*C1 gene copies were either 2 to 9 fold higher than cell counts from corresponding samples. Comparisons between *pks* and *rpo*C1 gene copies, revealed the abundance of *pks* varied between 2 to 167 fold lower than *rpo*C1 copy counts from equivalent samples. There are several plausible explanations for these observations. Disparity between gene targets and cell numbers are potentially introduced through errors in microscopy counts, due to the intrinsic difficulties distinguishing between cells within filaments and variable filament sizes. Cyanobacterial blooms are

heterogeneous; thereby not all strains of *C. raciborskii* within a natural population will necessarily be capable of producing CYN (Baker and Humpage 1994; Chorus and Bartram 1999). Moreover, blooms may also be an amalgamation of more than one cyanobacterial species (Chorus and Bartram 1999; Costa et al. 2006). It was therefore plausible that the presence of cylindrospermopsin *pks* gene within the bloom was not necessarily reflected as an equal quotient to that of the *C. raciborskii*–specific *rpo*C1 gene copy or *C. raciborskii* cell counts. An additional consideration is that *C. raciborskii* may have not been the only CYN producer within the bloom. However, if this latter scenario was evident, the *pks* gene copy estimations would have been expected to be higher than that of both *rpo*C1 gene copy and *C. raciborskii* cell counts. This was unlikely here, as *pks* gene copies were in lower abundance for the majority of the bloom. Conventional microscopy also supported this premise, as identification of alternate CYN producing species was not reported.

It appeared likely that the composition of the bloom was never more than 50% of the toxic genotype, as indicated by lower *pks* gene copy to cell count and higher *rpo*C1 gene to *pks* gene copies. In fact, by the end of the bloom period in late February, *pks* was not efficiently amplified, suggesting that the majority of the bloom consisted of the non-toxic genotype. The lack of *pks* gene amplification could be a result of genotypic shifts resulting in the *pks* target being below the limit of detection for this assay. Two dates, the 5th and 11th of March, from the latter stages of the bloom, the *rpo*C1 gene was detected, indicating that *C. raciborskii* genomic material was evident. Therefore, an underlying assumption could be the population succession between toxic and non toxic genotypes throughout the time course of the *C. raciborskii* bloom.

These findings are in contrast to those found by Rasmussen et al (2008b) when developing the CYN duplex assay during field based trials. Interestingly, in that study, the *C. raciborskii rpo*C1 and CYN *pks* target copies were reported as almost equivalent on all occasions from the same sample extracts at various sites. The focus of the Rasmussen work, however, was to demonstrate the applicability of the duplex assay with respect to detection and quantification of CYN *pks* and *C. raciborskii*-specific *rpo*C1 gene targets. For this reason, time course experiments were not a principal objective of that study and therefore not conducted. This however, does not sufficiently explain how the 1:1 relationship between the *pks* and *rpo*C1 gene targets was consistent throughout the trials at different locations. The sampling sites included both a sub-tropical (North Pine Dam, QLD, Australia) and semi-arid (Imperial

Lakes, NSW, Australia) location. The trials were conducted at locations where persistent *C. raciborskii* populations were known to exist. McGregor and Fabbro (2000) noted that CYN levels were always higher in water bodies dominated by long-term *C. raciborskii* populations, which promotes the possibility that genetic and environmental factors can affect a population's genotypic structure.

Conversely, the discrepancy between toxin and genus specific gene targets is not an uncommon observation and has been consistently recognised in respect to bloom-forming microcystin and non-microcystin producing *Microcystis* populations (Vaitomaa et al. 2003; Kardinaal et al. 2007; Welker 2007; Briand et al. 2009), and *Planktothrix agardhii* (Briand et al. 2008). The shifts in population dynamics between toxic and non-toxic *Microcystis* have been attributed to the overall fitness of non-toxic genotypes (Kardinaal *et al.* 2007; Welker 2007; Briand *et al.* 2009). This tends to correspond to bloom maxima, when competition for nutrient and light requirements has increased. The genotypic shifts in population are additionally reflected in the microcystin concentrations which are at high levels at the commencement of the bloom and low at the bloom maximum.

Cylindrospermopsin concentrations were available for the 6th of January sample (6000 cells/mL), corresponding to a period of exponential growth, the 15th of January (45100 cells/mL), which denotes the bloom maximum or stationary phase and a third, collected toward the end of the bloom, on the 24th of April (1300 cells/mL). The CYN (and deoxy-cylindrospermopsin) content were found to be 0.4 (N/A), 0.4 (0.5) and 0.1 (0.1) µg/L respectively, by HPLC (LOD = 0.2 µg/L).

Interestingly, there was no discernible difference between the toxin levels recorded for the bloom at exponential growth and at critical mass. This could be indicative of total CYN content, similarly reaching peak levels prior to the bloom's maximum, as often noted with microcystins as a consequence of genotypic shifts within the population (Vaitomaa *et al.* 2003; Kurmayer *et al.* 2004; Briand *et al.* 2008; Briand *et al.* 2009). The CYN level detected by the end of the bloom, on the 24th of April is negligible, as it was below the limit of HPLC detection, despite cell counts indicating there was at least 1330 cells/mL of *C. raciborskii* present.

Batch culture experiments have previously suggested that intracellular CYN is highest at exponential growth phase contributing to 85% of total CYN (Hawkins *et al.* 2001). Conversely, the same batch experiments found extracellular CYN concentrations are highest at the stationary phase of growth and can contribute up to 50% of total CYN present (Hawkins *et al.* 2001). These findings were supported by earlier field observations by McGregor and Fabbro (2000) and (Chiswell et al. 1999) that intracellular CYN concentrations were highest, when blooms were young, whilst up to 95% of total toxin could exist in the water column as the bloom declined. Therefore, it is likely the result from the 6th of January is reasonably representative of total toxin for that period. However, the fate of extracellular CYN, through biological and hydrological elements is unknown. CYN has demonstrated a relative short lifetime of 2 to 3 days, in natural surface waters, in the presence of sunlight and plant pigments (Chiswell *et al.* 1999). Moreover, recent findings suggest that CYN- biodegrading microorganisms exist in SA water bodies with no historical records of CYN (Smith *et al.* 2008). Therefore it is possible the result of 15th of January and 24th of April are underestimations of the actual total concentration.

The observation that the *rpo*C1 copy counts were considerably higher than given *C. raciborskii* cell counts was consistent with previous observations by Rasmussen *et al*. (2008b), and several previous studies investigating microcystin-producing cyanobacterial blooms (Vaitomaa *et al.* 2003; Rinta-Kanto *et al.* 2005). Possible reasons to explain this observation ranged from extraneous DNA persisting in the surrounding surface water as a result of lysed cells (Vaitomaa et al. 2003; Rinta-Kanto et al. 2005), an underestimation of the actual cell densities through microscopy (Rasmussen et al. 2008b), or evidence of several genomic copies within a cell (Vaitomaa et al. 2003; Rasmussen et al. 2008b). Genome content in cyanobacteria has been most extensively investigated in *Synechococcus* (Mann and Carr 1974; Asato 1979; Binder and Chisholm 1995). Reports of 2 to 6 genome copies have been documented (Binder and Chisholm 1995), and it has also been predicted that genome content can increase by 16 fold during periods of exponential growth (Mann and Carr 1974). (Kurmayer and Kutzenberger 2003) note that the most pronounced overestimation of gene targets when compared to cell counts of *Microcystis,* occurred when the populations were in transition between log and stationary phase.

The majority of estimates for the *rpo*C1 gene copy number did not exceed cell counts by more than two-fold during periods of exponential growth and four-fold for the transition to

stationary phase. These all seem reasonable estimates and support the presence of multiple chromosomal copies per cell at periods of sustained growth. This is also consistent with reports by Kurmayer and Kutzenberger (2003). The highest discrepancy between *rpo*C1 gene copies and cell counts was 9:1 observed on the 20th of February. Cell counts for this sample were 1440 cells/mL, which was near the lower LOD for the assay. Alternatively, abundant naked DNA potentially may have served as extraneous templates within the *rpo*C1 assay as this sampling date succeeded the bloom maxima.

Additional experimental considerations that may have impacted on detection of desired gene targets include the assumption that the extraction of the genomic DNA was absolute, and secondly there may have been as of yet unidentified cyanobacteria that exhibit homologous *rpo*C1 gene targets and potentially contributed to gene copy numbers. However this latter point was not evident in field trials performed by Rasmussen *et al*. (2008).

Real-time observations for the bloom examined here, suggest that environmental and competitive factors may also influence the dynamics of *C. raciborskii* blooms in respect to genotypic population structure. However, parallels to the microcystin conclusions cannot be entirely drawn for *C. raciborskii* blooms at this stage. Unfortunately, the genomic material available for this study commenced at time points that already indicated the bloom had entered the exponential phase. As a result, the naissance population structure was unknown. Toxin data was also unavailable for the full course of the bloom to help fully support the genotypic structure and dynamics over the bloom's lifetime. Any additional study undertaken in the dynamics of CYN producing *C. raciborskii* populations would greatly benefit by analysis of total toxin for the entire course of the bloom. However, work presented here indicates further investigation is warranted using qPCR and would benefit with comparison between blooms occurring in both tropical and temperate regions.

Studies comparing both temperate and tropical regions are highly desirable. Factors influencing the dynamics within, and between, groups of cyanobacteria are complex and often discerned factors that do not translate as an absolute requirement when reviewed between different environments (Dokulil and Teubner 2000; Wiedner et al. 2007). This CYN-producing, *C. raciborskii* bloom reported here was unique for the River Murray, as never before had a bloom with such high cell counts been witnessed. This was despite the observed presence of *C. raciborskii* in the River Murray since the 1980s (Baker 1996).

The River Murray is an important water resource for Southern Australia. In addition to agricultural and recreation demands placed on the river, it also supplies the state of South Australia with the majority of its raw drinking water. A shift in cyanobacterial communities, in particular the toxigenic species, has consequences in regard to public health and water management. With this in mind, a review of the environmental conditions during this novel bloom was undertaken.

5.3.3 Environmental conditions within the River Murray

Historically, the lower reaches of the River Murray are typically characterised by almost predictable blooms of *Anabaena circinalis*, *Microcystis aeruginosa* and *Nodularia spumigena* (Baker and Humpage 1994; Codd et al. 1994). However throughout the 2007/2008 summer period *C. raciborskii,* normally detected in tropical and sub-tropical regions of Australia, was shown to proliferate in the temperate River Murray environment. Review of historical records, made available by the SAWater water quality group,of the key parameters recorded to monitor the River Murray's water quality, permits speculation as to why *C. raciborskii* was able to dominate.

Cylindrospermopsis raciborskii is regarded as an invasive cyanobacterium (Padisák 1997). There have been several physical and chemical parameters identified that appear favourable for *C. raciborskii* in both tropical (Fabbro and Duivenvoorden 1996; Bouvy et al. 1999; McGregor and Fabbro 2000) and temperate (Padisák 1997; Hamilton et al. 2005; Wiedner et al. 2007) water bodies. These include:

- stratified water column;
- low turbidity;
- high temperature ≥ 25 ºC, at least 22 ºC-23 ºC for akinete germination;
- high irradiance
- high pH.

The River Murray is regarded as an eutrophic and turbid river. However, due to persistent drought conditions, the river system is currently experiencing some of the lowest flows ever recorded. Records obtained from the Murray Darling Basin Commission (MDBC) website indicate that that the median average flow into South Australia by mid December 2000 was

63400 ML/day. This had drastically fallen to 3500 ML/day by December 2007. As a result, stratification of the water column was persistent for up to two weeks (R. Daly personal communication). Also reflective of the slow flows were low turbidity measurements, reported as Nephelometric Turbidity Units (NTU), of 9.9 to 15 NTU over the time course of the *C. raciborskii* bloom. SA Water Corporation archives revealed turbidity measurements of around 40 to 100 NTU were frequently recorded for the Swan reach sampling point prior to the current drought.

It had been previously documented that the growth rate for *C. raciborskii* was greatest at temperatures ≥ 25 ºC (Saker and Griffiths 2000; Briand et al. 2004), often leading to the perennial dominance of *C. raciborskii* in tropical waters (Bouvy et al. 1999). In this study, *C. raciborskii* numbers were first observed in the second week of December 2007, after a month of reported water temperatures of 22 ºC to 23 ºC, which are ideal for akinete germination (Padisák 1997). The seasonal abundance of *C. raciborskii* coincided with the thermal maximum for the summer (January 8th, 25 ºC), a temperature consistent with more temperate regions in Europe (Padisák 1997; Briand et al. 2002; Wiedner et al. 2007) and North America (Hamilton et al. 2005). Interestingly, the thermal maximum for the previous summer (2006-2007) was also 25 ºC, but with no correlated mass occurrence of *C. raciborskii*. Water temperatures for the 2006-2007 summer season were also optimal for akinete germination, preceding the observed thermal maximum.

Cylindrospermopsis raciborskii is often referred to as a shade tolerant species (Padisák 1997; Briand et al. 2002). However, Wiedner et al (2007) demonstrated the importance of photosynthetic active radiation (PAR) in regulating the growth of *C. raciborskii* in German temperate water bodies. Photosynthetic active radiation accounts for global radiation which is often not considered in light intensity calculations. Low turbidity promotes increased PAR in water columns. Observations by Fabbro and Duivenvoorden (1996) found that populations of *C. raciborskii* increased in low turbid water coupled with high global radiation. It is likely that PAR was significant within the River Murray at the time of the bloom as low turbidity, a stable water column and warm temperatures would facilitate the penetration of incoming irradiance.

An ability to out compete other algal species at high pH, due to decreased CO_2 availability, has been reported to benefit *C. raciborskii* (Hamilton et al. 2005). The pH of the river water

was 7.7 to 8.1 prior to, and throughout the duration of the bloom. This is not a notably high pH compared to observations in Fitzroy River, 8.4 to 9.0 (Fabbro and Duivenvoorden 1996) and Lake Constance, Canada, 8.3 to 8.7 (Hamilton et al. 2005). The pH at the time of the bloom was similar to previous pH readings taken from the River Murray.

Phosphorous concentrations in the River Murray appeared to be considerably lower than past years. Phosphorous levels were 0.049 to 0.059 mg/L during the course of the bloom and had previously been as high as 0.111 to 0.396 mg/L. *C. raciborskii* has demonstrated a high affinity and storage capacity for phosphorus (Isvanovics et al. 2000) and therefore is a strong competitor in depleted conditions. Oxidised nitrogen levels were similar to previous years < 0.005 mg/L, ammonium concentrations were not available.

It appeared that several factors may have coincided and resulted in the dominance of *C. raciborskii* in the River Murray during this study. From a historical perspective, the water temperatures experienced in the River Murray's summer season of 2007/2008 were not exceptional. Therefore, temperature was potentially the initiator of growth; however the increased water stability, slow flows and low turbidity perhaps facilitated the light requirements to enhance growth. These parameters and *C. raciborskii* tolerance to low phosphorus levels all appear advantageous and supportive of a bloom event for this organism.

5.3.4 Cyanobacterial community composition using DGGE analysis

Communities are possibly more reliable indicators of water quality, rather than the presence or absence of an individual organism (Reynolds et al. 2002). Understanding cyanobacterial assemblages within a water body may aid in the development of genus-specific approaches to water management. Furthermore, an understanding of the growth responses of cyanobacterial assemblages to specific environmental conditions and nutrient loads could further enhance targeted and pre-emptive water management strategies and predications.

To evaluate the cyanobacterial community structure of the *C. raciborskii* bloom from the lower reaches of the River Murray, a culture independent approach was undertaken. Denaturing gradient gel electrophoresis (DGGE) is a powerful tool that permits discrimination of single nucleotide polymorphisms in a given DNA sequence (Fischer and Lerman 1983). DNA fragments are loaded onto a polyacrylamide gel with an ascending

gradient, achieved through variances in the urea/formamide concentrations, and mobilised utilising a constant charge and temperature. Denaturing gradient gel electrophoresis facilitates a culture independent method to analyse microbial communities and has been successfully applied to investigate cyanobacterial biodiversity (Janse et al. 2003; McGregor and Rasmussen 2008).

Five sampling points were selected from along the River Murray (Lock 9, Morgan, Swan Reach surface, Walkers Flat, Cowirra), for investigation through DGGE (Figure.5.1). The locations are representative of the span of *C. raciborskii* populations, observed within the lower River Murray during this time period. A similar profile was evident in each (Figure 5.6). Overall, 17 bands were excised for molecular characterisation, and 10 were successfully cloned to permit molecular characterisation. The subsequent findings are listed in Table 5.3. Overall, five different species were identified and all but one of these was filamentous and planktonic. Of the 10 bands successfully sequenced, 2 of the bands were identified as belonging to the order Oscillatoriales, 5 belonging to Nostocales and 3 belong to the order Chroococcales.

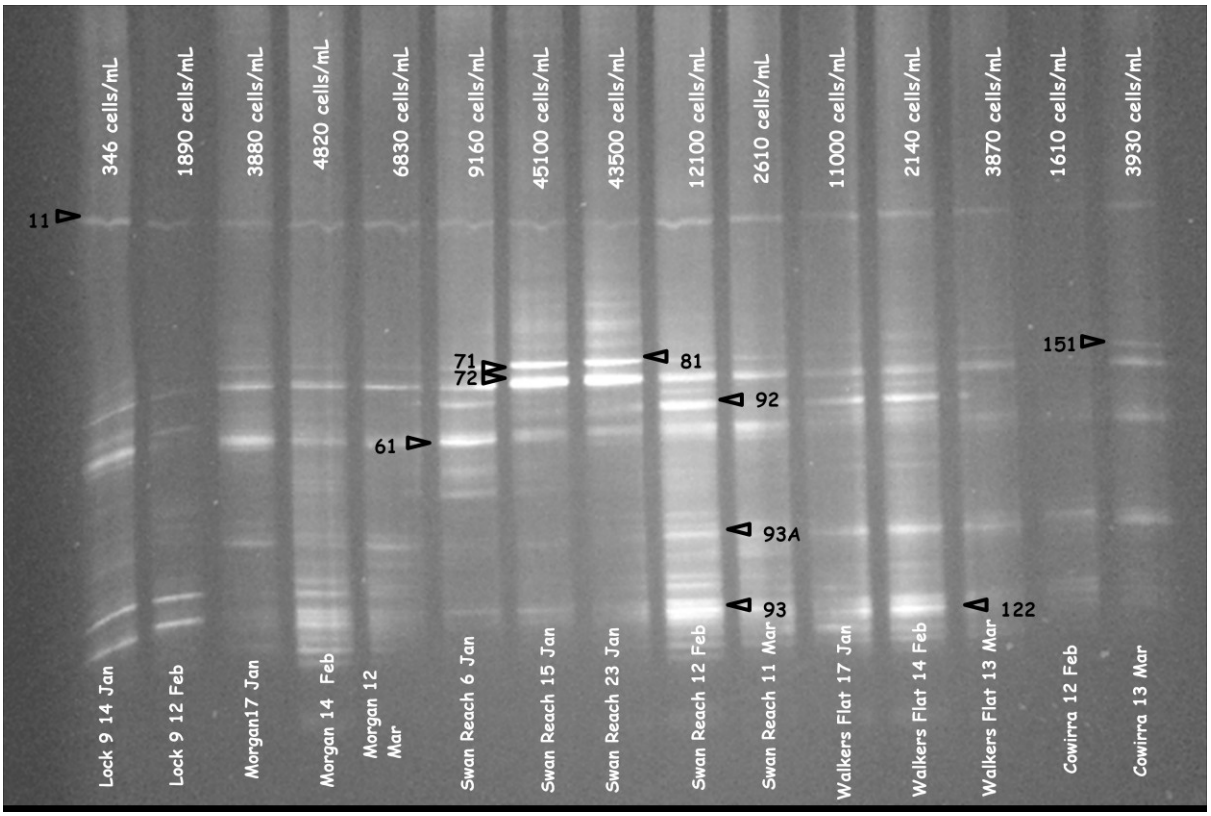


Figure 5.6 DGGE profiles from 5 sampling locations along the River Murray during the summer season 2007/2008.

Table 5.3 The molecular identity of DGGE bands excised from environmental waters of the River Murray

DGGE Band No.	Size (bp)	Match BlastN	BlastN Match Accession number and Geographical origin	BlastN Similarity ID	Match Band in DGGE	Similarity ID to matched band
11	475	*Leptolyngbya sp.*	**EF08333** Brazil	96%	-	-
71	480	*Cylindrospermopsis raciborskii*		100%	81,151	100%
72	480	*Cylindrospermopsis raciborskii*	**AF092504** Sydney, Australia	100%	71,81,151	99.7%
81	480	*Cylindrospermopsis raciborskii*		100%	71,151	100%
151	480	*Cylindrospermopsis raciborskii*		100%	71,81	100%
92	477	*Pseudanabaena sp.*	**AM259268** Lake Tuusuanjarvi, Finland	99%	-	-
61	478	*Anabaenopsis elenkenii*	**AM773309** Kazinga channel, Uganda	99%	-	-
93A	479	*Synechoccocus sp.*	**AF448067** Not specified	98%	93,122	94.3%,98.1%
93	478	*Synechoccocus sp.*	**AB60942** Lake Okutama, Japan	95%	93A,122	As above
122	479	*Synechoccocus sp.*	**AB60942** Lake Okutama, Japan	99%	93,93A	98.1%,95.8%

From the DGGE analysis, one of the identified bands represented the presence of a *Pseudanabaena* sp. The incidence of *Pseudanabaena* sp. has been associated with the production of 2 methyisoborneol (MIB) , a compound that sullies the water with a musty taste and odour (Izaguirre and Taylor 2004). MIB sensitive people are able to detect the compound at low levels (10 ng/L). *Pseudanabaena* sp., has been implicated in the production of MIB within the River Murray and for this reason the species is often routinely monitored for.

The association between *C. raciborskii* and *Pseudanabaena* sp. in cyanobacterial assemblages have previously been noted in tropical surface waters in Australia (Fabbro and Duivenvoorden 1996; McGregor and Fabbro 2000) and temperate waters of Europe (Nixdorf et al. 2003; Wiedner et al. 2007). This association has been primarily attributed to ecological adaptations that the two species are reputed to share, including shade and vertical mixing tolerances (McGregor and Fabbro 2000; Reynolds et al. 2002). As the River Murray is a large contributor to the raw drinking water for South Australia, the relationship between *C. raciborskii* populations and *Pseudanabaena* sp. were examined further on this occasion.

5.3.4.1.1 Relationship between *C. raciborskii* and *Pseudanabaena* sp.

Bio monitoring data demonstrated a positive correlation between the populations of *C. raciborskii* and *Pseudanabaena* sp. throughout the bloom in 2007/2008. *Pseudanabaena* sp. numbers appeared to reflect the population fluctuations of *C. raciborskii* over two successive summer seasons at the Swan Reach sampling point of the River Murray. Conversely, a negative correlation was evident between populations of *C. raciborskii*, *Pseudanabena* sp. with characteristically dominant *A. circinalis*. The population trends of the three species are represented in Figure 5.7.

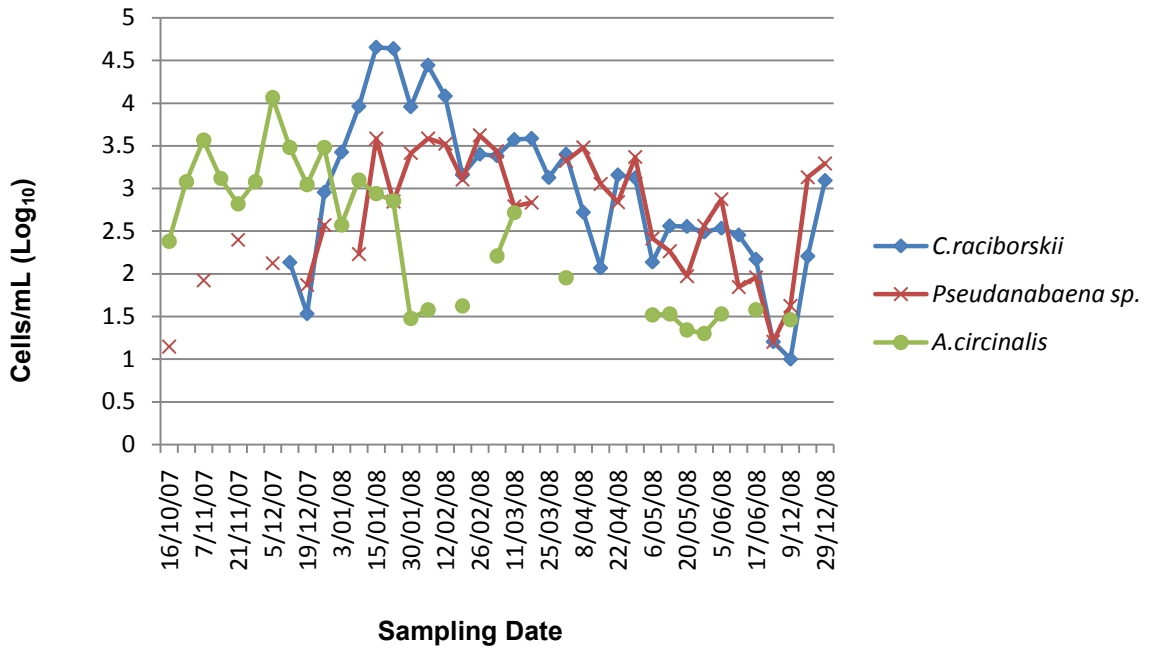


Figure 5.7 Cell numbers for three dominant cyanobacterial species, River Murray, Swan Reach sampling site 2007-2009

This trend between *C. raciborskii*, *Pseudanabaena* sp. and *A. circinalis* abundances was also observed at other locations within the River Murray, including Walkers Flat and Morgan that also experienced an increase in *C. raciborskii* numbers over 2007/2008. These results are represented in Figure 5.8 and Figure 5.9. However, it is less evident at the Morgan sample site which may be reflective of the limited availability of sampling data.

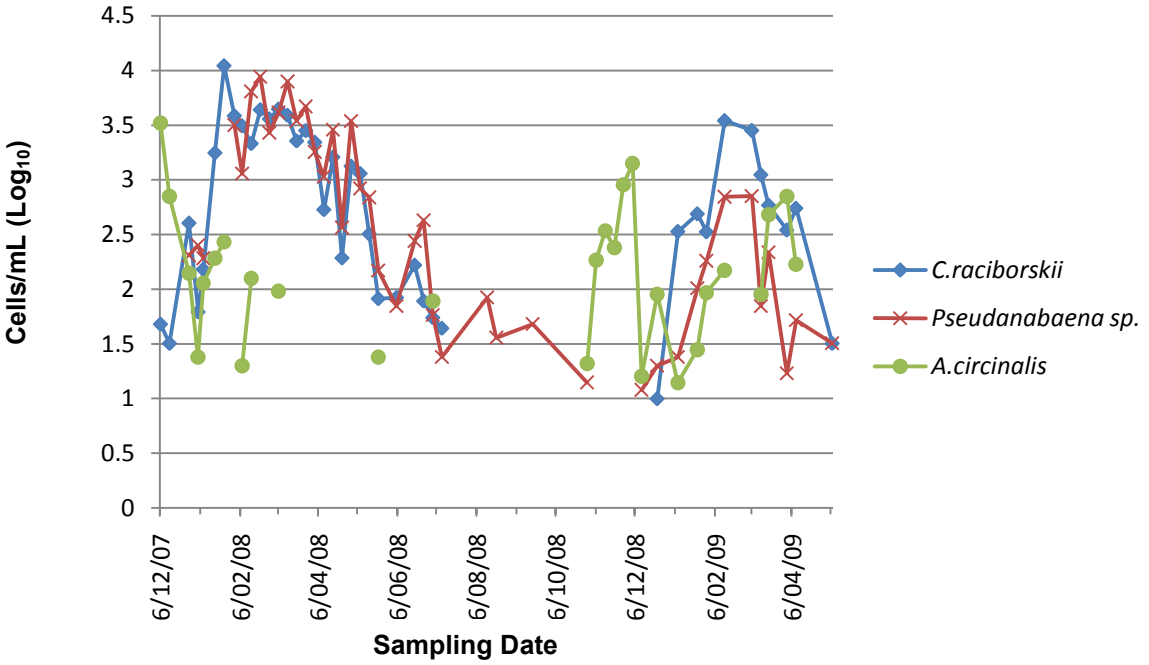


Figure 5.8 **Cell numbers for three dominant cyanobacterial species, River Murray, Walkers Flat sampling site 2007-2009.**

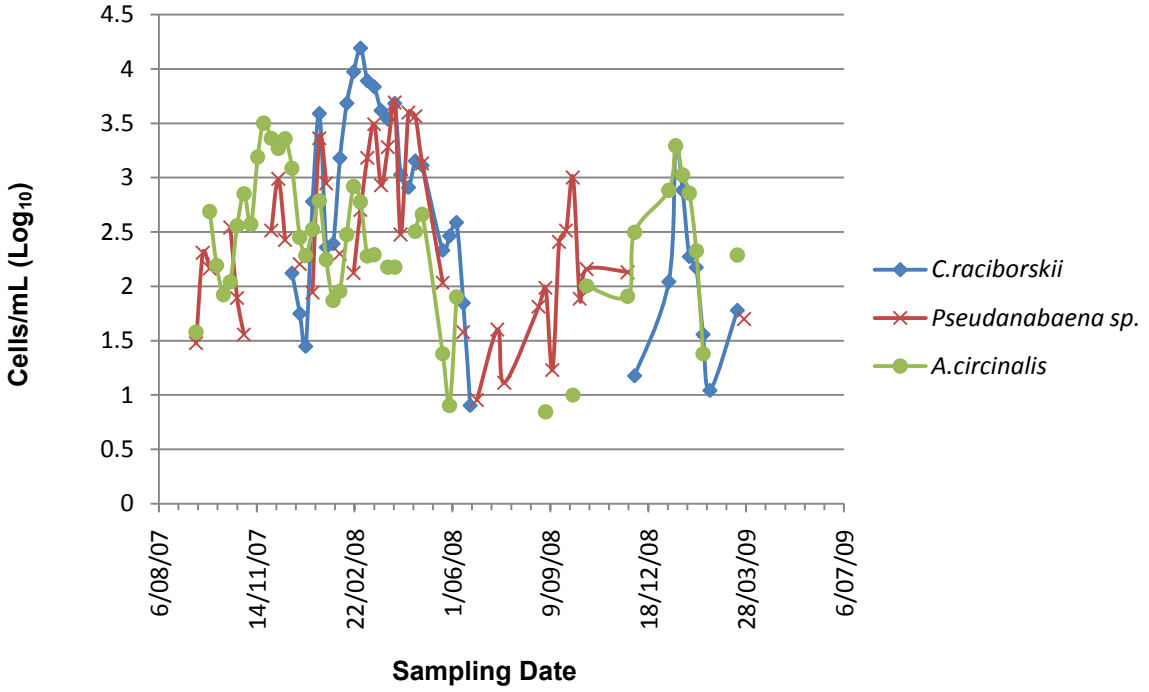


Figure 5.9 **Cell numbers for three dominant cyanobacterial species River Murray, Morgan Sample Pump, sampling site 2007-2009.**

Corresponding taste and odour data for these locations indicated significant levels of MIB at both Swan Reach (9 -141 ng/L) and Walkers Flat (10 - 78 ng/L) during the duration of the *C. raciborskii* bloom. No significant levels were detected at Morgan.

There have been resident populations of *C. raciborskii* and *Pseudanabaena* sp. within the Murray for some time (Baker and Humpage 1994; Baker 1996). However, it appears that environmental conditions within the Murray are now favouring the proliferation of *C. raciborskii.* It is unknown if the association between *C. raciborskii* and *Pseudanabaena* sp. is wholly an indication of similar niche requirements, or whether there is more complex synergistic relationship. What is evident is that the association may have implications in regards to water quality management, on both a public health and aesthetic level.

5.4 Summary

1. This work presented here reports the monitoring of a cylindrospermopsin-producing *C. raciborskii* bloom through quantitative real-time PCR. The application of this assay enabled preliminary insights into the structure of toxic and non-toxic *C. raciborskii* populations over the majority of its time course.
2. There appeared to be a population shift from CYN-producing to non-CYN- producing *C. raciborskii* over the time course of the bloom. This is consistent with observations from microcystin-producing cyanobacterial blooms (Vaitomaa et al. 2003; Briand et al. 2008; Ha et al. 2009). However, there was some discrepancy between these findings and those from preliminary investigations performed on CYN-producing *C. raciborskii* populations (Rasmussen et al. 2008b).
3. Additional research is required to establish why gene copies of *rpo*C1 and cell counts exhibit discrepancies from an expected 1:1 ratio at signifigant timepoints within the bloom. Until a greater understanding of these observations, also apparent in other cyanobacterial populations (Binder and Chisholm 1995, Vaitomaa et al 2003, Rasmussen et al 2008b) have been more fully addressed, qPCR can only be regarded as an interim semi-quantative tool.
4. The work within this chapter reviews the environmental conditions of the River Murray over the 2007/2008 summer season.

5. Multivariate parameters appeared to have favoured the dominance of *C. raciborskii* in the River Murray. These include,
    - persistent stratification of water column
    - optimal temperature for akinete germination
    - low turbidty
    - high irradiance
    - low phosphorous
6. A putative association between two problematic cyanobacteria, *C. raciborskii* and *Pseudanabaena* was observed in two out of three locations investigated within the River Murray. Both these organisms share similar tolerances and if current conditions prevail in the Murray, an increase in water quality issues on both public health and aesthetic levels may result.
7. The findings in this chapter offer water utilities preliminary insights into the population structure of CYN-producing cyanobacterial bloom. Further research to elucidate the breakdown between *rpo*C1 gene copy and cell numbers maybe used in the future to predict imminent blooms. This may guide water management strategies to be targeted for optimal time points to minimise costs involved in treating toxic cyanobacterial blooms. It may also lead to greater understanding of the conditions that favour the non-toxic genotype over toxic genotypes, leading toward the development of proactive and restorative practices to promote "good" cyanobacterial growth.
8. The next chapter explores the history of the CYN biosynthetic cluster in cyanobacterial lineages and the mechanisms that may be involved in disseminating the genes through cyanobacterial genomes.

6 EVOLUTIONARY INVESTIGATIONS INTO THE CYLINDROSPERMOPSIN BIOSYNTHETIC GENE CLUSTER

6.1 Introduction

The previous chapter demonstrated the applicability of using the molecular knowledge acquired in Chapter 3 on a molecular-based assay. The molecular assay permitted tracking of a cylindrospermopsin-producing *C. raciborskii* bloom which may assist in more proactive approaches to bloom management. An additional utility for the acquired molecular data is the exploration of the intrinsic information stored within the nucleotide sequences, through phylogenetic analyses. The work undertaken in this chapter was to reconstruct the evolutionary history of both the cylindrospermopsin biosynthetic cluster and the producing organisms.

Phylogenetic reconstructions of genes involved in secondary metabolism, such as those that promote toxicological responses, is an approach that is often used to investigate the evolutionary history of such genes within prokaryotic genomes. How these secondary metabolite clusters are disseminated through cyanobacterial genomes is of interest and may prove insightful in resolving the possible ecological role that these compounds fulfil, as well as improving our understanding of toxic bloom formation.

Current evolutionary investigations regarding cyanobacterial biosynthetic gene clusters have largely focussed on genes encoding the microcystin synthetase. Horizontal gene transfer (HGT) (Tillett 2000b; Moffitt and Neilan 2004; Tanabe et al. 2004), active intragenic recombination (Tanabe et al. 2004) or the presence of biosynthetic genes in the genome of the last common ancestor of all microcystin-producing species (Rantala et al. 2004) have all been suggested as likely explanations for their presence within cyanobacterial genomes. Horizontal transfer has also been purported as a major factor in the dispersal of the cylindrospermopsin biosynthetic cluster through cyanobacterial lineages (Kellmann et al. 2006; Stucken et al. 2009). These mechanisms have been discussed in greater detail in Chapter 1.

There are various methods available to investigate the probability that a HGT event has occurred within a genome. Observations of incongruence within gene phylogenies, evidence of patchy distribution of homologous genes throughout related genera, species and strains,

and uncharacteristic nucleotide composition are often indicative of HGT events (Ochman et al. 2000). The success of each of the available methodologies varies in their estimates of HGT events; therefore there remains some scepticism in the overall evolutionary impact of HGT (Kunin and Ouzounis 2003).

The evolutionary relationship of an individual gene (or cluster) within a genome is best resolved through comparisons with multiple loci and, in particular, with reference to genes that are considered constrained in their function and therefore highly conserved. Two key molecular markers that are often utilised in cyanobacterial phylogeny to benchmark congruency are the 16S rRNA (Woese 1987; Wilmotte 1994) and *rpoC1* genes (Palenik and Haselkorn 1992). The (16S) rRNAs, in conjunction with ribosomal proteins, comprise ribosomal subunits that are required in the translation of messenger RNA to protein. The DNA-dependant RNA polymerases play an important role in directing gene transcription and exist as a unique γ subunit, *rpo*C1, in cyanobacteria (Schneider et al. 1987; Xie et al. 1989). Both rRNA and *rpo*C1 genes are essential for primary cellular functions and as a consequence they are highly conserved and evolve relatively slowly.

There are many available methods to reconstruct evolutionary relationships between sequences of interest, where distance and discrete (character-based) methods are most frequently selected (Page and Holmes 1998). Distance methods convert sequence data into evolutionary distances through the construction of a pair-wise distance matrix. This information can then be processed to generate a phylogenetic tree through the implementation of algorithms, such as neighbour joining (NJ). While this method is computationally efficient, it is at the expense of losing the evolutionary information through the summation of sequence sets into matrices.

The two most commonly used discrete methods are cladistics and Maximum Likelihood (ML). Discrete methods differ from distance methods by retaining and analysing the information held within (cladistics), or as a function of (likelihood) the molecular data, rather than converting the sequences into a pair-wise matrix. Cladistics, which is based on the concept of cladogensis (that new species arise from branch splitting events in a phylogenetic tree), retains data held at individual nucleotide sites, treating each site as a character and the base at that site as a character state. Through the use of an outgroup, the character states for a given character are categorised as ancestral or derived and the sharing of derived character

states is used to demonstrate common ancestry; the end assumption is that the true phylogeny is reflected in the most parsimonious (shortest, requiring the least number of branching events to generate) tree. Maximum Likelihood applies an assumed model of evolution to calculate the probabilities of all possible trees for a given data set, with the tree with the highest probability being selected as the most likely tree to explain the observed molecular sequence data. Both are computationally expensive and in the case of parsimony analysis heuristic methods are often necessary to analyse large data sets, which may skew the optimal tree outcome (Page and Holmes 1998).

Both Distance and Maximum Likelihood methods are generally only applied to DNA or protein sequence data and they use evolutionary models, each with different assumptions that attempt to account for biases, such as differences in GC content and different rates of evolution at different nucleotide positions, in an effort to reconstruct a valid evolutionary tree.

Phylogenetic trees are representations of the evolutionary relationships estimated by the methods outlined above. Trees consist of internal nodes that represent the last hypothetical common ancestor for a group of taxa represented at the terminal nodes (also termed as leaves or OTUs [Operational Taxonomic Units]). The nodes are connected by edges, or branches, and in the case of distance-based or ML trees the evolutionary distance between two organisms is the summation of their horizontal branch lengths (Page and Holmes 1998). Often, an organism considered to be related but external to the group of organisms being studied (eg a related species if comparing strains of the same species, or a related genus when comparing species from the same genus) is utilised as an out-group to "root" the tree. In the case of cladistics, the outgroup is used to determine the characters states that are ancestral. Figure 6.1 illustrates the terminology used in reference to phylogenetic trees.

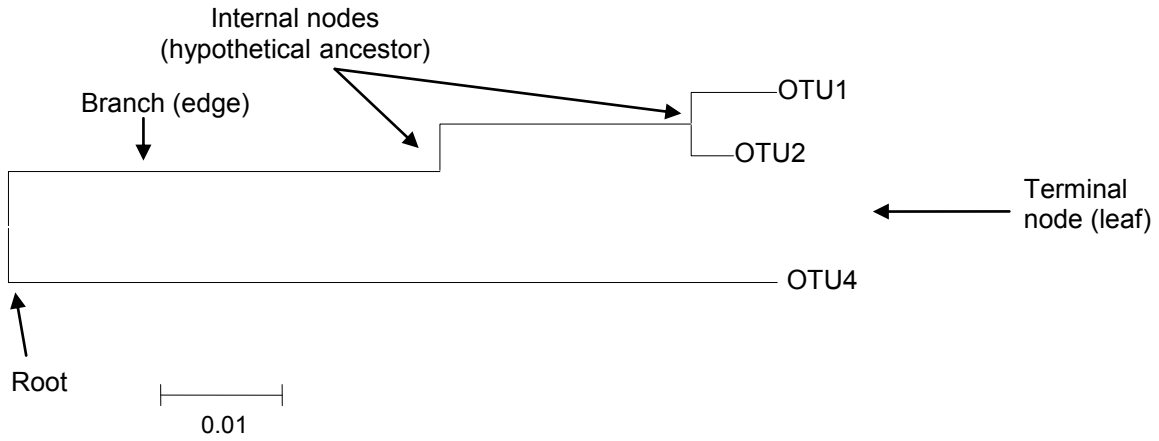


Figure 6.1 A representative tree and associated terms

It is recommended that tree branches be evaluated to assess confidence in the inferred phylogeny. Bootstrap analysis is often employed to evaluate the frequency of a given branching pattern through the random resampling of the dataset (Felsenstein 1988). Generally, only nodes with bootstrap values above 95% are considered strongly supported; however it is possible for an incorrect node within a given tree to yield robust values (and vice versa). Therefore, recovering trees with the same topologies using different methods and multiple loci, in conjunction with pseudo statistical methods, can lend support that the correct phylogeny has been resolved.

This chapter aims to extend upon previous investigations into the possible evolutionary history of the cylindrospermopsin biosynthetic cluster. It attempts to assess the involvement that HGT events may have had in disseminating the biosynthetic cluster between several cyanobacteria genera. It examines 21 taxa, 19 of which were identified as producers of the toxin cylindrospermopsin, including 3 *Aphanizomenon* sp. isolates from Germany (Stüken 2008) and an *Anabaena lapponica* from Finland (Spoof et al. 2006). In addition, an *Aphanizomenon issatchenkoi* isolate from New Zealand was analysed, for which PCR initially amplified CYN genes, prior to their subsequent loss from the genome soon after isolation (Wood et al. 2007b). There was no detection of CYN in any *Aphanizomenon issatchenkoi* extracts analysed by physicochemical methods.

6.2 Specific Materials and Methods

6.2.1 Summary of isolates

All isolates used for phylogenetic analyses performed in Chapter 6 are presented in Table 6.1.

6.2.2 Conventional PCR and sequence analysis

All conventional PCR reactions were prepared as outlined in Chapter 2 (section 2.5.3) and results verified on a 1.0 % agarose gel, as described in Chapter 2 (section 2.5.5) . Products of interest were purified, cloned and sequenced or sequenced directly as detailed in Chapter 2 (sections 2.6.1, 2.6.2 and 2.81 respectively).

6.2.3 Cylindrospermopsin biosynthetic gene-directed PCR

Conventional PCR reactions, using primers directed to genes within the CYN biosynthetic cluster were performed as outlined in Chapter 3 (section 3.2.4) and the applicable primer sets are listed in Table 3.5. Amplification products pertaining to AT domains within the *pks* modules of *cynA*, *cynB*, *cynD* and the amidinotransferase *cynC* were retained for sequencing and phylogenetic analysis.

Table 6.1 Summary of isolates analysed in Chapter 5. Sequences obtained from Genbank database are indicated in bold

Species	Strain	CYN	Geographical origin	Reference	Accession Number					
					16S gene	*rpoc1* gene	*cyn* gene			
							A	B	C	D
Anabaena bergii	ANA283A	+	Glenbawn Dam,NSW	AWQC/ Schembri et al. (2001)	**FJ234897**	**FJ234846**	+	+	+	+
	ANA283B	-	Glenbawn Dam,NSW	AWQC/ Stücken et al..(2009)	**FJ234889**	n/a	-	-	-	-
	ANA360B	-	Tanunda, SA	AWQC/ Stücken et al..(2009)	**FJ234887**	n/s	-	-	-	-
	ANA366A	-	Thornden Pk, SA	AWQC/ Stücken et al..(2009)	**FJ23489**	n/s	-	-	-	-
	ANA366B	+	Thornden Pk, SA	AWQC/ Schembri et al. (2001)	**FJ234890**	**FJ234847**	+	+	+	+
	ZIE26AB	-	Zierker See, Germany	Stücken et al..(2009)	**EF529487**	n/a	n/a	-	-	-
Anabaena flos-aquae	UTEX LB2557		Unknown	**GenBank**	**DQ234825**	-				
	NIES-73		Unknown	**GenBank**	**AB042858**	-				
	LMECYA165		Portugal	Valerio et al. (2009)	**EU078523**					
Anabaena kisseleviana	NIES-74		Unknown	**GenBank**	**AY701557**	-				
Anabaena lapponica	966	+	Lake Takajärvi, Finland	Jarkko Rapala/Lisa Spoof et al. (2000)	This wk	+	-	+	+	-
Anabaena (Aphanizomenon) aphanizomenoides	ANA259C	-	Darling Anabranch, NSW	AWQC	This wk	+	+	-	+	+
	ANA235C	-	Menindee Lakes, SA	AWQC/ Stücken et al..(2009)	**FJ234896**	n/a				

Table 6.1 continued.

Species	Strain	CYN	Geographical origin	Reference	Accession Number					
					16S gene	*rpoc1* gene	*cyn* gene			
							A	B	C	D
	ANA217A	-	Nildottie Lagoon,SA	AWQC/ Stücken et al..(2009)	**FJ234894**	n/a				
Aphanizomenon issatschenkoi	CAWBGO2	-	New Zealand	Wood et al. (2007b)	**DQ380334**	+	+	**DQ645532**	**DQ645534**	**DQ645533**
	LMECYA31		Portugal	Valerio et al. (2009)	**AY196088**	n/a				
Aphanizomenon ovalisporum	ILC-146 (KA1-1*)	+	Lake Kinneret, Israel	Shalev-Alon (2002)/ Stücken et al..(2009)*	**EF529482***	**FJ234859***	-	**AF358958 28**	**AF358958 28**	**AF358958 28**
	FAS-AP1	+	Florida, USA	Yilmaz et al. (2008)	**EU076457**	n/s				
	031E	+	Mildura, Vic	AWQC	+	+	+	+	+	+
	031G	+	Mildura, Vic	AWQC	+	+	+	+	+	+
	033B	+	Thornden Pk,SA	AWQC/ Stücken et al..(2009)	**FJ234895**	**FJ234845**	+	+	+	+
	033C	+	Thornden Pk,SA	AWQC	This wk	This wk	+	+	+	+
	033E	+	Thornden Pk,SA	AWQC	This wk	This wk	+	+	+	+
Aphanizomenon ovalisporum	035C	+	Thornden Pk,SA	AWQC/ Stücken et al..(2009)	**FJ234884**	**FJ234844**	+	+	+	+
	035D	+	Thornden Pk,SA	AWQC/ Stücken et al..(2009)	**FJ234885**	**FJ234843**	+	+	+	+
	035E	+	Thornden Pk,SA	AWQC	This wk	This wk	+	+	+	+
Aphanizomenon sp	10E6	+	Melangsee, Germany	Stücken et al..(2009)*/A. Stuken, PhD thesis (2008)	**EF529475***	**FJ234850***	+	+	+	+

Table 6.1 continued.

Species	Strain	CYN	Geographical origin	Reference	Accession Number					
					16S gene	*rpoc1* gene	*cyn* gene			
							A	B	C	D
	10E9	+	Melangsee, Germany	Stücken et al..(2009)*/ A. Stuken, PhD thesis 2008	**EF529476***	**FJ234855***	+	+	+	+
	22D11	+	Heiliger See, Potsdam, Germany	Stücken et al..(2009)*/ A. Stuken, PhD thesis 2008	**EF529480***	**FJ234852***	+	+	+	+
Cylindrospermopsis raciborskii	AWT205	+	Sydney, NSW	Hawkins et al. (1997)	This wk	This wk	+	+	+	+
	023E	+	Coen Dam, QLD	AWQC	This wk	This wk	+	+	+	+
	023D	+	Coen Dam, QLD	AWQC	This wk	This wk	+	+	+	+
	024C	+	Fitzroy River, QLD	AWQC	This wk	This wk	+	+	+	+
	026J	+	Bourke, NSW	AWQC	This wk	This wk	+	+	+	+
Cylindrospermopsis raciborskii	030A	+	Malparinka Farm,NSW	AWQC	This wk	This wk	+	+	+	+
	032	+	Fitzroy River, QLD	AWQC	This wk	This wk	+	+	+	+
	034	+	Billabong, QLD	AWQC	This wk	This wk	+	+	+	+
	LMECYA 132	-	Portugal	Valerio et al. (2009)	**EU078547**					
	T3	-	Brazil	GENBANK	**EU439566**	n/s				
Fischerella muscicola	PCC7414		Unknown	GENBANK	**AB075986**	**AB074804**				

Table 6.1 continued.

Species	Strain	CYN	Geographical origin	Reference	Accession Number					
					16S gene	*rpoc1* gene	*cyn* gene A	B	C	D
Fischerella sp.	MV9		Costa Rica	GENBANK	**DQ786169**	n/s				
Gloeobacter violaceus	PCC7421		Unknown	GENBANK	**AF132790**	**U52340**				
Microcystis aeruginosa	PCC7806		Unknown	GENBANK	**AF139299**	**AY425000**				
Nodularia spumigena	PCC73104		Unknown	GENBANK	**DQ185241**	n/a				
	BY1		Unknown	GENBANK	**AF268004**	n/s				
Nodularia spumigena	HEM		Unknown	GENBANK	n/d	n/a				
	SAG 36.92		Unknown	GENBANK	**DQ185251**	n/s				
Nostoc sp.	SAG 36.92		Unknown	GENBANK	**DQ185251**	n/s				
	I0-102-I		Unknown	GENBANK	**AY566855**	**n/a**				
Nostochopsis sp.	89-45		Corsica, France	GENBANK (Gugger and Hoffmann 2004)	**AJ544081**	**n/s**				
Planktothrix agardhii	213		Unknown	GENBANK	**EU151902**	**n/a**				
	NIVA-CYA 127		Unknown	GENBANK	**AB045915**	**n/a**				
Raphidiopsis curvata	HB1	+	Wuhan, China	GENBANK, Li et al. (2001)	**AY763116**	**n/s**				
Trichormus variabilis	KCTC AG10269		Unknown	GENBANK	**DQ234833**	**n/s**				

Table 6.1 continued.

Species	Strain	CYN	Geographical origin	Reference	Accession Number					
					16S gene	*rpoc1* gene	*cyn* gene			
							A	B	C	D
	HINDAK 2001/4		Russia	GENBANK	**AJ630456**	**n/s**				
Umezakia natans	TAC101	+	Lake Mikata, Fukui, Japan	Harada et al. (1994)/Kellmann et al. (2006)*	**AY897614 ***	**n/s**			**AY8 976 05**	

6.2.3.1 Cyanobacterial specific 16S rRNA gene PCR

The hypervariable region of the 16S rRNA gene was amplified using oligonucleotide primers, PF1 and PR1 as previously described by Shaw *et. al* 1999. The expected amplification product is approximately 474 bp.

PF1, 5'-TGTAAAACGACGGCCAGTCCAGACTCCTACG-3'
PR1, 5'-CGCGTTAGCTACGGCACGGCTCG-3'

A touchdown protocol was employed to minimise any non-specific amplification. Thermal cycling conditions for the PCR were 94 °C for 10 min then 94 °C for 30 s, 68 °C for 30 s and 72 °C for 2 min where the annealing temperature was decreased by 1 °C each cycle followed by 25 cycles of 94 °C for 30 s, 58 °C for 30 s and 72 °C for 2 min with a final extension of 72 °C for 7 min followed by a 4 °C hold.

6.2.3.2 Cyanobacterial *rpo*C1 PCR

Oligonucleotide primers for the amplification of the *rpo*C1 locus are outlined in Table 6.2. A touchdown thermal cycling protocol was established to minimise non-specific products and comprised of an initial denaturation step of 94 °C for 10 min followed by 6 cycles of 94 °C for 30 s, initial annealing of 55 °C for 30 s and extension of 72 °C for 2 min. The annealing temperature was subsequently decreased by 1 °C over the next 6 cycles with a final 32 cycles performed at 94 °C for 30 s, 50 °C for 30 s and 72 °C for 2 min followed by a final extension of 72 °C for 5 min, 4 °C hold . The anticipated amplification products were expected to be approximately 606 bp.

Table 6.2 Oligonucleotide primers used to amplify the *rpo*C1 gene locus

Primer	Sequence (5'-3')[a]	Description	Reference
rpoC1-1	GAGCTCYAWNACCATCCAYTCNGG	Cyanobacterial specific *rpo*C1 gene forward primer used for *C. raciborskii* strains.	Palenik and Haselkorn 1992
rpoC1-T	GGTACCNAAYGGNSARRTNGTTGG	Cyanobacterial specific *rpo*C1 gene, reverse primer, used for *C. raciborskii* strains.	Palenik and Haselkorn 1992
rpoCAphF	CWGAAACNATTAAYTACCGNAC	Cyanobacterial specific *rpo*C1, forward primer, used to amplify from *Aphanizomenon/Anabaena* sp.	Wood *et. al.* 2007b
rpoCAphR	YAWNACCATCCANTCYGG	Cyanobacterial specific *rpo*C1, reverse primer, used to amplify from *Aphanizomenon/* *Anabaena* sp.	Wood *et. al.* 2007b

[a] B = G/C/T, N = A/G/C/T, R = A/G, S = C/G, W = A/T, Y = C/T

6.2.4 Long template PCR

Long template PCR utilising the Elongase *taq* polymerase system was prepared as previously described in Chapter 2 (section 2.5.4). Primers used are detailed in Chapter 3 (Table 3.1) and involved the following primer combinations: CA1/CD4 with predicted product size of 7.3 Kb in relation to *C. raciborskii* AWT205 sequence and CC2/CD3 as a positive control generating a 5 Kb product. Touchdown thermal cycling protocols were utilised to reduce non-specific amplification and consisted of 94 °C for 30 s; 6 cycles of 94 °C for 30 s, 58 °C for 30 s and 68 °C for 7 min, with the annealing temperature decreasing by 1 °C each cycle; followed by 35 cycles of 94 °C for 30 s, 53 °C for 30 s and 68 °C for 7 min. Products were verified by 1% agarose gel electrophoresis as described in Chapter 2 (section 2.5.5).

6.2.5 Phylogenetic analyses

Phylogenetic trees were recovered from molecular data using three distinct methods: estimation of evolutionary distances followed by Neighbour Joining (NJ) analysis; cladistics by Maximum Parsimony (MP); Maximum likelihood (ML). NJ trees were constructed using distances estimated using the corrected LogDet algorithm (Kumar 2002), which accounts for heterogeneity in base composition that may exist between the taxa (Lockhart et al. 1994). Phylogenetic analyses were conducted with programs and conditions described in Chapter 2 (section 2.8). Primer sequences and ambiguous sequences were removed from the alignments and the input order of all sequences was randomised.

6.2.6 Metrics used to determine Horizontal Gene Transfer events.

Of the methods available to investigate the likelihood of HGT events within a genome, observations of phylogenetic incogruence and anomalous nucleotide compostion were selected to investigate these events within CYN-producing cyanobacterial linegaes.To ascertain whether there was evidence of disparity amongst gene phylogenies between genera and species, several phylogenetic trees were constructed (as detailed in 6.2.5) to permit comparisons between the distribution of multiple loci from the CYN gene cluster and stable reference markers *rpo*C1 and 16S rRNA. Evidence of uncharacteristic nucleotide distribution was determined by evaluating Relative Synonymous Codon Usage (RSCU) between CYN biosynthetic genes and key primary metabolism genes *rpo*C1, *nif*H and the phycocyanin alpha and beta loci, isolated from *Cylindrospermopsis raciborskii* AWT205. Relative Synonymous Codon Usage was calculated within the MEGA 4 program (Tamura et al. 2007).

6.2.7 Other Statistical methods

Selection pressures were inferred by calculating the ratio of nonsynonymous substitutions per nonsynonymous (d_N) site to synonymous substitutions per synonymous site (d_S), based on the modified Nei-Gojobori method within the MEGA 4 program (Nei and Gojobori 1986).

6.3 Results and Discussion

6.3.1 Sequence and phylogenetic analysis of cylindrospermospin-producing cyanobacteria

6.3.1.1 Distribution of cylindrospermopsin-producing genera through analysis of 16S rRNA gene phylogeny.

A multiple alignment of 44 partial 16S rRNA gene sequences (457 bp) was performed (Appendix 4). Of the total sequences used within the alignment, 21 sequences were obtained through this study. The remaining sequences were sourced either from Genbank or through collaborative work with A. Stüken (Stüken 2008; Stüken et al. 2009). The alignment was then used to infer the phylogeny of the sequences and determine the overall distribution of the genera and species of recognised cylindrospermopsin producers among the cyanobacterial lineages. Sporadic distribution of morphological or physiological traits amongst closely related organisms has been attributed as evidence of possible HGT events (Ochman et al. 2000; Boucher et al. 2003).

The 16S rRNA gene sequences included 9 CYN-producing and 2 non-CYN-producing *C. raciborskii.;* a confirmed saxitoxin producing *C. raciborskii* (T3) and the LMECYA132 (Valerio et al. 2009) isolate that has no demonstrated toxicity. Also included were 11 CYN-producing *Aph ovalisporum*, 2 CYN-producing and 4 non-CYN-producing *A. bergii* and a CYN-producing *A. lapponica* obtained from Finland (Spoof et al. 2006). An anatoxin-a producing *Aph. issatschenkoi* from New Zealand was also included in the study as the majority of genes associated with CYN toxicity were amplified through PCR, prior to their apparent loss soon after isolation (Wood et al. 2007b). Three CYN-producing *Aphanizomenon* sp. isolated from Germany and initially morphologically identified as *Anabaena flos-aquae* were also included. Other sequences available on Genbank from reported CYN-producers *Umezakia natans* TAC101(Harada et al. 1994) and *Raphidopsis curvata* HB1 (Li et al. 2001) were included to help evaluate the distribution of CYN producing species.

The 16S rRNA sequence available from the Genbank database pertaining to the CYN-producing *Aph. ovalisporum* isolate ILC-146 from Israel (Shalev-Alon et al. 2002), revealed a single nucleotide polymorphism (SNP) at position 65 (from a G to a T) in comparison to all other *Aph. ovalisporum* species. This is likely to be a sequencing artefact, as sequence available at the same locus from a different culture of the same strain, re-labelled as KA1.1(Stüken et al. 2009) is also available on Genbank and this SNP is not evident. Therefore the second submission of the 16S rRNA gene sequence for this isolate has been used throughout these phylogenetic analyses.

The Australian *Aph. ovalisporum* isolates, APH 035C, APH 035D, APH 035E and APH 035F displayed one SNP at position 12, where an A was substituted for a C. This however, was not perceived as an artefact as all were isolated from the same bloom. Therefore it appears reasonable to assume that they could display an identical yet individual genotype at this locus compared to *Aph. ovalisporum* isolated elsewhere.

The topology of phylogenetic trees generated by all methods did not differ in their geometry, and are therefore represented by the NJ tree in Figure 6.2. Bootstrap values obtained through NJ, MP, and ML are only given at nodes demonstrating support of 50% and above. The trees supported the filamentous, heterocyst- forming cyanobacteria; which includes orders of both Nostocales and Stigonematales; as a monophyletic cluster. *Planktothrix agardhii* 213 and NIVA-CYA 127 sequences were used as representatives of the order Oscillatoriales, which are filamentous but lack heterocytes and akinetes, and cluster distantly from the Nostocales.

Within the filamentous heterocyst-forming cluster, two distinct sub clusters are discernible and are representative of *Aph. ovalisporum* and *C. raciborskii*. The observation that these two species form tight monophyletic clusters within the order of Nostocales has been reported elsewhere (Wilson et al. 2000; Stüken et al. 2009). All *C. raciborskii* strains revealed 100% sequence identity with each other, despite differences in geographical location of isolation or toxicity.

Phylogenetic analysis of the remaining sequences highlighted the polyphyletic nature of the genera *Anabaena* and *Aphanizomenon* (Gugger et al. 2002; Rajaniemi et al. 2005; Stüken et al. 2009) and is perhaps reflective of the inherent difficulties when systematically classifying organisms based on morphological characters alone. Polyphyletism is also observed within the order of Stigonematales (Gugger and Hoffmann 2004; Valerio et al. 2009) as demonstrated by *U. natans* TAC101, which clustered more closely with *Aph. ovalisporum* than with other members assigned to this order, including *Fischerella* sp. MV-9, PCC7414 and *Nostochopsis* sp. 89-45.

The CYN producing *Anabaena bergii* strains ANA 283A and ANA 366B cluster with CYN producing *Aph. ovalisporum,* which has been reported previously (Fergusson and Saint 2003; Kellmann et al. 2006; Stüken et al. 2009) and was more thoroughly investigated by Stüken et al. (2009). Through Stüken's work it is evident that *A. bergii* and *Aph. ovalisporum* are

closely related but genetically distinct, as isolates taxonomically identified as *A. bergii* from Germany, Senegal and Australia (with exception of ANA 283A and ANA 366B) clustered together but apart from *Aph. ovalisporum*. Interestingly, all *A. bergii* in this group were not identified as CYN producers and it is thus felt that isolates from these taxa are unlikely to be major CYN producers. These observations are therefore supportive of the suggestion that *A. bergii* isolates ANA 283A and ANA 366B were morphologically misclassified and throughout the course of this work these strains will be analysed within the dataset obtained for *Aph. ovalisporum*

The strains that were identified as CYN producers have been indicated on the 16S rRNA gene tree (Figure6.2) and display a sporadic distribution amongst the heterocyst and akinete-forming filamentous cyanobacteria. It is unclear whether this sporadic distribution could also extend to include the filamentous heterocyst and akinete-lacking genera; to date only a *Lyngbya wollei* (Seifert et al. 2007) has been identified as a potential producer but no sequence from this organism is available for analysis.

The observed distribution could be regarded as indicative that HGT has played a role in disseminating the cylindrospermopsin genes amongst different cyanobacterial lineages; however there is an alternate hypothesis that patchy distribution could also be representative of an ancient origin for the genes and a result of cumulative gene loss (Woese 1987; Kurmayer et al. 2004; Rantala et al. 2004). To evaluate which of the above hypotheses is representative of cylindrospermopsin evolutionary history, additional phylogenetic information was warranted and therefore congruency between genes implicated in primary metabolism and toxicity was undertaken. The following six sections relate to the sequence and phylogenetic analysis and comparisons of the primary metabolism 16S rRNA and *rpo*C1 genes and 5 genes implicated in the biosynthesis of CYN.

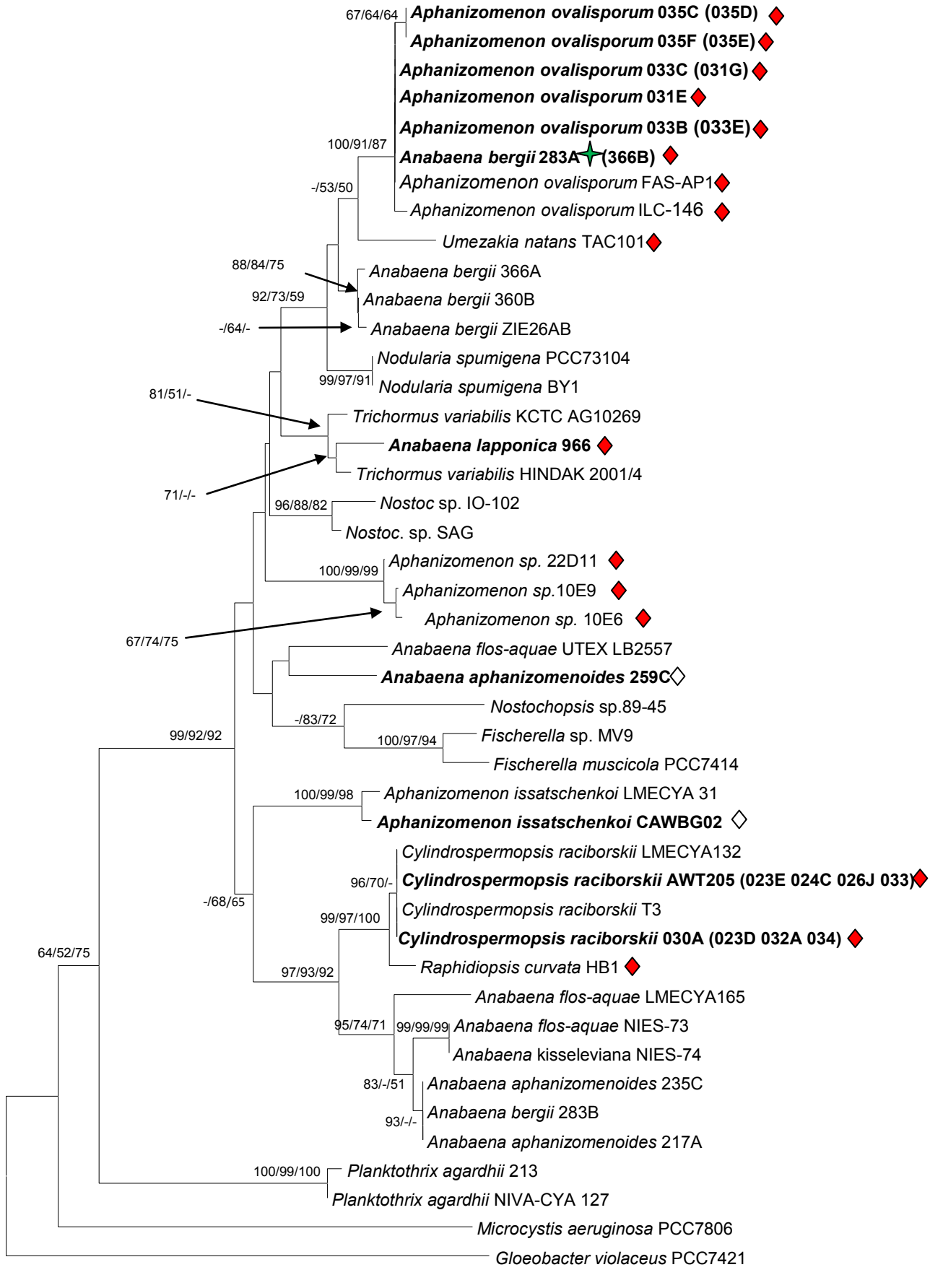

67/64/64
Aphanizomenon ovalisporum 035C (035D)
Aphanizomenon ovalisporum 035F (035E)
Aphanizomenon ovalisporum 033C (031G)
Aphanizomenon ovalisporum 031E
Aphanizomenon ovalisporum 033B (033E)
100/91/87
Anabaena bergii 283A (366B)
-/53/50
Aphanizomenon ovalisporum FAS-AP1
Aphanizomenon ovalisporum ILC-146
Umezakia natans TAC101
88/84/75
Anabaena bergii 366A
92/73/59
Anabaena bergii 360B
-/64/-
Anabaena bergii ZIE26AB
Nodularia spumigena PCC73104
99/97/91
Nodularia spumigena BY1
81/51/-
Trichormus variabilis KCTC AG10269
Anabaena lapponica 966
Trichormus variabilis HINDAK 2001/4
71/-/-
96/88/82
Nostoc sp. IO-102
Nostoc. sp. SAG
Aphanizomenon sp. 22D11
100/99/99
Aphanizomenon sp.10E9
Aphanizomenon sp. 10E6
67/74/75
Anabaena flos-aquae UTEX LB2557
Anabaena aphanizomenoides 259C
Nostochopsis sp.89-45
-/83/72
99/92/92
100/97/94
Fischerella sp. MV9
Fischerella muscicola PCC7414
100/99/98
Aphanizomenon issatschenkoi LMECYA 31
Aphanizomenon issatschenkoi CAWBG02
Cylindrospermopsis raciborskii LMECYA132
96/70/-
Cylindrospermopsis raciborskii AWT205 (023E 024C 026J 033)
-/68/65
Cylindrospermopsis raciborskii T3
99/97/100
Cylindrospermopsis raciborskii 030A (023D 032A 034)
64/52/75
Raphidiopsis curvata HB1
97/93/92
Anabaena flos-aquae LMECYA165
95/74/71
99/99/99
Anabaena flos-aquae NIES-73
Anabaena kisseleviana NIES-74
83/-/51
Anabaena aphanizomenoides 235C
93/-/-
Anabaena bergii 283B
Anabaena aphanizomenoides 217A
100/99/100
Planktothrix agardhii 213
Planktothrix agardhii NIVA-CYA 127
Microcystis aeruginosa PCC7806
Gloeobacter violaceus PCC7421
0.02 nucleotide substitutions per site

Figure 6.2 Phylogenetic relationships of selected filamentous cyanobacteria genera and strains using partial 16S rRNA gene sequences.

Gloeobacter violaceus PCC7421 and Microcystis aeruginosa PCC7806 sequences were included as out-groups. Taxa that were known to produce cylindrospermopsin are denoted by ♦. Taxa that had amplified remnants of CYN genes homologues are denoted by ◇ A ✦ represents a strain that had since lost the ability to produce CYN within the course of this work. All sequences obtained through this work are indicated in bold face. Bootstrap percentages are indicated at nodes and are given as follows: NJ, MP and ML. Some isolates whose sequences demonstrate 100% similarity were grouped to facilitate computational efficiency and are listed in brackets following a representative isolate. *Cylindrospermopsis raciborskii* strains were grouped in accordance to the two types identified by Wilson et al. (1998).

6.3.1.2 Analysis of 16S rRNA gene sequences and phylogeny between cylindrospermopsin producing cyanobacteria

A multiple sequence alignment of partial sequences obtained from within the 16S rRNA gene locus was complied for producers of CYN or species that amplified homologues to CYN genes, but were not necessarily CYN-producers themselves (Appendix 5). This was to permit the evaluation of CYN gene phylogeny in comparison to the primary metabolic genes of the same producer organisms. A complete description of the sequence analysis pertaining to *C. raciborskii* and *Aph. ovalisporum* 16S rRNA genes has been already detailed in section 6.3.1.1.

Alignment of the German *Aphanizomenon* isolates showed a sequence identity of 99.6 % to 99.8 % in respect to each other, with base differences at 2/457 positions. The lone isolates of *A. lapponica*, *A. issatschenkoi* and *A. aphanizomenoides* all revealed closest sequence identities to *Aph. issatschenkoi* (95.6 %), *A. lapponica* (95.6 %) and *A.lapponica* (95.2 %) respectively.

The phylogenetic tree (Figure. 6.3) reveals three distinct clusters, representing the main identified CYN-producers, *Aph. ovalisporum* (Group A), *Aphanizomenon sp* (Group B) and *C. raciborskii* (Group C). One other cluster is evident that is weakly supported, for the clustering of the *A. lapponica, A. issatschenkoi and A. aphanizomenoides.*

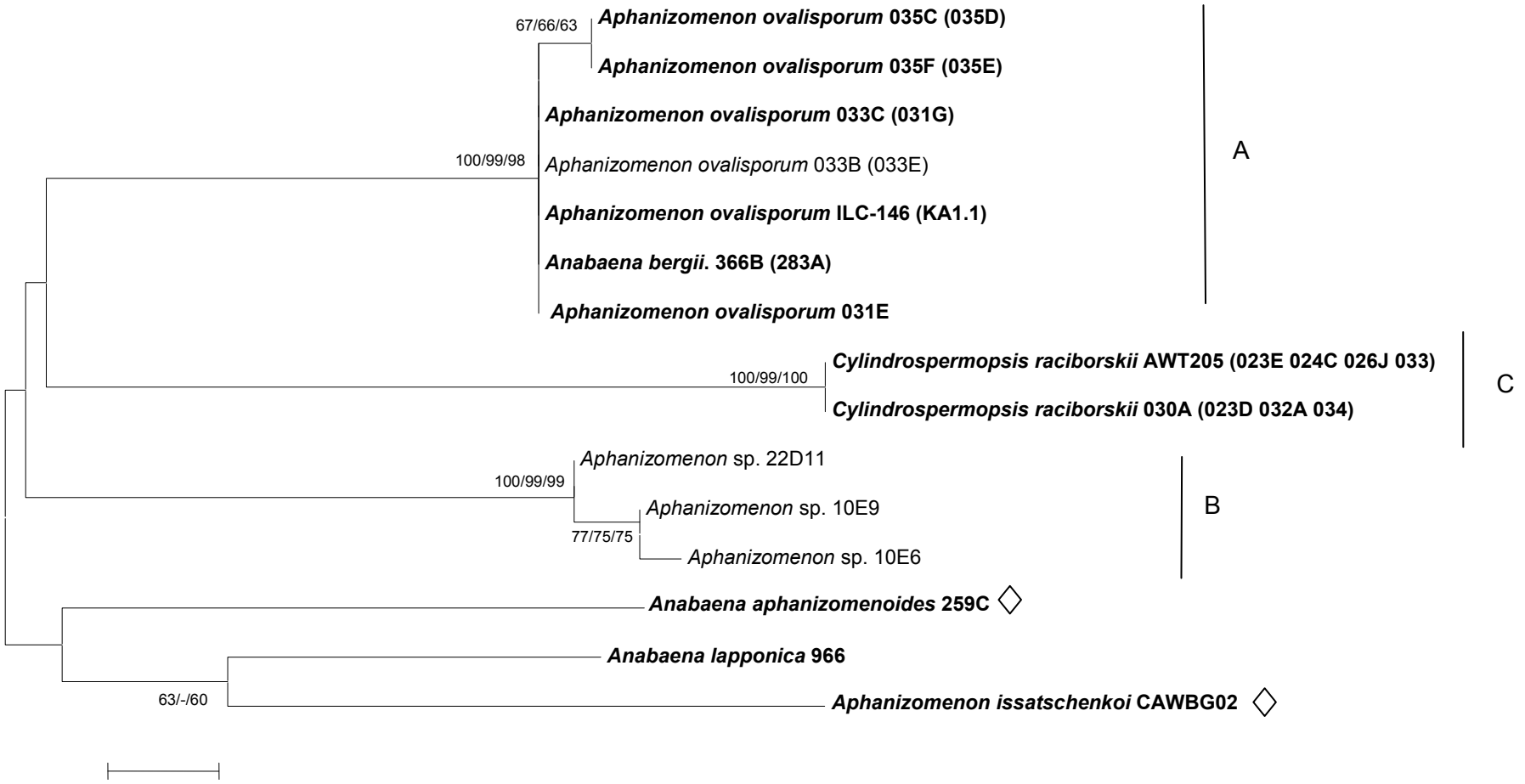

67/66/63
Aphanizomenon ovalisporum 035C (035D)
Aphanizomenon ovalisporum 035F (035E)
Aphanizomenon ovalisporum 033C (031G)
100/99/98
Aphanizomenon ovalisporum 033B (033E)
Aphanizomenon ovalisporum ILC-146 (KA1.1)
Anabaena bergii. 366B (283A)
Aphanizomenon ovalisporum 031E
A
Cylindrospermopsis raciborskii AWT205 (023E 024C 026J 033)
100/99/100
Cylindrospermopsis raciborskii 030A (023D 032A 034)
C
Aphanizomenon sp. 22D11
100/99/99
Aphanizomenon sp. 10E9
77/75/75
Aphanizomenon sp. 10E6
B
Anabaena aphanizomenoides 259C
Anabaena lapponica 966
63/-/60
Aphanizomenon issatschenkoi CAWBG02
0.02 nucleotide substitutions per site

Figure 6.3 The NJ phylogenetic tree reconstructed for the 16S rRNA gene of CYN-producing isolates.

The genetic distances were calculated from a 457 bp region using the Log Det (corrected) algorithm (Tamura and Kumar 2002). Bootstrap percentages demonstrating a confidence of ≥ 50% are indicated at the nodes and are given as follows: NJ, MP, ML. Taxa that have amplified remnants of CYN genes homologues but are non-CYN producers are denoted by ◇. Sequences obtained from this study are in bold face. Some isolates whose sequences demonstrate 100% similarity were grouped to facilitate computational efficiency and are listed in brackets following a representative isolate. *Cylindrospermopsis raciborskii* strains were grouped in accordance to the two types identified by Wilson et al. (1998)

6.3.1.3 Analysis of *rpo*C1 gene sequences and phylogeny between cylindrospermopsin producing cyanobacteria

Oligonucleotide primers listed in Table 6.2 (section 6.2.3.2) amplified a 600-612 bp fragment targeting the *rpo*C1 locus, 457 bp was used in the subsequent alignment (Appendix 6). The data set consisted of 26 isolates from various CYN-producing cyanobacteria that were also included in the 16S rRNA gene analysis presented in Figure 6.3 (section 6.3.1.2).

Sequence analysis reveals two genotypes for *C. raciborskii,* correlating to type I and type II as previously identified by (Wilson 1998). *C. raciborskii* Type II showed a non synonymous base change from C to A at position 451, which results in an amino acid change from glutamine to lysine. Overall gene sequences from the *rpo*C1 locus for *C. raciborskii* are highly conserved with a 99.8 % to 100 % sequence identity observed.

The multiple alignment of *rpo*C1 sequences for *Aph. ovalisporum* and *A. bergii* also revealed high conservation, 99.8 % to 100 % sequence identity, in respect to each other (Appendix 6). For *Aph. ovalisporum* and the two *A. bergii* sequences examined there appears to be two possible genotypes, revolving around a SNP occurring at position 561. For 9 out of 12 isolates (including *A. bergii*, ANA 283B) an identical sequence was observed (Type I), two isolates had a base change from C to T (Type IIA) and the *A. bergii* isolate ANA 366B (Type IIB) a base change from C to G. The polymorphisms are synonymous for all sequences.

The three German *Aphanizomenon* sp. isolates showed a 99.5 % to 100 % identity with respect to each other. Isolate 22D11 revealed 3 base changes in respect to the sequences of isolates 10E6 and 10E9, all base changes were synonymous. It has previously been demonstrated that the DNA-dependent RNA polymerase gene, *rpoC1*, has greater discriminatory powers at the strain level for cyanobacteria compared with 16S rRNA gene sequences (Toledo and Palenik 1997; Wilson 1998). These observed sequence variances for the German *Aphanizomenon* may therefore be reflective of strain discrimination based on geographical origin as isolate 22D11 was initially isolated from Heiliger See, Potsdam whilst the other strains where isolated from Melangsee (Preussel et al. 2006). Of the other *Aphanizomenon* represented in this study, the German isolates showed the closest sequence identity to the *Aphanizomenon issatschenkoi* (82.9 %) and were closer in sequence identity to *C. raciborskii* (77.9 %) than *Aph. ovalisporum* (74.9 % to 75.1 %).

The lone representative isolates of *A. lapponica, Aph issatchenkoi and A. aphanizomenoides* revealed closest sequence identity to the German *Aphanizomenon*, (81.6 %), *A.lapponica* (80.8%), *Aph. ovalisporum* (82.4 % to 82.5 %), respectively.

Phylogenetic comparisons between the 16S rRNA (Figure 6.3) and *rpoC1* (Figure 6.4) gene trees are overall similar in their topology, though the branching patterns between *Aph. issatchenkoi*, *A. lapponica* and *A. aphanizomenoides* are not always consistent nor well supported. These inconsistencies in branching pattern and poor support are not uncommon amongst closely related genera, and could be a result of the abovementioned species being "loner" sequences, due to the lack of available sequence data from close relatives to help determine their true phylogenetic affinities within this study (Wilmotte and Herdman 2001). The internal nodes were not well supported in general for either tree between any of the different tree constructing methods used, with exception to the 3 distinct clusters (Groups I-III) pertaining to the *Aph. ovalisporum*, *C. raciborskii* and *Aphanizomenon* sp. clusters. It was only in the *rpoC1* tree that there appeared to be some resolution of the position of the *A. aphanizomenoides* ANA 259C, which was significantly supported and clustered closely to the *Aph. ovalisporum*.

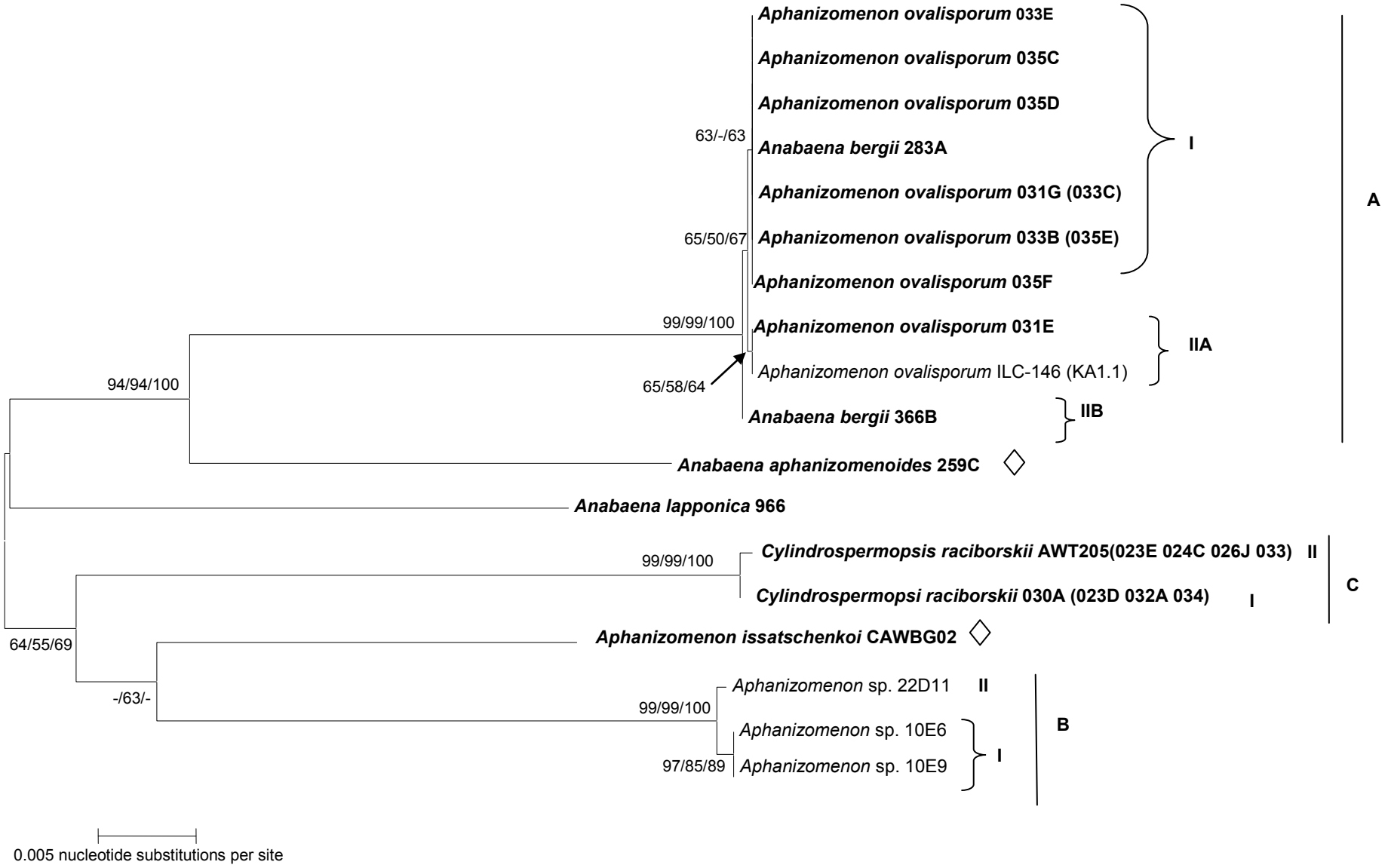

Aphanizomenon ovalisporum 033E
Aphanizomenon ovalisporum 035C
Aphanizomenon ovalisporum 035D
63/-/63
Anabaena bergii 283A
Aphanizomenon ovalisporum 031G (033C)
65/50/67
Aphanizomenon ovalisporum 033B (035E)
Aphanizomenon ovalisporum 035F
I
A
99/99/100
Aphanizomenon ovalisporum 031E
Aphanizomenon ovalisporum ILC-146 (KA1.1)
IIA
94/94/100
65/58/64
Anabaena bergii 366B
IIB
Anabaena aphanizomenoides 259C
Anabaena lapponica 966
99/99/100
Cylindrospermopsis raciborskii AWT205(023E 024C 026J 033) II
Cylindrospermopsi raciborskii 030A (023D 032A 034) I
C
64/55/69
Aphanizomenon issatschenkoi CAWBG02
-/63/-
Aphanizomenon sp. 22D11 II
99/99/100
Aphanizomenon sp. 10E6
97/85/89
Aphanizomenon sp. 10E9
I
B
0.005 nucleotide substitutions per site

Figure 6.4 The NJ phylogenetic tree reconstructed for the *rpo*C1 gene of CYN-producing isolates.

The genetic distances were calculated from a 567 bp region using the Log Det (corrected) algorithm (Tamura and Kumar 2002). Bootstrap percentages demonstrating a confidence of ≥ 50% are indicated at the nodes and are given as follows: NJ, MP, ML. Taxa that have amplified remnants of CYN genes homologues but are non-CYN producers are denoted by ◇ .Sequences obtained from this study are in bold face. Some isolates whose sequences demonstrate 100% similarity were grouped to facilitate computational efficiency and are listed in brackets following a representative isolate. *Cylindrospermopsis raciborskii* strains were grouped in accordance to the two types identified by Wilson et al. (1998).

6.3.1.4 Cylindrospermospin amidinotransferase gene, *cynC*

The oligonucleotide primers CC1 and CC2 were designed to amplify a 752 bp region of the amidinotransferase gene *cynC*, identified within the cylindrospermopsin biosynthetic cluster (Chapter 3, Table 3.5). A total sequence length of 670 bp was used to perform a multiple sequence alignment (Appendix 7). An additional *cynC* sequence was available in Genbank, pertaining to the *U. natans* (Kellman 2006), and was excluded from this study, as there is a potential that it could be a PCR hybrid.

All CYN producing *C. raciborskii* appeared generic and the non-CYN producing *A. aphanizomenoides* ANA 259C demonstrated a 100% nucleotide sequence identity with the *C. raciborskii* strains.

The multiple alignment revealed two possible genotypes within the CYN-producing *Aph. ovalisporum*. *Aph. ovalisporum* type I represented 8 of the 11 taxa and showed a non synonymous base change from A to G at position 607, resulting in an amino acid change from isoleucine to valine. *Aph. ovalisporum* type II revealed two non synonymous base changes. The first at position 224 represented a base change from C to T and resulted in an amino acid change of serine to phenylalanine; while the second was at position 616 where a base change of T to C resulted in the corresponding amino acid change from tyrosine to histidine. There was no evidence that the genotypes represented segregation based on geographical location of the isolate, nor morphology as the two putative strains initially identified by traditional classification as *A. bergii* were represented in each.

The *Aphanizomenon* isolates from Germany revealed 100 % identity with each other and demonstrated a higher degree of nucleotide sequence similarity with *C. raciborskii* (98.9 %) then to that of *Aph. ovalisporum* (95.1%-94.9%). The non-CYN producing *Aph. issatschenkoi* also had closer nucleotide sequence similarities with *C. raciborskii* (99.5 %) than with the other *Aphanizomenon* species (98.5% with the German isolates, 95.2 % to 95.4 % for *Aph. ovalisporum*).

The isolate from Finland, *Anabaena lapponica* 966, demonstrated the highest degree of sequence divergence to all other CYN-producing strains presented here. It had an 88.4 % sequence identity to the consensus sequence of *C. raciborskii* and *Aph. ovalisporum* type II,

88.5% sequence similarity to *Aph. ovalisporum* type I, and 88.8% nucleotide sequence identity to the German *Aphanizomenon* sp. This divergence was further evident by the difficulty experienced in achieving effective amplification in gene specific PCR reactions targeting regions of the cylindrospermopsin gene cluster within the *A. lapponica* genome (Chapter 3).

The *cynC* tree revealed a deeply bifurcated tree (Figure 6.5), with two distinct major clusters, one representing the *Aph. ovalisporum* as a strongly supported group (Group A) and the other comprising a series of sub-clusters consisting of *C. raciborskii* (Group C) and *Aphanizomenon* sp. (Group B) The high sequence divergence observed with the *A. lapponica* sequence was represented as a long branch effectively "rooting" the tree. The statistical support for the two major clusters was high with bootstrap values ranging from 99-100%. The *Aphanizomenon* sp. and *C. raciborskii* are also well supported clusters but the genotypes identified within the *Aph. ovalisporum* group reveal moderate support for all three tree methods used.

Comparisons between the tree topologies inferred by analysing the primary metabolic associated 16S rRNA (Figure6.3) and *rpo*C1 (Figure6.4) genes to that of the *cynC* revealed a general overall congruency, particularly in regard to the CYN-producing organisms. It appeared that there were potentially two distinct genotypes for *Aph. ovalisporum* that are reflected in both the *rpoC1* and the *cynC* loci (designated type I and type II). Clustering of the *A. bergii* ANA 283A and ANA 336B *cynC* gene data further supported the premise that these two isolates are morphological variants of *Aph. ovalisporum* and belong within the same genus. The *cynC* sequenced from the non-CYN producing *A. aphanizomenoides* ANA 259C demonstrated strong similarity to *C. raciborskii*, possibly indicative of a recent HGT event.

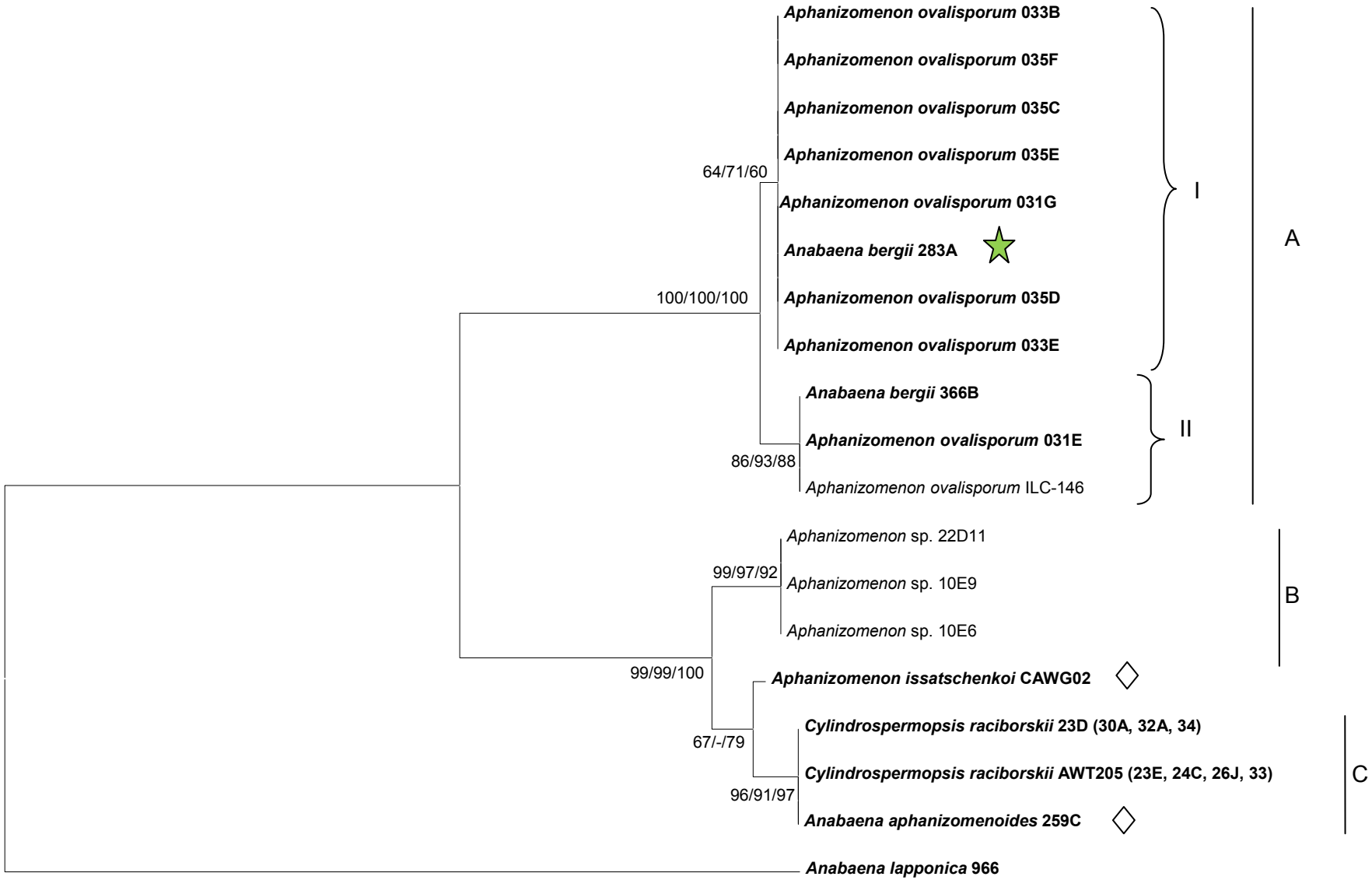

Aphanizomenon ovalisporum 033B
Aphanizomenon ovalisporum 035F
Aphanizomenon ovalisporum 035C
Aphanizomenon ovalisporum 035E
64/71/60
Aphanizomenon ovalisporum 031G
Anabaena bergii 283A
100/100/100
Aphanizomenon ovalisporum 035D
Aphanizomenon ovalisporum 033E
I
Anabaena bergii 366B
Aphanizomenon ovalisporum 031E
86/93/88
Aphanizomenon ovalisporum ILC-146
II
A
Aphanizomenon sp. 22D11
99/97/92
Aphanizomenon sp. 10E9
Aphanizomenon sp. 10E6
B
99/99/100
Aphanizomenon issatschenkoi CAWG02
Cylindrospermopsis raciborskii 23D (30A, 32A, 34)
67/-/79
Cylindrospermopsis raciborskii AWT205 (23E, 24C, 26J, 33)
96/91/97
Anabaena aphanizomenoides 259C
C
Anabaena lapponica 966
0.005 nucleotide substitutions per site

Figure 6.5 The NJ phylogenetic tree reconstructed for the Cylindrospermopsin amidinotransferase, *cynC* gene.

The genetic distances were calculated from a 670 bp region using the Log Det (corrected) algorithm (Tamura and Kumar 2002). Bootstrap percentages demonstrating a confidence of ≥ 50% are indicated at the nodes and are given as follows: NJ, MP, ML. Taxa that are not CYN-producers are denoted by ◇. *Anabaena bergii* ANA 283A lost CYN toxigenicity through the course of this work is flagged with a ★. Sequences directly obtained through this study are highlighted in bold type. Some isolates whose sequences demonstrate 100% similarity were grouped to facilitate computational efficiency and are listed in brackets following a representative isolate. *Cylindrospermopsis raciborskii* strains were grouped in accordance to the two types identified by Wilson et al (1998).

cynD

The oligonucleotide primers CD3 and CD4 were designed to amplify a 689 bp region of the acyltransferase domain (AT) within the chimeric PS/PKS gene *cynD*, of the cylindrospermopsin biosynthetic cluster (Chapter 3, Table 3.5). A total sequence length of 615 bp was used to perform a multiple sequence alignment (Appendix 8).

The alignment of the *C. raciborskii cynD* sequences revealed 100 % sequence identity between the two sequence types indentified within the *rpoC1* gene (Appendix 6, Figure 6.4) and is consistent with the sequence analysis and phylogeny of the *cynC* gene (Appendix 7, Figure 6.5). Multiple alignment and sequence comparisons of the *cynD* homologue, isolated from *A. aphanizomenides* ANA 259C, with the *C. raciborskii* sequence again displayed 100 % sequence identity. This was consistent with the findings in the *cynC* sequence and phylogenetic analysis.

Sequence and phylogenetic analysis (Figure 6.6) further divided the two genotypes, identified previously within the *Aph. ovalisporum rpo*C1 (Figure 6.4) and *cynC* (Figure 6.5) phylogenies, into a putative additional cluster (depicted as IB Figure5.5). The *Aph. ovalisporum* cluster is bifurcated as evident also in *cynC*. Two isolates of *Aph. ovalisporum*, APH 031G and APH 033C, do not cluster with the other *Aph. ovalisporum,* and revealed sequence identities ranging from 98.9 % (with *Aph. ovalisporum* type II) to 99.5 % (with *Aph. ovalisporum* type IB). The sequences of these two isolates, *Aph. ovalisporum* APH 031G and APH 033C, demonstrated 100 % sequence identity to *Aph. issatschenkoi*. This is in contrast to the *cynC* sequence and phylogeny findings where *Aph issatschenkoi* demonstrated higher sequence identity (99.5 %) and clustered closer to *C. raciborskii*.

Multiple alignments of the German *Aphanizomenon* isolates revealed a generic sequence between the three, though the strain 22D11 was incomplete and 37 bp shorter than all other sequences in the alignment. The German isolate's sequence differed to the *C. raciborskii* by only one base at position 98 resulting in a non-synonymous change from glycine to valine.

The phylogenetic comparisons between *cynD*, *cynC* and the primary metabolism genes (16S and *rpo*C1) reveal an underlying congruence between the principal CYN-producers and their primary metabolism associated genes. There is evidence that the *Aph. issatschenkoi* may have

a mosaic composition of the CYN biosynthetic homologues, perhaps suggestive of at least two separate HGT events. The *Aph. issatschenkoi cynD* gene sequence displayed a higher sequence identity with *Aph. ovalisporum* compared to the *cynC* gene data where the similarity was closer to *C. raciborskii* sequence.

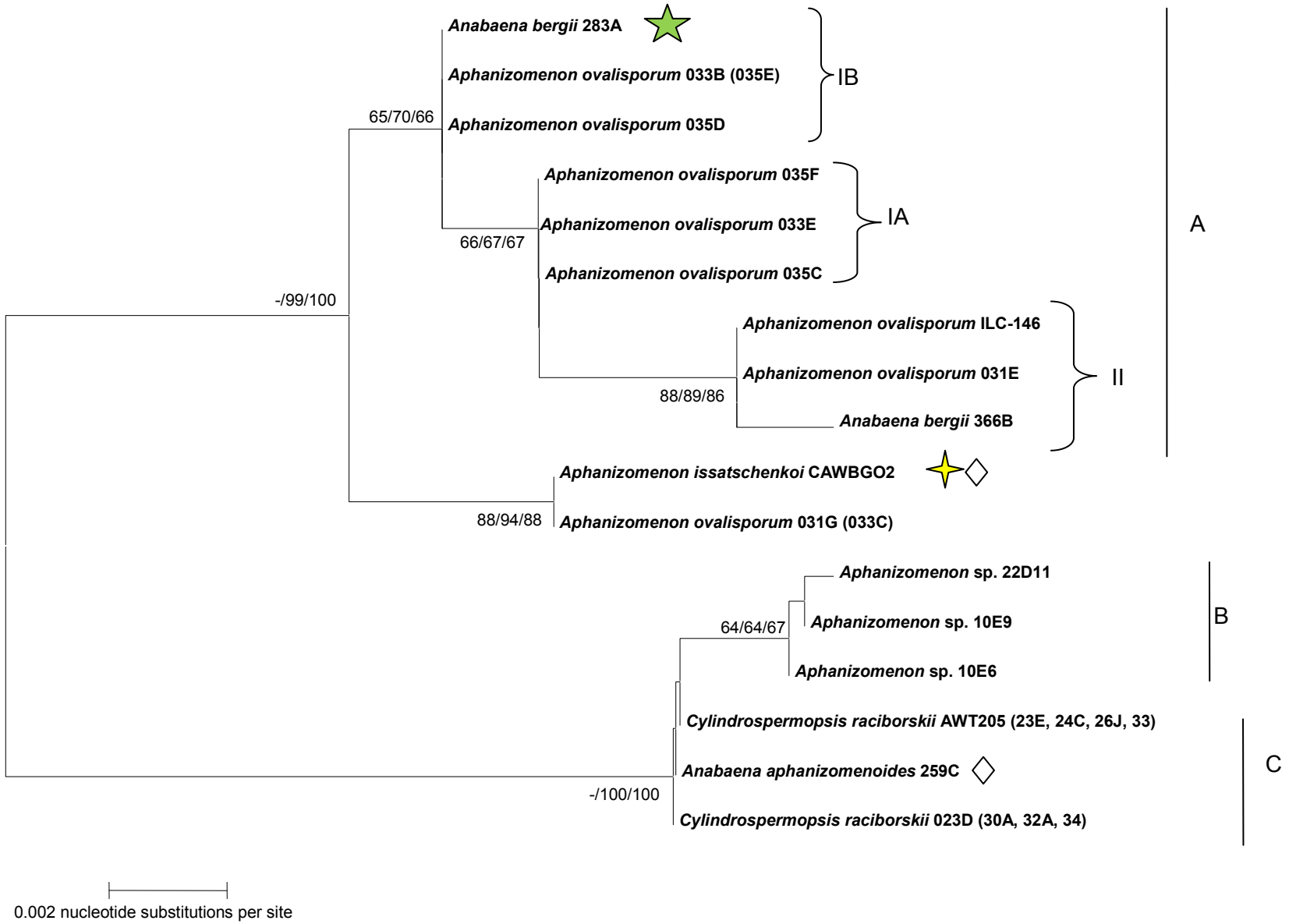

Anabaena bergii 283A
Aphanizomenon ovalisporum 033B (035E)
Aphanizomenon ovalisporum 035D
IB
65/70/66
Aphanizomenon ovalisporum 035F
Aphanizomenon ovalisporum 033E
Aphanizomenon ovalisporum 035C
IA
66/67/67
A
-/99/100
Aphanizomenon ovalisporum ILC-146
Aphanizomenon ovalisporum 031E
Anabaena bergii 366B
II
88/89/86
Aphanizomenon issatschenkoi CAWBGO2
Aphanizomenon ovalisporum 031G (033C)
88/94/88
Aphanizomenon sp. 22D11
Aphanizomenon sp. 10E9
Aphanizomenon sp. 10E6
64/64/67
B
Cylindrospermopsis raciborskii AWT205 (23E, 24C, 26J, 33)
Anabaena aphanizomenoides 259C
Cylindrospermopsis raciborskii 023D (30A, 32A, 34)
-/100/100
C
0.002 nucleotide substitutions per site

Figure 6.6 The NJ phylogenetic tree reconstructed for the Cylindrospermopsin acyltransferase domain within the PKS, *cynD* gene.

The genetic distances were calculated from a 615 bp region using the Log Det (corrected) algorithm (Tamura and Kumar 2002). Bootstrap percentages demonstrating a confidence of ≥ 50% are indicated at the nodes and are given as follows: NJ, MP, ML. Taxa that are not CYN-producers are denoted by ◇. *Anabaena bergii* 283A lost CYN toxigenicity through the course of this work is flagged with a ★. Sequences demonstrating branch swapping are denoted with a ✦Sequences directly obtained through this study are highlighted in bold type. Some isolates whose sequences demonstrate 100% similarity were grouped to facilitate computational efficiency and are listed in brackets following a representative isolate. *Cylindrospermopsis raciborskii* strains were grouped in accordance to the two types identified by Wilson et al. (1998).

cynA

The oligonucleotide primers CA1 and CA2 were designed to amplify a 689 bp region of the acyltransferase domain (AT) within the PKS gene, *cynA*, of the cylindrospermopsin biosynthetic cluster (Chapter 3, Table 3.5). A total sequence length of 655 bp was used to perform a multiple sequence alignment (Appendix 9).

Multiple sequence alignment of the *C. raciborskii* isolates revealed a generic sequence that again displays 100% sequence identity with the *A. aphanizomenoides* ANA 259C. Interestingly, 100% sequence identity was also observed between *C. raciborskii* isolates and *Aph. issatschenkoi*.

The *Aph. ovalisporum* isolates again showed that they had three sequence variants, but only type II was consistent with the same members belonging to the group in all CYN genes analysed. Sequence identities between type II and the other sequences varied from 99.3 % to 99.7 % and all base changes were non synonymous. All *Aph. ovalisporum* isolated showed a diagnostic insertion sequence of 30 bp in length.

Again the phylogenetic tree generated was deeply bifurcated and strong support was evident for the cluster consisting of the *C. raciborskii* and the *Aphanizomenon* sequence from Germany (Figure 6.7). The tree again reflected the overall congruence between the primary metabolism associated genes and the CYN-producers. There was consistency also observed in phylogeny for all amplicons isolated from the *A. aphanizomenoides,* which were all homologous to the genes within the CYN operon of *cynC*, *cynD* and *cynA*. Conflicts were apparent, however, in regards to the non-CYN producing *Aph. issatschenkoi*, which appears to be a mosaic CYN genes of different origins.

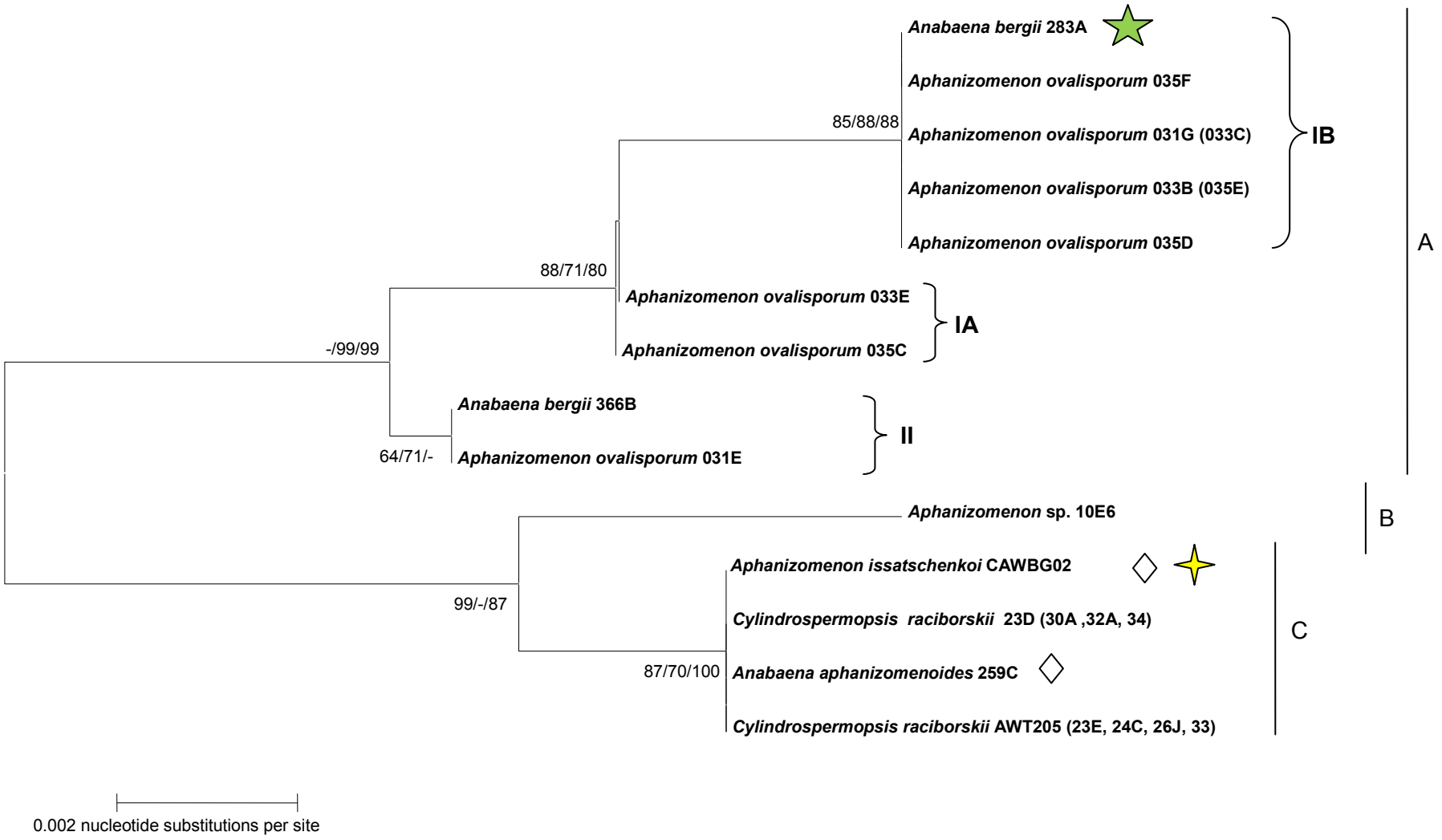

Anabaena bergii 283A
Aphanizomenon ovalisporum 035F
85/88/88
Aphanizomenon ovalisporum 031G (033C)
Aphanizomenon ovalisporum 033B (035E)
Aphanizomenon ovalisporum 035D
IB
A
88/71/80
Aphanizomenon ovalisporum 033E
IA
-/99/99
Aphanizomenon ovalisporum 035C
Anabaena bergii 366B
II
64/71/-
Aphanizomenon ovalisporum 031E
Aphanizomenon sp. 10E6
B
Aphanizomenon issatschenkoi CAWBG02
99/-/87
Cylindrospermopsis raciborskii 23D (30A ,32A, 34)
C
87/70/100
Anabaena aphanizomenoides 259C
Cylindrospermopsis raciborskii AWT205 (23E, 24C, 26J, 33)
0.002 nucleotide substitutions per site

Figure 6.7 The NJ phylogenetic tree reconstructed for the Cylindrospermopsin acyltransferase domain within the PKS, *cynA* gene.

The genetic distances were calculated from a 655 bp region using the Log Det (corrected) algorithm (Tamura and Kumar 2002). Bootstrap percentages demonstrating a confidence of ≥ 50% are indicated at the nodes and are given as follows: NJ, MP, ML. Taxa that are not CYN-producers are denoted by ◇. *Anabaena bergii* ANA 283A lost CYN toxigenicity through the course of this work is flagged with a ★Sequences demonstrating branch swapping are denoted with a ✦ Sequences directly obtained through this study are highlighted in bold type. *Cylindrospermopsis raciborskii* strains were grouped in accordance to the two types identified by Wilson et al. (1998) Some *Aph. ovalisporum* isolates whose sequences demonstrate 100% similarity were grouped to facilitate computational efficiency and are listed in brackets following a representative isolate.

cyn B

The oligonucleotide primers CB1 and CB2 were designed to amplify a 609 bp region of the acyltransferase domain (AT) within the PKS gene, *cynB*, of the cylindrospermopsin biosynthetic cluster (Chapter 3, Table 3.5). A total sequence length of 561 bp was used to perform a multiple sequence alignment (Appendix 10).

Multiple sequence alignments of the *C. raciborskii* isolates again revealed a generic sequence, with 100% sequence identity observed between *C. raciborskii* and *A. aphanizomenoides* ANA259C, a consistent observation for all homologous CYN gene sequences amplified from the non-CYN producing *A. aphanizomenoides.* There was, additionally, 100% sequence identity with the *cynB* homologue amplified from the New Zealand *Aph. issatschenkoi*, which was similarly observed for the *Aph. issatschenkoi cynA* amplicon.

The multiple sequence alignments of the *Aph. ovalisporum* isolates showed possibly three sequence variants, with type II again showing consistency in regards to its members. There was only one SNP delineating sequence variants IA and IB, with a non synonymous base change at position 400 from G to C and resulted in an amino acid change from alanine to proline. Sequence identities between sequence type II and the other *Aphanizomenon* sequences varied from 99.5% to 99.6%. There were two non synonymous changes for type II. The first at position 379 represented a base change from T to C, which leads to an amino acid change from tyrosine to arginine, whilst the second was identical to that observed for sequence type IB at position 400. Compared to the *C. raciborskii* isolates, *Aph. ovalisporum* demonstrated sequence identities ranging from 97.7% to 98%.

The *Aphanizomenon* isolates, 10E6 and 10E9, again revealed 100% sequence identity with each other, though isolate 10E9 was 117 base pairs shorter. A higher degree of nucleotide sequence similarity was again observed with *C. raciborskii* (which additionally includes *Aph. issatchenkoi* and *A. aphanizomenoides*) at 99.6%. Sequence identities with the *Aph. ovalisporum* sequence variants ranged from 97.9% to 98%.

Once again the CYN gene homologue isolated from the Finnish *A. lapponica* 966, revealed a highly divergent sequence, compared to all other CYN (and non-CYN) producing

cyanobacteria. An 88.4% sequence identity was observed between *A. lapponica* and *C. raciborskii,* similarly for the German *Aphanizomenon* isolates. The sequence identity to the *Aph. ovalisporum* genotypes ranged from 87.7% to 88%. Such high sequence divergence, observed with the *A. lapponica* CYN gene fragments, could possibly be a consequence of accelerated adaptive evolution. *Anabaena lapponica* was isolated from the coldest northern latitude for a CYN producing organism thus far (Spoof et al. 2006). Studies investigating amino acid composition of genomes from cryophilic Archaea, have identified preferential use of amino acidsthat putatively assist in cold adaptation of the organisms by optimising the structure of functioning proteins (Saunders et al. 2003). A level of cold adaptation for the boreal *A. lapponica* may also be essential to permit efficient function. To assess the selection pressures enacting on the *A. lapponica* cluster, an analysis of d_N/d_S ratios (outlined in section 6.2.6) was undertaken. Pairwise comparisons between the *A. lapponica cynB* sequence and those obtained from *Aph. ovalisporum* and *C. raciborskii* revealed d_N/d_S ratios of 1.7 and 1.4 respectively, ratios > 1 are expected under adaptive evolution. Therefore, the obtained ratios are supportive of the hypothesis that positive selection is influential in the sequence diversity observed for *A. lapponica*.

The *cynB* tree (Figure 6.8) revealed three stable groups, one representing *Aph. ovalisporum,* the other comprising two subclusters consisting of *C. raciborskii* and *Aphanizomenon sp.* The highly divergent *A. lapponica* is again revealed as a long branch effectively rooting the tree. Throughout this work, all comparisons between the phylogeny of toxin-associated genes (*cynC*,*cynD*, *cynA* and *cynB*) with those associated with primary metabolism (16S rRNA and *rpo*C1genes), revealed an underlying congruence. These findings are comparable to phylogenetic conclusions drawn for the microcystin biosynthetic gene cluster (Rantala et al. 2004).

This is suggestive that the biosynthetic gene cluster, involved in the production of CYN, was evident in the genome of the last common ancestor of the heterocystous, akinete-forming cyanobacteria, of the order Nostocales. Further, this suggests that there has been subsequent loss of the gene cluster from the genome, resulting in non-toxic strains. This premise is perhaps further supported by the coincidental loss of CYN toxicity that was observed with isolate *A. bergii* ANA283A, determined by both absence of amplifiable CYN gene targets and lack of CYN detected in all extracts analysed by HPLC.

However, there were incidences of incongruence that may be reflective of recent HGT events of all or even partial segments of the CYN gene cluster. The New Zealand isolate *Aph. issatschenkoi* amplified CYN genes by PCR, despite no detectable CYN in any of its extracts. The amplified sequences displayed high sequence identity to and clustered with *C. raciborskii.* An exception for this isolate, was the *cynD* amplicon, which clustered with *Aph. ovalisporum*. This potentially implies that the chromosomal composition of these genes within *Aph. issatschenkoi* are mosaic and possibly indicative of 2 to 3 independent HGT events.

Anabaena aphanizomenoides phylogeny also revealed incongruence with respect to toxin-related genes and those associated with primary metabolism. Interestingly, *A. aphanizomenoides* ANA 259C, which amplified the CYN operon (*cynC*, *cynD* and *cynA*) and *cynB* homologues but no other CYN related genes (Chapter 3, Figure 3.12), was isolated during a survey of cyanobacterial blooms within the Murray Darling Basin (Baker and Humpage 1994). Whilst this particular isolate was not originally analysed for toxicity, as *A. aphanizomenoides* are not classically regarded as a toxic species within Australia (P Baker, personal communication), some *A. aphanizomenoides* isolated during the same survey were. Of the 16 *A. aphanizomenoides* strains analysed through mouse bioassay, 5 produced subacute hepatotoxic injury. None of these 5 isolates were viable at the time of this study, to permit further analysis. Therefore, two possible scenarios potentially exist to explain the presence of the CYN operon *(cynC*, *cynD* and *cynA*) and *cynB* genes in the *A. aphanizomenoides* genome.

1. *Anabaena aphanizomenoides* ANA259C was once a producer of CYN and acquired the suite of genes through HGT but, is eventually losing the biosynthetic cluster from its genome.

2. The presence of the genes displaying high identity to the CYN operon *(cynC*, *cynD, cynA*) and *cynB* genes, may have arisen through a gene duplication event (paralogous evolution) with a subsequent evolutionary development of a new function. The genes may be involved in an alternative metabolic pathway. The alternate compound may potentially be similar structurally to that of CYN, but potentially any toxicity has been modified or nulled and even differs in mode of action.

To further examine the evolutionary history of the CYN biosynthetic cluster in *C. raciborskii*, and to potentially strengthen the ancient origin of the gene cluster premise, heterogeneity in codon bias within the *C. raciborskii* AWT205 genome, was evaluated through relative synonymous codon usage (RSCU).

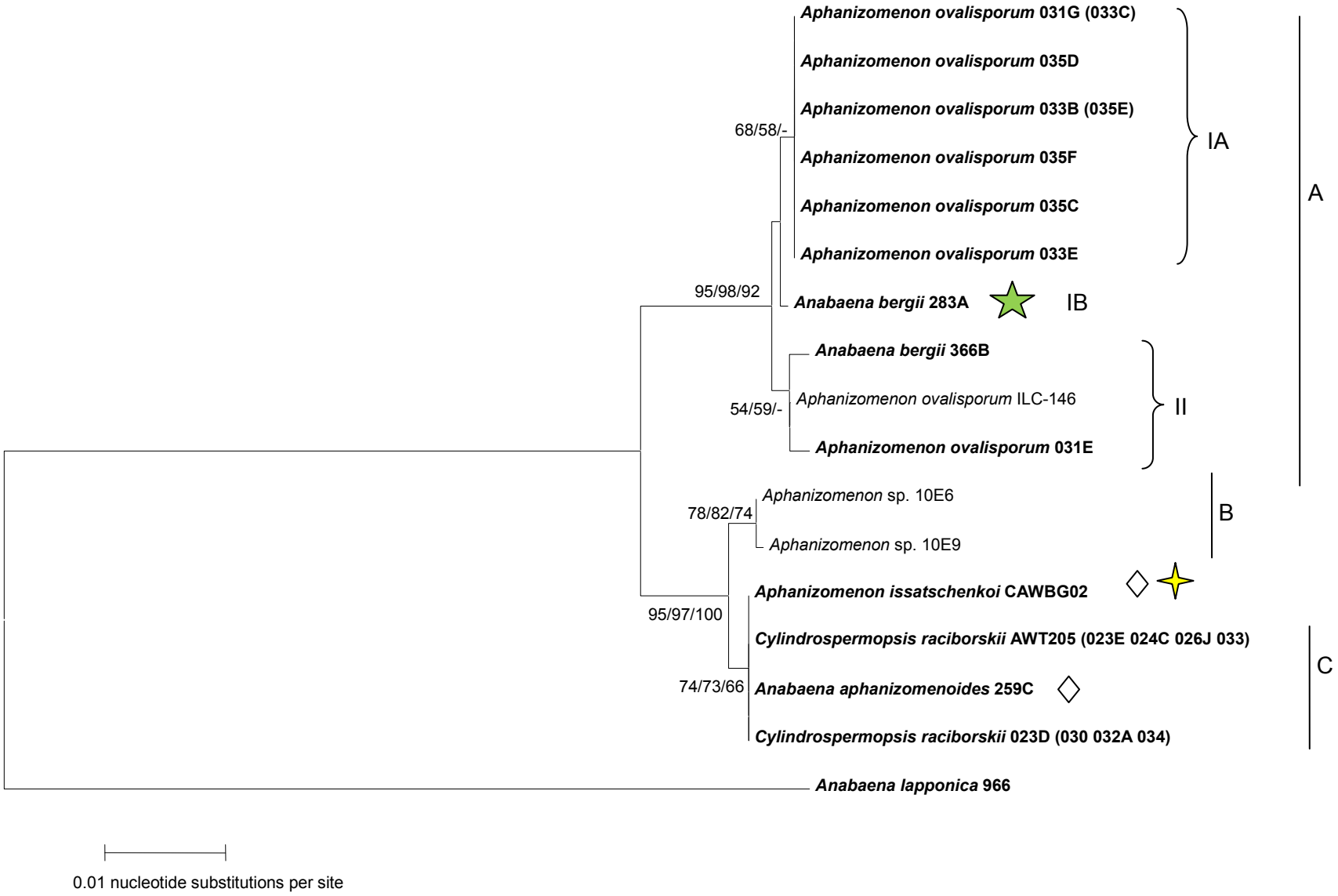

Aphanizomenon ovalisporum 031G (033C)
Aphanizomenon ovalisporum 035D
Aphanizomenon ovalisporum 033B (035E)
68/58/-
Aphanizomenon ovalisporum 035F
Aphanizomenon ovalisporum 035C
Aphanizomenon ovalisporum 033E
IA
A
95/98/92
Anabaena bergii 283A
IB
Anabaena bergii 366B
54/59/-
Aphanizomenon ovalisporum ILC-146
Aphanizomenon ovalisporum 031E
II
Aphanizomenon sp. 10E6
78/82/74
Aphanizomenon sp. 10E9
B
Aphanizomenon issatschenkoi CAWBG02
95/97/100
Cylindrospermopsis raciborskii AWT205 (023E 024C 026J 033)
74/73/66
Anabaena aphanizomenoides 259C
Cylindrospermopsis raciborskii 023D (030 032A 034)
C
Anabaena lapponica 966
0.01 nucleotide substitutions per site

Figure 6.8 The NJ phylogenetic tree reconstructed for the Cylindrospermopsin acyltransferase domain within the PKS, *cynB* gene.

The genetic distances were calculated from a 561 bp region using the Log Det (corrected) algorithm (Tamura and Kumar 2002). Bootstrap percentages demonstrating a confidence of ≥ 50% are indicated at the nodes and are given as follows: NJ, MP, ML. Taxa that are not CYN-producers are denoted by ◇ . *Anabaena bergii* ANA283A lost CYN toxigenicity through the course of this work is flagged with a ★ . Sequences directly obtained through this study are highlighted in bold type. Sequences demonstrating branch swapping are denoted with a ✦ *Cylindrospermopsis raciborskii* strains were grouped in accordance to the two types identified by Wilson et al. (1998) Some *Aph. ovalisporum* isolates whose sequences demonstrate 100% similarity were grouped to facilitate computational efficiency and are listed in brackets following a representative isolate.

6.3.2 Codon usage in *C.raciborskii* AWT205

Genes that are suspected to be acquired by HGT events can potentially be detected by the comparison of codon usage to genes that are considered indigenous to the recipient. For any given genome, codon usage may be heterogeneous from gene to gene and it is recognised that an overall codon bias is species specific (Kaplan and Fine 1998; Garcia-Vallve et al. 2000). Relative synonymous codon usage (RSCU) is recommended to account for differences in sequence length and amino acid composition of different gene products (Sharp et al. 1986). Converting codon usage into RSCU values effectively permits a "normalised" comparison between the gene sets.

Coding regions of genes identified within the cylindrospermopsin biosynthetic cluster (Chapter 3) were concatenated into two separate groups based on their location either internal or external of the putative transposase identified between the *cynF* and *cynH* ORFs (Chapter 3, Figure 3.11). The organisation of the concatenation was randomised and assessed in various orders of sequence input to assess any impact on the RSCU calculations. There was no detectable consequence with regard to concatenation order to the overall output of the RSCU calculations.

RSCU calculations were organised, on the assumption that if the transposase had any role in facilitating the insertion of gene sequences with the CYN biosynthetic cluster of *C. raciborskii* AWT205, anomalous codon usage may be detected from those genes internal and potentially imported by the transposase in comparison to genes situated externally to the transposase. The RSCU of the toxin genes were compared to primary metabolism associated genes available on Genbank or from this study. Partial sequences pertaining to *rpoC1* (this work), *nifH* (accession number AF426768) and the phycocyanin alpha and beta loci, after removal of the interspacer sequence (accession number AF426787), were available and concatenated resulting in an average of 452 codons for analysis. All sequences were isolated from *C. raciborskii* AWT205. The RSCU comparisons are presented in Table 6.3.

In general, codon usage across the toxin gene subsets, and those involved in primary metabolism appeared to be highly similar. There were observations of significant codon bias only within the primary metabolism gene set (data not shown), but for only three of the 20 amino acids (arginine, glutamine and phenylalanine). For example, GAA was preferred 100%

over GAG for the glutamine codons amongst the primary metabolism genes, but the relative frequency of the bias, was weak for the CYN toxin genes within the transposase. Whilst, it is recognised that direction of codon bias is largely uniform within the same taxa, observations of bias differences have been noted between genes within a genome that are not necessarily attributable to HGT events. It is thought that some genes in *E. coli* that demonstrate significant codon bias correlates to genes that are constitutively expressed at higher levels (Medigue et al. 1991). Similarly, the codon bias observed for the primary metabolism associated genes of *C.raciborskii,* may also be indicative of the constraints imposed to maximise translational efficiency.

There are only two codons preferred between primary metabolism associated genes and those CYN toxin genes found external of the transposase in comparison to the preferred codon dominating those CYN genes that were internal to the transposase (Table 6.3).This can be indicative of two likely scenarios, 1) Horizontal transfer was not a factor in the history of CYN, thereby supporting the proposal that the cluster is an ancient ancestral trait or 2) the HGT event was ancient and over the passage of time ameriolation of the exogenous sequence has become evident or the donor had a similar codon usage to that of the recipient, *C. raciborskii* AWT205.

Therefore the evidence gathered throughout this work is not wholly supportive that HGT events have played a major role in the transmission of the CYN biosynthetic cluster in recent times. That is not to be completely dismissive of HGT, as detection of ancient events is problematic due to the inherent issues outlined above and phylogenetic conflicts were observed for the non-CYN producing *A. issatschenkoi* and *A. aphanizomenoides*.

The results presented here are to some extent, inconsistent to those found by Kellman (2006) and Stucken (2009). Whilst Kellman et al. (2006) does recognise that there is observed phylogenetic stability between the identified CYN producing cyanobacterial groups of *C. raciborskii* and *Aph. ovalisporum*, the authors concluded that HGT was influential factor in disseminating the toxic gene cluster through cyanobacterial lineages. The support for these conclusions was two-fold:

1. the high sequence identity of the toxin gene homologues between genera compared to their primary metabolism 16S rRNA genes; and
2. observed phylogenetic incongruency of the *U. natans* and a CYN homologue isolated from a non-CYN producing Hungarian *C. raciborskii.*

However, high sequence identity between generas with respect to the toxin genes could also be viewed as evidence that they were acquired from a common ancestor. Whilst the sequence identity of the CYN gene fragments isolated from non-CYN producing cyanobacteria and are suggestive of natural deletion mutants, convergent evolution needs to be excluded.

Stucken's (2009) work was unable to determine the presence of natural deletion mutants and on that principle felt the acquisition and loss of the toxic gene cluster was potentially akin to a pathogenic island and thus supportive of HGT. However, both Kellman (2006) and the work presented within this thesis suggest otherwise, as in both studies putative natural deletion mutants were detected. Additionally the studies undertaken here include a larger data set then previous investigations, which permits greater resolution and presents a stronger case for the findings detailed within this chapter.

Intragenic recombination has played a significant role toward the diversity observed within the microcystin biosynthetic cluster (Tanabe et al. 2004; Fewer et al. 2007). From the molecular characterisation of the CYN biosynthetic cluster detailed in chapter 3, it was evident that intergenic recombination was also apparent. To explore the extent of recombination within the CYN gene cluster, structural comparisons between *C. raciborskii* and *Aph. ovalisporum* were examined.

Table 6.3 **The amino acids (excluding termination and initiation codons) and their preferred codon for coding regions of the CYN biosynthetic genes, internal and external to the putative transposase *orf* boundaries and primary metabolism associated genes.**

Amino acid	CYN toxin genes internal to transposase		CYN toxin genes external to transposase		Housekeeping Genes	
Phe	UUU	178 (1.27)	UUU	77 (1.00)	UUU	12(1.50)
Leu	UUG	227 (1.64)	UUG	138 (1.68)	CUG	16 (1.71)
iLeu	AUU	178 (1.20)	AUA	81 (1.04)	AUU	10 (1.50)
Val	GUG	140 (1.20)	GUG	92 (1.28)	GUG	8.0 (1.68)
Ser	UCU	122 (1.22)	UCA	97 (1.45)	AGU	8.0 (1.50)
Pro	CCA	140 (1.26)	CCA	73 (1.13)	CCA	12 (1.60)
Thr	ACA	130 (1.35)	ACA	62 (1.08)	ACG	9 (1.50)
Ala	GCA	148 (1.11)	GCU	57 (1.40)	GCU /GCA	7 (1.22)
Tyr	UAU	137 (1.30)	UAC	77 (1.05)	UAU	7 (1.17)
His	CAU	110 (1.06)	CAC	71 (1.05)	CAU	9(1.38)
Gln	CAA	226 (1.20)	CAG	108 (1.11)	CAG	6(1.20)
Asn	AAU	167 (1.18)	AAC	104 (1.14)	AAU	10 (1.33)
Lys	AAA	179 (1.01)	AAA	131(1.03)	AAA	10 (1.25)
Asp	GAU	176 (1.31)	GAC	42 (1.33)	GAU	4 (1.33)
Glu	GAA	207 (1.16)	GAG	54 (1.21)	GAA	1 (2.00)
Cys	UGC	94 (1.11)	UGC	96 (1.20)	UGC	10 (1.05)
Arg	CGC	91 (1.20)	AGA	84 (1.56)	AGA	26 (3.25)
Gly	GGU	113 (1.06)	GGG	66 (1.44)	GGA	9.0 (1.57)

6.3.3 Structural comparison between *Aphanizomenon ovalisporum* and *Cylindrospermopsis raciborskii* CYN biosynthetic cluster

A plasmid construct incorporating a genomic fragment from a cylindrospermopsin producing *Aphanizomenon ovalisporum* was sequenced and revealed three open reading frames *(orf)* annotated *aoaA* (amindinotransferase), *aoaB* (a chimeric PKS/NRPS) and *aoaC* (PKS) (Shalev-Alon et al. 2002). The genes were implicated in the biosynthesis of CYN and their organisation is outlined in Figure 6.9A. Organisation of homologous CYN biosynthetic genes, identified from the genome of *Cylindrospermopsis raciborskii* strain AWT205 (presented in this thesis) demonstrated a more uni-linear organisation, (Figure 6.9B). The *orf* pertaining to *C. raciborskii* homologues, *cynC*, *cynD* and *cynA* appeared to be transcribed as a single unit and a homologue of the *cynA* gene was not identified in *Aph. ovalisporum.*

Investigating the structural difference between *C. raciborskii* and *Aph. ovalisporum* CYN biosynthetic clusters was attempted on two strains of an Australian *Aph. Ovalisporum* (APH 031G and APH 031E) utilising long template PCR. Firstly, an attempt was directed to ascertain whether the operon (*cynC*, *cynD* and *cynA*) was also evident in the *Aph. ovalisporum*. Gene-specific primers designed from the sequence obtained from *C. raciborskii* AWT205, flanking the AT domains in the *cynD* (*aoaB*) to *cynA* genes were utilised with a predicted PCR amplification product of 7.3 KB. A positive control fragment of 5 KB was designed spanning regions homologous to the AT domain within *cynD* (*aoaB*) and the amidinotransferase gene *cynC* (*aoaA*). A 7.3 KB PCR product was amplified from both *Aph. ovalisporum* strains (Figure 6.10).

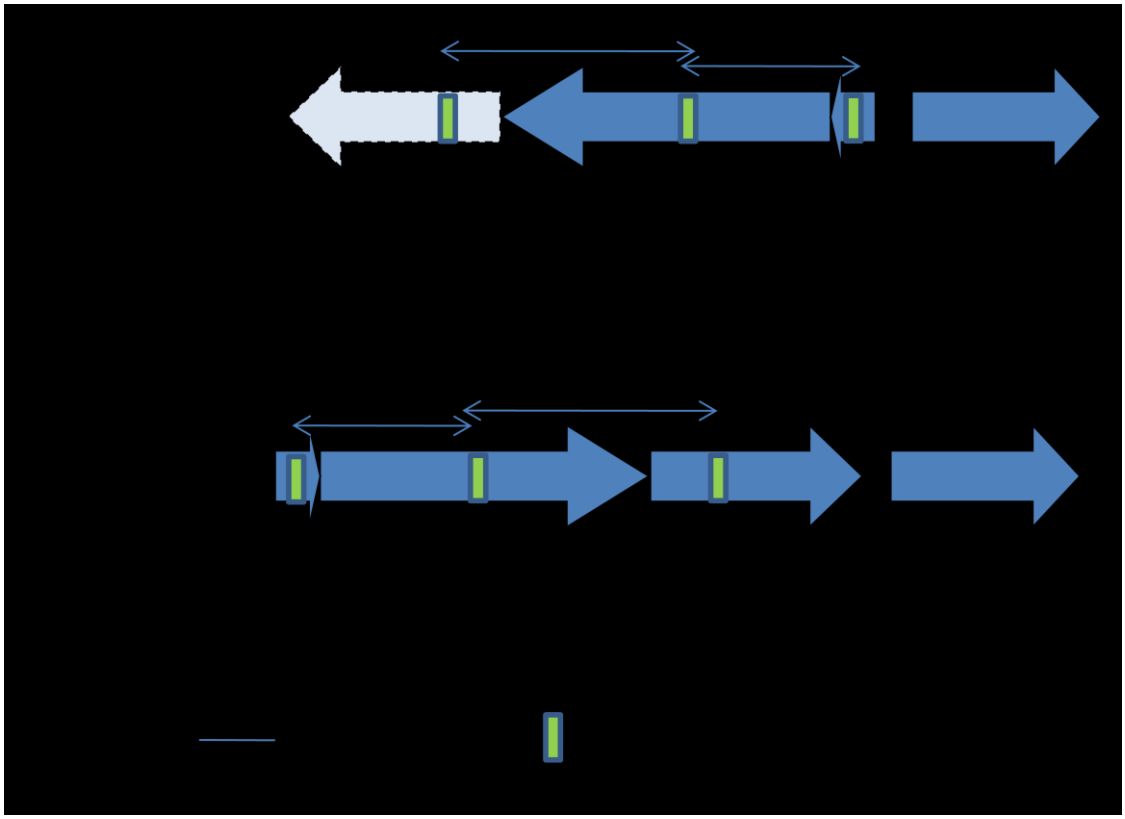


Figure 6.9 Structural organisation of a region of a CYN biosynthetic cluster from *Aph. ovalisporum* ILC-146 (A) and *C. raciborskii* AWT 205 (B).

Dashed lines indicates unknown gene. Predicted fragments sizes of amplified regions are indicated

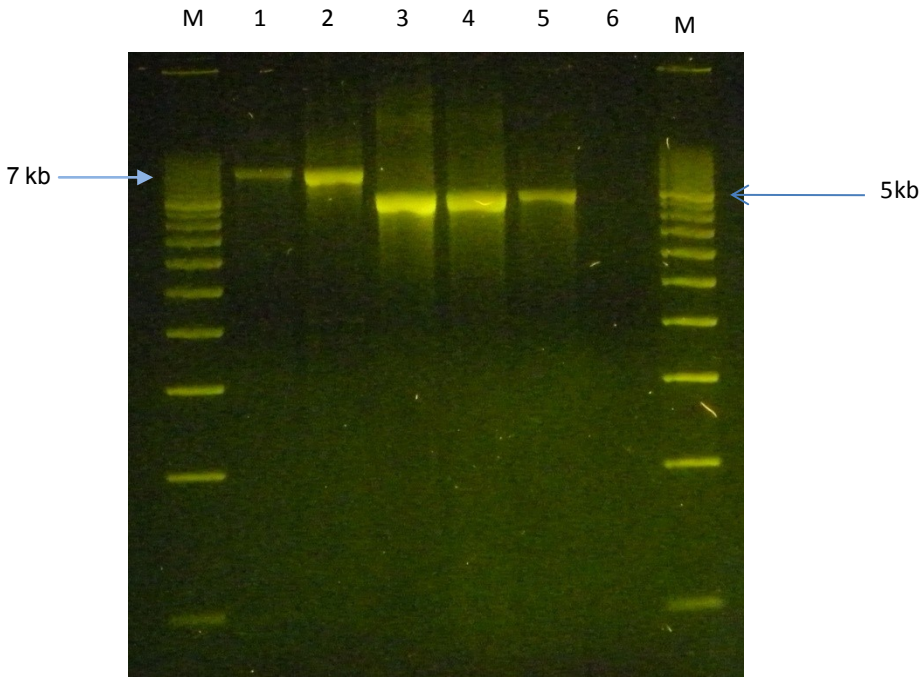


Figure 6.10 PCR amplification of the genomic region spanning the *cynD* (*aoaB*) and *cynA* acyltransferase (AT) domains.

Lane 1 - *Aph. ovalisporum* APH 031G; Lane 2 - *Aph. ovalisporum* APH 031E; the genomic region spanning the *cynD* (*aoaB*) and *cynC* (*aoaA*) as a positive control, Lane 3 - *Aph. ovalisporum* APH31G;Lane 4 - *Aph. ovalisporum* APH31E ; Lane 5 - *C. raciborskii* AWT 205; Lane 6 - negative control; Lanes - M 500 bp molecular weight marker (Geneworks).

Additional long template PCR reactions were performed in an effort to ascertain the structural organisation of other genes identified in the *C. raciborskii* relative to the *Aph. ovalisporum* cluster. Specific oligonucleotide primers (Chapter 3, Table 3.5) designed to amplify from within the putative aminotransferase (*cynJ*) and amidohydrolase (*cynF*) genes were used in combination with primers directed to the AT domain of *cynB* and *cynA* (Figure 6.12). However, there was no success in amplifying a product from any of the combinations attempted and therefore the chromosomal location of *cynJ* or *cynF* in *Aph. ovalisporum* remains unclear.

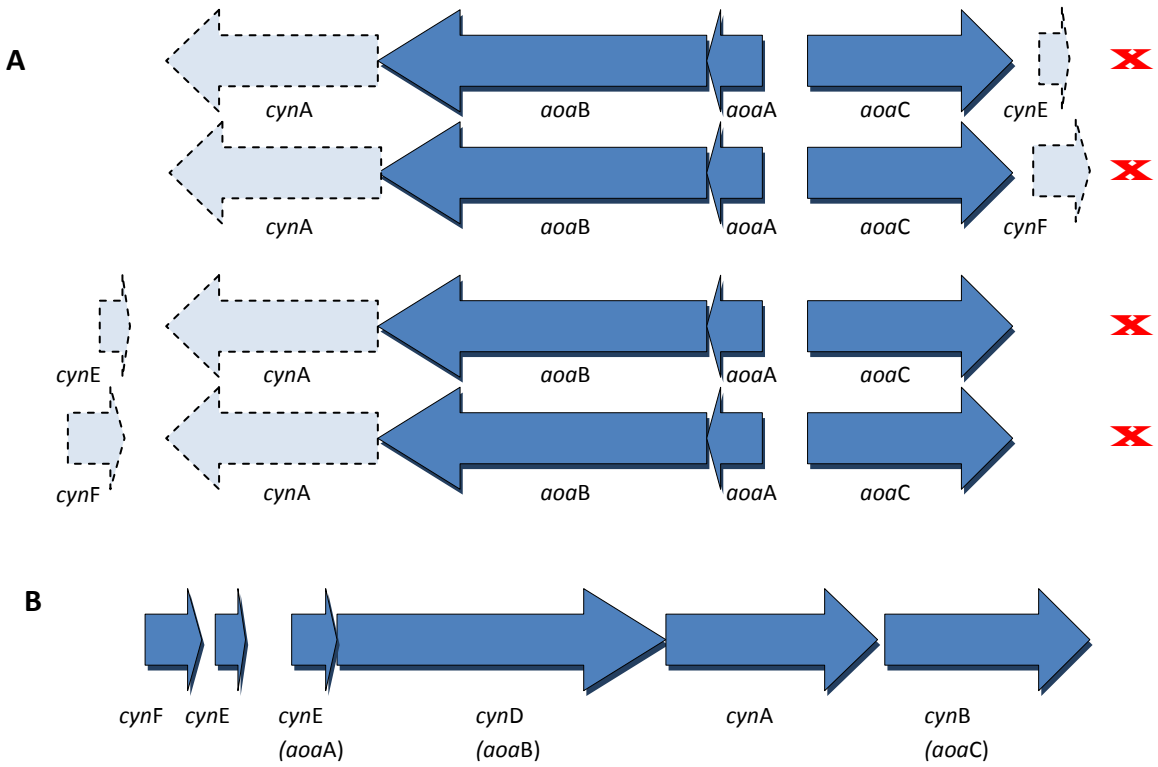


Figure 6.11 Structural organisation of a region of a CYN biosynthetic cluster (in solid) for *Aph. ovalisporum* ILC-146 (A) and *C. raciborskii* AWT 205 (B).

Dashed lines indicates unknown gene. Red cross indicates unsuccessful, long-templated PCR efforts to determine gene organisation in Aph. ovalisporum.

Therefore, it appears that the architecture of the CYN biosynthetic cluster has undergone at least one apparent recombination event involving the *cynB* homologue, *aoaC* within the *Aph. ovalisporum* genome. Intragenic recombination is considered prolific within the microcystin biosynthetic cluster (Tanabe et al. 2004), and has played an important role in engineering novel variants of the compound . To date, there have only been two other structural variants of CYN, deoxy- CYN (Norris et al. 1999) and 7-epi-CYN (Banker et al. 2000) identified. Whether these are a result of reengineering the CYN biosynthetic cluster or a product of external processes post synthesis of CYN is yet to be elucidated.

6.4 Summary

1. In this study, congruency was demonstrated between the phylogeny of primary metabolism associated genes (16S rRNA (Figure 6.3) and rpoC1 (Figure 6.4)), with that of the 4 CYN-toxic genes (*cynC*, *cynD*, *cynA* and *cynB* (Figures 6.5-6.8)),

isolated from known CYN producing cyanobacteria. This is suggestive that the CYN gene cluster was found in a common ancestor of filamentous, heterocystous cyanobacteria of the order Nostocales. Without the availability of the Oscillatoriales, *Lyngbya wollei* sequence it is impossible to determine whether the common ancestor encompasses all filamentous cyanobacteria.

2. Further data supporting the premise that the CYN biosynthetic cluster was an ancient ancestral trait were:
    - a patchy distribution of the CYN producing cyanobacterial lineages through the 16S rRNA gene tree (Figure 6.2)
    - the coincidental loss of CYN toxicity by *A. bergii* ANA283A during the course of this study, determined by both HPLC and lack of toxin gene targets amplified during PCR
    - no evidence of codon bias between primary metabolism and CYN toxin genes within the *C. raciborskii* AWT205 genome (Table 6.3).
3. Positive selection pressures were predicated by determining the dN/dS ratios for the *A. lapponica* isolates *cynB* homologue. This is suggestive that the cluster's highly divergent sequence is a consequence of adaptive evolutionary influences, which may facilitate the organisms' functioning at colder water temperatures.
4. There was at least one incident of intragenic recombination of chromosomal gene organisation of the CYN biosynthetic cluster for the *Aph. ovalisporum*. Evidence of gene reorganisation may have implications in the evolutionary history of the CYN gene cluster and the engineering of novel compounds.
5. It appears as if the CYN gene cluster is slowly being lost from the cyanobacterial lineages. This is does not discount novel CYN-producers still being identifed in the future but suggests the promicious uptake of DNA fragments containing the CYN gene cluster is not a significant occurence.

7 GENERAL DISCUSSION

Cylindrospermopsin is a potent secondary metabolite produced by numerous freshwater cyanobacterial species. The presence of CYN and CYN-producing cyanobacteria in global water bodies poses significant risks to the consumers, as the toxin has been implicated in acute toxicity episodes (Hawkins et al. 1985; Saker et al. 1999). Further to this, preliminary evidence suggests long-term exposure may lead to tumour promotion (Falconer and Humpage 2001). Currently, water management practices rely on routine monitoring of intensively utilised water bodies. Identification and enumeration using microscopy is employed to identify problematic cyanobacterial populations; however it is recognised that phenotypically, toxin and non toxin producing strains are indistinguishable. Where warranted (cyanobacterial cell counts > 2000 cells/mL), additional analytical chemistry-based methods are employed to determine the presence and concentrations of cyanotoxins. Whilst the current methodology is generally effective, it can be time consuming, laborious and expensive. Often the determination of a toxin problem can be delayed whilst awaiting results; meanwhile the bloom population and toxin potential can escalate and reactive resolution of the situation may be necessary. The consequential economic losses due to cyanobacterial blooms can be wide-ranging, therefore there is a need for rapid detection and management, to minimise the time affected water bodies are unavailable for the reticulated supply.

Due to increasing eutrophication of surface waters and global warming impacts, the occurrences of cyanobacterial blooms are escalating in their prevalence. Further to this identification of novel CYN-producing species appear more frequent (Preussel et al. 2006; Spoof et al. 2006; Seifert et al. 2007), perhaps reflective of more intensive screening activity. To aid water quality managers to rapidly identify problematic cyanobacterial populations, and thus ensure appropriate measures are initiated early to guarantee public health, molecular based approaches are proving attractive alternatives to current detection strategies. Molecular probes can be designed to be specific, sensitive and lend themselves to greater portable and rapid technology platforms. However, prior to the commencing this work there was a paucity of molecular data available for many of the biosynthetic mechanisms involved for toxic cyanobacterial secondary metabolites.

At the commencement of this thesis, the molecular machinery implicated in CYN production had only been partially elucidated (Schembri et al. 2001; Shalev-Alon et al. 2002; Fergusson 2003). The molecular knowledge available had led to the development of molecular assays to permit rapid detection of CYN-producing cyanobacteria and more specifically toward identification of the prevalent producer in Australia, *C. raciborskii* (Fergusson and Saint 2003; Rasmussen et al. 2008b). However, without full consideration of all the genes involved in CYN biogenesis, it is perilous to assume the most robust determinant of CYN toxicity has been chosen. To resolve this knowledge gap, Chapter 3 of this thesis detailed the sequencing strategies undertaken in an effort to fully resolve the biosynthetic pathway.

The sequencing (Chapter 3) and bioinformatic analysis (Chapter 4) of a 44.2 KB region, revealed an organisation of the cylindrospermopsin assembly (43 KB), as a predominantly polyketide pathway with an integrated peptide synthetase, which featured a novel adenylation domain predicted to upload a guandinoacetic acid starter unit. Also represented in the pathway are a number of unusual structural and enzymatic features. These include two genes (*cynI* and *cynF*) that encode homodimeric polypeptides with putative involvement toward the unique cyclisation of the uracil moiety and *cynE*, which is putatively implicated in the sulphonation of the final compound. Recognition of unique genes within the cluster, promotes the possibility of identifying an ideal molecular marker that permits discrimination of CYN-producing cyanobacteria in environmental waters.

Gene-directed PCR was utilised in Chapter 3 to achieve two specific aims, firstly to identify genes that were correlated to CYN toxicity and secondly to determine which molecular markers were unambiguous in their detection toward CYN-producing cyanobacteria. Previous studies targeting the *pks* (*cynB*) and *ps* (*cynD*) suggested these genes were robust markers for the detection of CYN-producers. However, findings from Chapter 3 highlighted that false negatives threaten the long term potential of these molecular assays. Further to this the multiplex assay when trialled on the German *Aphanizomenon* CYN-producing species, failed to detect the *pks* determinant, however the *ps* region was readily detected (Preussel et al. 2006). The prospect of failed detection may have considerable impact in regards to water quality management and public safety issues. The findings from Chapter 3 and observations by Preussel et al (2006) demonstrates the importance of identifying unambiguous toxicity determinants to promote confidence in molecular based assays and perhaps suggest that relying on a single molecular target for detection may be risky.

The overall investigations by gene-directed PCR in Chapter 3, indicated that the AMT primers, directed toward the *cynE* encoding a putative sulfontransferase, appeared most promising as molecular probe candidates toward CYN detection across genera. These probes robustly detected all CYN producers, even the recently isolated *A. lapponica*, which proved elusive toward PCR detection utilising the majority of molecular probes targeting CYN genes designed within this study. This was also true for the benthic *L. wollei*. The AMT probes will need to be further challenged with a range of bacterial and cyanobacterial extracts from environmental waters to ultimately determine its reliability as definitive CYN toxic determinants. However thus far, the *cynE* region appears to be a robust CYN determinant, and is recommended for inclusion in future molecular based assays.

Ultimately, characterisation of the CYN biosynthetic cluster is desirable from more than one genus. This will facilitate the detailed description of all genes implicated in CYN biosynthesis across a broad range of species and permit increased confidence in designing not only CYN-specific molecular probes but also species-specific probes. To date, the German *Aphanizomenon* sp. CYN biosynthetic cluster has been fully sequenced and annotated and is soon to be released to the wider scientific community (A. Stüken, personal communication). However, of particular interest would be the elucidation of the CYN cluster from *A. lapponica* and *L. wollei*, as they appear to be distinctive in their coding preferences compared to all other CYN-producing cyanobacteria.

Whilst evidence of natural deletion mutants is perhaps sufficient to implicate the gene cluster in the biosynthesis of CYN, targeted mutagenesis may still be desirable. Only then will it be possible to alleviate any doubts that persist about the biosynthetic role of the cluster and its association with CYN-producing cyanobacteria. In addition, targeted mutagenesis experiments will permit further clarification of several of the ORFs, whose predicated biochemical roles are still ambiguous in respect to CYN production. Of particular interest is the polypeptide encoded by *cynE*, whose hypothetical biochemical role is predicted to be a sulfotransferase. Interestingly, the sulphated end-products of sulfotranserases in other organisms, are involved in a range of biological processes including cellular signalling, hormone regulation, and molecular recognition; perhaps functional studies of this gene can elucidate the importance of cylindrospermopsin to CYN-producing cyanobacteria. In comparison to the 60 identified variants of microcystin, cylindrospermopsin has to date

demonstrated limited analogous structures. This is suggestive that the utility of cylindrospermopsin to the cyanobacteria that produce it must be specific in its role.

Overall, the sequencing (Chapter 3) and preliminary analysis (Chapter 4) of a putative CYN gene cluster, has provided the basic information essential for the future development of robust DNA or even RNA-based detection systems for cylindrospermopsin-producing cyanobacteria. Furthermore, access to this genetic information enables the biosynthesis of cylindrospermopsin to be finally studied holistically. The availability of the sequence allows for the identification and characterisation of gene promoters, facilitates expression studies and understanding the mechanisms of regulation for toxin production. Thus far, investigations into promoter regions associated with CYN production have been detailed for *Aph. ovalisporum*, these investigations may prove constructive with directing similar investigations in CYN–producing *C. raciborskii* (Shalev-Malul et al. 2008; Lieman-Hurwitz et al. 2009). Finally, a better understanding of the molecular mechanics of CYN production could be very important in predicting which environmental conditions promote or reduce toxin production. Investigations in Chapter 5 were directed toward beginning to understand this process in a bloom setting.

The tracking of a CYN-producing *C. raciborskii* bloom in the lower River Murray (Chapter 5) permitted novel insights into the molecular mechanics of bloom-forming *C. raciborskii* and the CYN toxin potential through the blooms time course. Interestingly, the toxic genotype appeared to dominate in the preliminary stages of the bloom with an observed shift to the non-toxic genotype, during stationary growth. Further time course experiments are necessary to strengthen these observations and eventually may find utility in arming water quality managers with a quantifiable tool of a blooms toxic potential. This may assit in mangement decisons with respect to efficiently dosing algal blooms with copper sulphate dosages or guiding water treatment plants with activated carbon requirements, minimising environmental consequences and costs. Recent methods in rapid preparation of environmental samples, support the future of qPCR methodologies being deployed for onsite analysis (Rasmussen et al. 2008a). However, refinement of qPCR is required before it can truly be utilised on a routine monitoring basis. There is a need to overcome PCR inhibition and develop robust internal controls. A better understanding of replication of the genome, cell cycle and ribosomal copy numbers in filamentous cyanobacteria is desirable, which may resolve some of the discrepancies observed between cyanobacterial cell counts and gene copy

numbers. Further to this, whilst the qPCR developed by Rasmussen at al (2008) is able to detect CYN-producing *Cylindrospermopsis*, well below the alert levels outlined by the ANZDWG framework, it would be desirable to improve the sensitivity toward detection to facilitate understanding of CYN-toxic bloom formation from as close to the initial seed as possible.

The shifts in genotypic populations over the bloom's time course, has similarly been observed in microcystin-producing cyanobacterial populations. These comparable findings may contribute to the resolution of the key importance these compounds play amongst cyanobacterial communities. The preliminary findings detailed in Chapter 5, appear suggestive that CYN may support founding populations, to establish a community by assisting acclimation to altered environmental conditions, or exists as a deterrent to zooplankton grazing. Toward the peak of the bloom a shift toward the non toxic genotype suggests CYN-producing cyanobacteria may represent weaker competitors for the available resources and the metabolic cost of producing the secondary metabolites may prove too expensive. Similar hypotheses exist to explain the existence of microcystin (Jungmann and Benndorf 1994) and non-microcystin producing strains have demonstrated a competitive edge when exposed to altered light conditions (Kardinaal et al. 2007). Further studies comparing growth rates between toxic and non-toxic *C. raciborskii* isolates, grown axenically, under identical conditions may help clarify the observations in genotype succession.

The reporting of the *C. raciborskii* bloom in Chapter 5 represented the first significant mass occurrence of *C. raciborskii* in this region. Whilst it is easy to explain the origin of *C. raciborskii* in the River Murray, questions arise with regard to the recent success of the population. Indeed, the success of *C. raciborskii* is a worldwide phenomenon that has been generating intense investigation in an effort to understand the attributes this species may hold to assist its global proliferation (Padisák 1997; McGregor and Fabbro 2000; Neilan et al. 2003; Briand et al. 2004; Chonudomkul et al. 2004; Gugger et al. 2005). Is the recent success due to the existence of an eco-physically adapted strain to temperate waters? Or is it solely because *C. raciborskii* has finally found the desirable environment to thrive in the River Murray due to altered environmental conditions, contributed by the recent slow flows and rising global temperatures? The isolation of a *C. raciborskii* isolate, CYN43 during the work undertaken in Chapter 5, presents the opportunity to explore some of these possibilities in greater detail.

Other questions present themselves with respect to the long-term consequences for the River Murray. Will the population continue to grow and became entrenched as a dominant feature of the cyanobacterial community, thus contributing to a shift in toxin composition of temperate Murray River waters? And in doing so, will, mass occurrences of CYN-producing *C. raciborskii* become a common water quality management issue for water authorities of the lower Murray? An additional consideration lies with the positive relationship observed and presented in Chapter 5, between *C. raciborskii* and putative MIB -producing *Pseudanabaena* sp. If this demonstrates to be more than just a circumstantial observation, it appears that water quality managers may face additional challenges on both public health and an aesthetic prospective. Of course, more detailed temporal and spatial ecological studies are required to form any solid basis for the synergy of the relationship between the two cyanobacterial species.

Finally, the molecular data acquired and presented in this thesis, was used to investigate the evolutionary history of the CYN biosynthetic cluster in the cyanobacterial genome (Chapter 6). Previous findings nominated horizontal gene transfer (HGT) as the likely distributor of the CYN cluster between genera (Kellmann et al. 2006). However, findings detailed in this thesis suggested that the CYN gene cluster was found in a common ancestor of filamentous, heterocystous cyanobacteria of the order Nostocales. There is also evidence that the cluster is being gradually lost from derived lineages, though this does not discount the potential novel CYN-producing cyanobacteria. Intragenic recombination of chromosomal gene organisation of the CYN biosynthetic cluster is evident for *Aph. ovalisporum*. Evidence of gene reorganisation may have implications in the evolutionary history of the CYN gene cluster and the engineering of novel compounds. The extent of this recombination will be fully recognised when the organisation of the CYN biosynthetic cluster is characterised from other genera. Examination of the nucleotide sequence obtained from *C. raciborskii* AWT205 reveals the presence of significant repetitive sequences; in *cynE* alone there exists a 43 bp direct tandem repeat. Whilst nucleotide repeats can be attributed to secondary structures, they also are recognised as contributing to the plasticity of cyanobacterial genomes.

In conclusion, at the commencement of this work, there was a lack of knowledge about the molecular machinery for the majority of cyanobacterial toxc secondary metabolites rountinely found in global water bodies. Since then the saxitoxin biosynthetic cluster has

been identifed in three distinct genra (Kellmann et al. 2008a; Mihali et al. 2009) and the anatoxin-a biosynthetic cluster has also been sequenced (Cadel-Six et al. 2009). The work presented in this thesis has provided the basic information for the biosynthesis of CYN. Together these findings herald a new era in cyanobacterial research. In particular, the molecular data from this work enhances the development of robust and specific identification of CYN-toxigenic species. Knowledge of the molecular architecture unleashes the inherent information this molecular data contains to contribute to a greater understanding of expression and regulation of the toxin. This will find a future in enabling more proactive approaches to bloom management, as environmental conditions that promote or reduce toxin production are determined. The information presented in this thesis will lead to advancement in effective water management strategies to reduce public health risks, posed by the presence of CYN-producing cyanobacteria.

Bibliography

Asato, Y. (1979). "Macromolecular synthesis in synchronized cultures of *Anacystis nidulans*." J Bacteriol **140**(1): 65-72.

Avise, J. C. (1994). Molecular Markers, Natural History and Evolution. New York, Chapman & Hall.

Baker, J. A., et al. (2001). "Identification of cyanobacteria and their toxigenicity in environmental samples by rapid molecular analysis." Environmental Toxicology **16**: 472-482.

Baker, P. D. (1996). Occurrence of *Cylindrospermopsis* in South-Eastern Australia. Cylindrospermopsis Symposium - A New Toxic Algal Bloom Challenge for Australia, South-East Queensland Water Board/Brisbane City Council/ARMCANZ, Brisbane.

Baker, P. D. and A. R. Humpage (1994). "Toxicity associated with commonly occurring Cyanobacteria in surface waters of the Murray-Darling basin, Australia." Australian Journal of Marine and Freshwater Research **45**: 773-786.

Banker, R. and S. Carmeli (1997). "Identification of cylindrospermopsin in *Aphanizomenon ovalisporum* (Cyanophyceae) isolated from Lake Kinneret, Israel." Journal of Phycology **33**: 613-616.

Banker, R., et al. (2000). "7-Epicylindrospermopsin, a toxic minor metabolite of the cyanobacterium *Aphanizomenon ovalisporum* from lake Kinneret, Israel." J Nat Prod **63**(3): 387-9.

Bedekar, A., et al. (1998). "The comparative amino acid sequences, substrate specificities and gene or cDNA nucleotide sequences of some prokaryote and eukaryote amidinotransferases: implications for evolution." Comp Biochem Physiol B Biochem Mol Biol **119**(4): 677-90.

Berry, J. P., et al. (2008). "Cyanobacterial toxins as allelochemicals with potential applications as algaecides, herbicides and insecticides." Mar Drugs **6**(2): 117-46.

Binder, B. J. and S. W. Chisholm (1995). "Cell cycle regulation in marine *Synechococcus* sp. strains." Appl Environ Microbiol **61**(2): 708-717.

Bindseil, K. U. and A. Zeeck (1993). "Prelactones C and B, oligoketides from *Streptomyces* producing concanamycins and bafilomycins." Helvetica Chimica Acta **76**: 150-157.

Blackburn, S., et al. (1997). Cyanobacterial blooms: why are they toxic? (The regulation of toxin production and persistence). Managing Algal Blooms. J. R. Davis. Canberra, CSIRO Land and Water**:** 67-77.

Bonfils, J., et al. (2000). "Sheep cytosolic branched-chain amino acid aminotransferase: cDNA cloning, primary structure and molecular modelling and its unique expression in muscles." Biochim Biophys Acta **1494**(1-2): 129-36.

Boucher, Y., et al. (2003). "Lateral gene transfer and the origins of prokaryotic groups." Annu Rev Genet **37**: 283-328.

Bourke, A. T. C., et al. (1983). "An outbreak of hepatoenteritis (the Palm Island mystery disease) possibly caused by algal intoxication." Toxicon **3**: 45-48.

Bouvy, M., et al. (1999). "Dynamics of a toxic cyanobacterial bloom (*Cylindrospermopsis raciborskii*) in a shallow reservoir in the semi-arid region of northeast Brazil." Aquatic Microbial Ecology **20**(3): 285-297.

Bouvy, M., et al. (2001). "Effects of a cyanobacterial bloom (*Cylindrospermopsis raciborskii*) on bacteria and zooplankton communities in Ingazeira reservoir (northeast Brazil)." Aquatic Microbial Ecology **25**: 215-227.

Boyer, S. L., et al. (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.

Briand, E., et al. (2009). "Spatiotemporal changes in the genetic diversity of a bloom-forming *Microcystis aeruginosa* (cyanobacteria) population." Isme J **3**(4): 419-29.

Briand, E., et al. (2008). "Temporal variations in the dynamics of potentially microcystin-producing strains in a bloom-forming *Planktothrix agardhii* (Cyanobacterium) population." Appl Environ Microbiol **74**(12): 3839-48.

Briand, J. F., et al. (2004). "*Cylindrospermopsis raciborskii* (Cyanobacteria) invasion at mid-latitudes: selection, wide physiological tolerance, or global warming?" J.Phycol **40**: 231-238.

Briand, J. F., et al. (2002). "Environmental context of *Cylindrospermopsis raciborskii* (Cyanobacteria) blooms in a shallow pond in France." Water Res **36**(13): 3183-92.

Brown, M. H., et al. (1999). "The multidrug efflux protein NorM is a prototype of a new family of transporters." Mol Microbiol **31**(1): 394-5.

Bruno, M., et al. (1994). "Anatoxin-A and a previously unknown toxin in *Anabaena planctonica* from blooms found in Lake Mulargia (Italy)." Toxicon **32**(3): 369-73.

Burch, M. D. (2008). "Effective doses, guidelines & regulations." Adv Exp Med Biol **619**: 831-53.

Burgoyne, D. L., et al. (2000). "Biosynthesis of cylindrospermopsin." Journal of Organic Chemistry **65**: 152-156.

Burja, A. M., et al. (2001). "Marine cyanobacteria- a prolific source of natural products." Tetrahedron **57**: 9347-9377.

Burton, R. A., et al. (2004). "The CesA gene family of barley. quantitative analysis of transcripts reveals two groups of co-expressed genes." Plant Physiol **134**(1): 224-36.

Butler, A. R., et al. (1999). "Impact of thioesterase activity on tylosin biosynthesis in *Streptomyces fradiae*." Chem Biol **6**(5): 287-92.

Byth, S. (1980). "Palm Island mystery disease." Medical Journal of Australia **2**: 40-42.

Cadel-Six, S., et al. (2009). "Identification of a polyketide synthase coding sequence specific for anatoxin-a-producing *Oscillatoria* cyanobacteria." Appl Environ Microbiol **75**(14): 4909-12.

Caffrey, P. (2003). "Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases." Chembiochem **4**(7): 654-7.

Cai, Y. P. and C. P. Wolk (1990). "Use of a conditionally lethal gene in *Anabaena* sp. strain PCC 7120 to select for double recombinants and to entrap insertion sequences." J Bacteriol **172**(6): 3138-45.

Cane, D. E. and C. T. Walsh (1999). "The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases." Chem Biol **6**(12): R319-25.

Cane, D. E., et al. (1998). "Harnessing the biosynthetic code: combinations, permutations, and mutations." Science **282**(5386): 63-8.

Carmichael, W. W. (1992). "Cyanobacterial secondary metabolites - the cyanotoxins." Journal of Applied Bacteriology **72**: 445-459.

Carmichael, W. W. (2001). "Health effects of toxin-producing cyanobacteria: "The CyanoHABs"." Human and Ecological Risk Assessment **7**(5): 1393-1407.

Carmichael, W. W. and J. An (1999). "Using an enzyme linked immunosorbent assay (ELISA) and a protein phosphatase inhibition assay (PPIA) for the detection of microcystins and nodularins." Nat Toxins **7**(6): 377-85.

Challis, G. L., et al. (2000). "Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains." Chem Biol **7**(3): 211-24.
Chang, Z., et al. (2002). "The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit." Gene **296**(1-2): 235-47.
Chang, Z., et al. (2004). "Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **67**(8): 1356-67.
Chapman, E., et al. (2004). "Sulfotransferases: structure, mechanism, biological activity, inhibition, and synthetic utility." Angew Chem Int Ed Engl **43**(27): 3526-48.
Cheng, Y. Q., et al. (1999). "A putative branched-chain-amino-acid transaminase gene required for HC-toxin biosynthesis and pathogenicity in *Cochliobolus carbonum*." Microbiology **145 (Pt 12)**: 3539-46.
Cheng, Y. Q., et al. (2003). "Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis." Proc Natl Acad Sci U S A **100**(6): 3149-54.
Chiswell, R. K., et al. (1999). "Stability of cylindrospermopsin, the toxin from the cyanobacterium, *Cylindrospermopsis raciborskii*: Effect of pH, temperature, and sunlight on decomposition." Environmental Toxicology **14**(1): 155-161.
Chonudomkul, D., et al. (2004). "Morphology, genetic diversity, temperature tolerance and toxicity of *Cylindrospermopsis raciborskii* (Nostocales, Cyanobacteria) strains from Thailand and Japan." FEMS Microbiology Ecology **48**: 345–355.
Chorus, I. and J. Bartram (1999). Toxic cyanobacteria in water. A guide to their public health consequences, monitoring and management. London, E & FN Spon.
Chorus, I., et al. (2000). "Health risks caused by freshwater cyanobacteria in recreational waters." Journal of Toxicology and Environmental Health. Part B. Critical Reviews **3**(4): 323-347.
Christiansen, G., et al. (2003). "Microcystin biosynthesis in *Planktothrix*: genes, evolution, and manipulation." Journal of Bacteriology **185**(2): 564-572.
Christiansen, G., et al. (2006). "Transposons inactivate biosynthesis of the nonribosomal peptide microcystin in naturally occurring *Planktothrix* spp." Appl Environ Microbiol **72**(1): 117-23.
Codd, G. A., et al. (1994). "Toxic Blooms of Cyanobacteria in Lake Alexandrina, South Australia-Learning from History." Aust. J. Mar. Freshwater Res. **45**: 731-736.
Cohen, M. F., et al. (1998). "Transposon Mutagenesis of Heterocyst-Forming Filamentous Cyanobacteria." Methods in Enzymology **297**.
Cortes, J., et al. (1990). "An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea." Nature **348**(6297): 176-8.
Cortes, J., et al. (1995). "Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage." Science **268**(5216): 1487-9.
Costa, I. A., et al. (2006). "Occurrence of toxin-producing cyanobacteria blooms in a Brazilian semiarid reservoir." Braz J Biol **66**(1B): 211-9.
Csako, G. (2006). "Present and future of rapid and/or high-throughput methods for nucleic acid testing." Clin Chim Acta **363**(1-2): 6-31.
Deckers, H. M. and G. Voordouw (1994). "Identification of a large family of genes for putative chemoreceptor proteins in an ordered library of the *Desulfovibrio vulgaris* Hildenborough genome." J Bacteriol **176**(2): 351-8.

Dittmann, E., et al. (1997). "Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806." Molecular Microbiology **26**(4): 779-787.

Dokulil, M. T. and K. Teubner (2000). "Cyanobacterial dominance in lakes." Hydrobiologia **438**: 1–12.

Donadio, S., et al. (1991). "Modular organization of genes required for complex polyketide biosynthesis." Science **252**: 675-679.

Du, L., et al. (2001). "Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules." Metab Eng **3**(1): 78-95.

Dyble, J., et al. (2002). "Genetic characterization of *Cylindrospermopsis raciborskii* (Cyanobacteria) isolates from diverse geographic origins based on *nifH* and *cpcBA*-IGS nucleotide sequence analysis." Applied and Environmental Microbiology **68**(5): 2567-2571.

Eaglesham, G. K., et al. (1999). "Use of HPLC-MS / MS to monitor cylindrospermopsin, a blue-green algal toxin for public health purposes." Environmental Toxicology **14**(1): 151-154.

Edwards, D. J. and W. H. Gerwick (2004). "Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase." J Am Chem Soc **126**(37): 11432-3.

Edwards, D. J., et al. (2004). "Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*." Chem Biol **11**(6): 817-33.

Eichler, E. E., et al. (2004). "An assessment of the sequence gaps: unfinished business in a finished human genome." Nat Rev Genet **5**(5): 345-54.

Erhard, M., et al. (1999). "Rapid identification of the new anabaenopeptin G from Planktothrix agardhii HUB 011 using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry." Rapid Commun Mass Spectrom **13**(5): 337-43.

Fabbro, L. D. and L. J. Duivenvoorden (1996). "Profile of a bloom of the Cyanobacterium *Cylindrospermpsis raciborskii* (Woloszynska) Seenaya and Subba Raju in the Fitzroy River in Tropical Central Queensland." Mar. Freshwater Res. **47**: 685-694.

Falconer, I. R. (1999). "An overview of problems caused by toxic blue-green algae (cyanobacteria) in drinking and recreational water." Environmental Toxicology **14**(1): 5-12.

Falconer, I. R. (2001). "Toxic cyanobacterial bloom problems in Australian waters: risks and impacts on human health." Phycologia **40**(3): 228-233.

Falconer, I. R. and A. R. Humpage (2001). "Preliminary evidence for *in vivo* tumour initiation by oral administration of extracts of the blue-green alga *Cylindrospermopsis raciborskii* containing the toxin cylindrospermopsin." Environmental Toxicology **16**(2): 192-195.

Falconer, I. R. and A. R. Humpage (2005). "Health risk assessment of cyanobacterial (blue-green algal) toxins in drinking water." Int J Environ Res Public Health **2**(1): 43-50.

Falconer, I. R. and A. R. Humpage (2006). "Cyanobacterial (blue-green algal) toxins in water supplies: Cylindrospermopsins." Environ Toxicol **21**(4): 299-304.

Falconer, I. R., et al. (1988). "Oral toxicity of a bloom of the Cyanobacterium *Microcystis aeruginosa* administered to mice over periods up to 1 year." J Toxicol Environ Health **24**(3): 291-305.

Felsenstein, J. (1988). "Phylogenies from Molecular Sequences: Inference and Reliability." Annual Review of Genetics **22**: 521-565

Fergusson, K. M. (2003). Rapid methods for the detection of toxic cyanobacteria. School of Pharmaceutical, Molecular and Biomedical Sciences. Adelaide, University of South Australia.

Fergusson, K. M. and C. P. Saint (2003). "Multiplex PCR assay for *Cylindrospermopsis raciborskii* and cylindrospermopsin-producing cyanobacteria." Environ Toxicol **18**(2): 120-5.

Fewer, D. P., et al. (2007). "Recurrent adenylation domain replacement in the microcystin synthetase gene cluster." BMC Evol Biol **7**: 183.

Fewer, D. P., et al. (2008). "Natural occurrence of microcystin synthetase deletion mutants capable of producing microcystins in strains of the genus *Anabaena* (Cyanobacteria)." Microbiology **154**(Pt 4): 1007-14.

Finking, R. and M. A. Marahiel (2004). "Biosynthesis of nonribosomal peptides1." Annu Rev Microbiol **58**: 453-88.

Fischer, S. G. and L. S. Lerman (1983). "DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory." Proc Natl Acad Sci U S A **80**(6): 1579-83.

Frangeul, L., et al. (2008). "Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium." BMC Genomics **9**: 274.

Froscio, S. M., et al. (2001). "Cell-free protein synthesis inhibition assay for the cyanobacterial toxin cylindrospermopsin." Environmental Toxicology **16**(5): 408-412.

Garber, M., et al. (2009). "Closing gaps in the human genome using sequencing by synthesis." Genome Biol **10**(6): R60.

Garcia-Vallve, S., et al. (2000). "Horizontal gene transfer in bacterial and archaeal complete genomes." Genome Res **10**(11): 1719-25.

Garner, D. L., et al. (1994). "Dual DNA staining assessment of bovine sperm viability using SYBR-14 and propidium iodide." J Androl **15**(6): 620-9.

Giglio, S., et al. (2008). "Isolation and characterization of the gene associated with geosmin production in cyanobacteria." Environ Sci Technol **42**(21): 8027-32.

Giovannoni, S. J., et al. (1988). "Evolutionary relationships among cyanobacteria and green chloroplasts." Journal of Bacteriology **170**(8): 3584-3592.

Goff, S. A., et al. (2002). "A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. japonica). (cover story)." Science **296**(5565): 92.

Gophna, U., et al. (2006). "Ancient lateral gene transfer in the evolution of *Bdellovibrio bacteriovorus*." Trends Microbiol **14**(2): 64-9.

Gorham, P. R., et al. (1964). "Isolation and culture of toxic strains of *Anabaena flos-aquae* (Lyngb.) deBréb." Verhandlungen der Internationalen Vereinigung fur theoretische und angewandte Limnologie **15**: 796-804.

Grunewald, J. and M. A. Marahiel (2006). "Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides." Microbiol Mol Biol Rev **70**(1): 121-46.

Gugger, M., et al. (2005). "First report in a river in France of the benthic cyanobacterium *Phormidium favosum* producing anatoxin-a associated with dog neurotoxicosis." Toxicon **45**(7): 919-28.

Gugger, M., et al. (2002). "Phylogenetic comparison of the cyanobacterial genera *Anabaena* and *Aphanizomenon*." International Journal of Systematic and Evolutionary Microbiology **52**: 1867-1880.

Gugger, M., et al. (2005). "Genetic diversity of *Cylindrospermopsis* strains (cyanobacteria) isolated from four continents." Appl Environ Microbiol **71**(2): 1097-100.

Gugger, M. F. and L. Hoffmann (2004). "Polyphyly of true branching cyanobacteria (Stigonematales)." Int J Syst Evol Microbiol **54**(Pt 2): 349-57.

Ha, J. H., et al. (2009). "Quantification of toxic *Microcystis* and evaluation of its dominance ratio in blooms using real-time PCR." Environ Sci Technol **43**(3): 812-8.
Haande, S., et al. (2007). "Diversity of *Microcystis aeruginosa* isolates (Chroococcales, Cyanobacteria) from East-African water bodies." Arch Microbiol **188**(1): 15-25.
Hamilton, P. B., et al. (2005). "The Occurrence of the cyanobacterium *Cylindrospermopsis raciborskii* in Constance lake:an exotic cyanoprokaryote new to Canada." Phycologia **44**(1): 17-25.
Harada, K., et al. (1991). "Microcystins from *Anabaena flos-aquae* NRC 525-17." Chem Res Toxicol **4**(5): 535-40.
Harada, K., et al. (1994). "Isolation of cylindrospermopsin from a cyanobacterium *Umezakia natans* and its screening method." Toxicon **32**(1): 73-84.
Harvey, B. M., et al. (2006). "Evidence that a novel thioesterase is responsible for polyketide chain release during biosynthesis of the polyether ionophore monensin." Chembiochem **7**(9): 1435-42.
Havens, K. E. (2008). "Cyanobacteria blooms: effects on aquatic ecosystems." Adv Exp Med Biol **619**: 733-47.
Hawkins, P. R., et al. (1997). "Isolation and toxicity of *Cylindrospermopsis raciborskii* from an ornamental lake." Toxicon **35**(3): 341-346.
Hawkins, P. R., et al. (2001). "Phenotypical variation in a toxic strain of the phytoplankter, *Cylindrospermopsis raciborskii* (Nostocales, Cyanophyceae) during batch culture." Environ Toxicol **16**(6): 460-7.
Hawkins, P. R., et al. (1985). "Severe hepatotoxicity caused by the tropical cyanobacterium (Blue-Green Alga) *Cylindrospermopsis raciborskii* (Woloszynska) Seenaya and Subba Raju isolated from a domestic water supply reservoir." Applied and Environmental Microbiology **50**(5): 1292-1295.
Haydock, S. F., et al. (1995). "Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases." FEBS Lett **374**(2): 246-8.
Heintzelman, G. R., et al. (2002). "Stereoselective total syntheses and reassignment of stereochemistry of the freshwater cyanobacterial hepatotoxins cylindrospermopsin and 7-epicylindrospermopsin." J Am Chem Soc **124**(15): 3939-45.
Hopwood, D. A. and D. H. Sherman (1990). "Molecular genetics of polyketides and its comparison to fatty acid biosynthesis." Annual Review of Genetics **24**: 37-66.
Humpage, A. (2008). "Toxin types, toxicokinetics and toxicodynamics." Adv Exp Med Biol **619**: 383-415.
Humpage, A. R. and I. R. Falconer (2003). "Oral toxicity of the cyanobacterial toxin cylindrospermopsin in male Swiss albino mice: determination of no observed adverse effect level for deriving a drinking water guideline value." Environ Toxicol **18**(2): 94-103.
Humpage, A. R., et al. (2000). "Micronucleus induction and chromosome loss in transformed human white cells indicate clastogenic and aneugenic action of the cyanobacterial toxin, cylindrospermopsin." Mutation Research **472**(1-2): 155-161.
Humpage, A. R., et al. (1994). "Paralytic shellfish poisons from Australian cyanobacterial blooms." Australian Journal of Marine and Freshwater Research **45**: 761-771.
Hutson, S. (2001). "Structure and function of branched chain aminotransferases." Prog Nucleic Acid Res Mol Biol **70**: 175-206.
Hyenstrand, P., et al. (1998). "Factors determining cyanobacterial success in aquatic systems - a literature review." Archiv für Hydrobiologie **51**: 41-62.

Ikeda, H., et al. (1999). "Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*." Proc Natl Acad Sci, USA **96**: 9509-9514.

Isvanovics, V., et al. (2000). "Growth and phosphate uptake kinetics of the cyanobacterium, *Cylindrospermopsis raciborskii* (Cyanophyceae) in throughflow cultures." Freshwater biology **43**: 257-275.

Iteman, I., et al. (2000). "Comparison of conserved structural and regulatory domains within divergent 16S rRNA-23S rRNA spacer sequences of cyanobacteria." Microbiology **146**: 1275-1286.

Izaguirre, G., et al. (2007). "Benthic cyanobacteria (Oscillatoriaceae) that produce microcystin-LR, isolated from four reservoirs in southern California." Water Res **41**(2): 492-8.

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

Janse, I., et al. (2003). "High-resolution differentiation of cyanobacteria by using rRNA-internal transcribed spacer denaturing gradient gel electrophoresis." Appl Environ Microbiol **69**(11): 6634-43.

Jungmann, J. and J. Benndorf (1994). "Toxicity of *Daphnia* of a compound extracted from laboratory and natural *Microcystis* spp., and the role of microcystin." Freshwater biology **32**: 13-20.

Kakuta, Y., et al. (1998). "Conserved structural motifs in the sulfotransferase family." Trends Biochem Sci **23**(4): 129-30.

Kaneko, T., et al. (2001). "Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120." DNA Res **8**(5): 205-13; 227-53.

Kao, C. M., et al. (1995). "Manipulation of macrolide ring size by directed mutagenesis of a modular polyketide synthase." J. Am. Chem. Soc. **117**: 9105-9106.

Kaplan, J. B. and D. H. Fine (1998). "Codon usage in *Actinobacillus* actinomycetemcomitans." FEMS Microbiol Lett **163**(1): 31-6.

Kardinaal, W. E., et al. (2007). "Competition for light between toxic and nontoxic strains of the harmful cyanobacterium *Microcystis*." Appl Environ Microbiol **73**(9): 2939-46.

Kasting, J. F. and J. L. Siefert (2002). "Life and the evolution of Earth's atmosphere." Science **296**(5570): 1066-8.

Kellmann, R., et al. (2008b). "Identification of a saxitoxin biosynthesis gene with a history of frequent horizontal gene transfers." J Mol Evol **67**(5): 526-38.

Kellmann, R., et al. (2008a). "Biosynthetic intermediate analysis and functional homology reveal a saxitoxin gene cluster in cyanobacteria." Appl Environ Microbiol **74**(13): 4044-53.

Kellmann, R., et al. (2006). "Functional modeling and phylogenetic distribution of putative cylindrospermopsin biosynthesis enzymes." J Mol Evol **62**(3): 267-80.

Khosla, C., et al. (1999). "Tolerance and specificity of polyketide synthases." Annu Rev Biochem **68**: 219-53.

Khosla, C., et al. (2007). "Structure and mechanism of the 6-deoxyerythronolide B synthase." Annu Rev Biochem **76**: 195-221.

Kleinkauf, H. and H. von Döhren (1996). "A non-ribosomal system of peptide biosynthesis." Eur. J. Biochem. **236**: 335-351.

Kohara, Y., et al. (1987). "The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library." Cell **50**(3): 495-508.

Kokocinski, M., et al. (2009). "First report of the cyanobacterial toxin cylindrospermopsin in the shallow, eutrophic lakes of western Poland." Chemosphere **74**(5): 669-75.

Koonin, E. V., et al. (2001). "Horizontal gene transfer in prokaryotes: quantification and classification." Annu Rev Microbiol **55**: 709-42.
Koski, M. K., et al. (2009). "The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif." J Biol Chem.
Kuhstoss, S., et al. (1996). "Production of a novel polyketide through the construction of a hybrid polyketide synthase." Gene **183**(1-2): 231-6.
Kumar, K. T. a. S. (2002). "Evolutionary distance estimation under heterogeneous substitution pattern among lineages " Molecular biology and Evolution **19**: 1727-1736.
Kunin, V., et al. (2005). "The net of life: reconstructing the microbial phylogenetic network." Genome Res **15**: 954-959.
Kunin, V. and C. A. Ouzounis (2003). "The balance of driving forces during genome evolution in prokaryotes." Genome Res **13**(7): 1589-94.
Kurmayer, R., et al. (2004). "Abundance of active and inactive microcystin genotypes in populations of the toxic cyanobacterium *Planktothrix* spp." Environ Microbiol **6**(8): 831-41.
Kurmayer, R., et al. (2002). "Diversity of microcystin genes within a population of the toxic cyanobacterium *Microcystis* spp. in Lake Wannsee (Berlin, Germany)." Microbial Ecology **43**: 107-118.
Kurmayer, R. and T. Kutzenberger (2003). "Application of real-time PCR for quantification of microcystin genotypes in a population of the toxic cyanobacterium *Microcystis* sp." Appl Environ Microbiol **69**(11): 6723-30.
Lagos, N., et al. (1999). "The first evidence of paralytic shellfish toxins in the freshwater cyanobacterium *Cylindrospermopsis raciborskii*, isolated from Brazil." Toxicon **37**: 1359-1373.
Leitao, E., et al. (2006). "Genes involved in the maturation of hydrogenase(s) in the nonheterocystous cyanobacterium *Lyngbya majuscula* CCAP 1446/4." International Journal of Hydrogen Energy **31**: 1469-1477.
Li, R., et al. (2001). "First report of the cyanotoxins cylindrospermopsin and deoxycylindrospermopsin from *Raphidiopsis curvata* (Cyanobacteria)." Journal of Phycology **37**: 1121-1126.
Li, R. H. and M. M. Watanabe (2002). "DNA base composition of planktonic species of *Anabaena* (Cyanobacteria) and its taxonomic value." J Gen Appl Microbiol **48**(2): 77-82.
Lieman-Hurwitz, J., et al. (2009). "A cyanobacterial AbrB-like protein affects the apparent photosynthetic affinity for CO2 by modulating low-CO2-induced gene expression." Environ Microbiol **11**(4): 927-36.
Liu, T., et al. (2006). "Identification of NanE as the thioesterase for polyether chain release in nanchangmycin biosynthesis." Chem Biol **13**(9): 945-55.
Lockhart, P. J., et al. (1994). "Recovering evolutionary trees under a more realistic model of sequence evolution." Mol Biol Evol **11**(4): 605-12.
Looper, R. E., et al. (2005). "Synthesis of the putative structure of 7-deoxycylindrospermopsin: C7 oxygenation is not required for the inhibition of protein synthesis." Angew Chem Int Ed Engl **44**(25): 3879-81.
Lyra, C., et al. (2000). "Site-specific restriction endonucleases in cyanobacteria." J Appl Microbiol **89**(6): 979-91.
MacRae, I. J., et al. (2000). "Crystal structure of adenosine 5'-phosphosulfate kinase from *Penicillium chrysogenum*." Biochemistry **39**(7): 1613-21.

Mahmood, N. A. and W. W. Carmichael (1986). "Paralytic shellfish poisons produced by the freshwater cyanobacterium *Aphanizomenon flos-aquae* NH-5." Toxicon **24**: 175-186.
Maier, H. R., et al. (2001). "Flow management strategies to control blooms of the cyanobacterium, *Anabaena circinalis*, in the River Murray at Morgan, South Australia." Research and management **17**: 637-650.
Mann, N. and N. G. Carr (1974). "Control of macromolecular composition and cell division in the blue-green algae *Anacystis nidulans*." J Gen Microbiol **83**(2): 399-405.
Marahiel, M. A. and L. O. Essen (2009). "Chapter 13. Nonribosomal peptide synthetases mechanistic and structural aspects of essential domains." Methods Enzymol **458**: 337-51.
Marahiel, M. A., et al. (1997). "Modular peptide synthetases involved in nonribosomal peptide synthesis." Chem Rev **97**(7): 2651-2674.
Martin, C., et al. (1993). "Cyanopeptolins, new depsipeptides from the cyanobacterium *Microcystis* sp. PCC 7806." J Antibiot (Tokyo) **46**(10): 1550-6.
Mbedi, S., et al. (2005). "Variability of the microcystin synthetase gene cluster in the genus *Planktothrix* (Oscillatoriales, Cyanobacteria)." FEMS Microbiol Lett **245**(2): 299-306.
McGregor, G. B. and L. D. Fabbro (2000). "Dominance of *Cylindrospermopsis raciborskii* (Nostocales, Cyanoprokaryota) in Queensland tropical and subtropical reservoirs: Implications for monitoring and management." Lakes & Reservoirs: Research and Management **5**: 195–205.
McGregor, G. B. and J. P. Rasmussen (2008). "Cyanobacterial composition of microbial mats from an Australian thermal spring: a polyphasic evaluation." FEMS Microbiol Ecol **63**(1): 23-35.
Medigue, C., et al. (1991). "Evidence for horizontal gene transfer in *Escherichia coli* speciation." J Mol Biol **222**(4): 851-6.
Meißner, K., et al. (1996). "Toxic and non-toxic strains of the cyanobacterium *Microcystis aeruginosa* contain sequences homologous to peptide synthetase genes." FEMS Microbiology Letters **135**: 295-303.
Meriluoto, J., et al. (2000). "Isolation and detection of microcystins and nodularins, cyanobacterial peptide hepatotoxins." Methods Mol Biol **145**: 65-87.
Metcalf, J. S., et al. (2001). "Colorimetric immuno-protein phosphatase inhibition assay for specific detection of microcystins and nodularins of cyanobacteria." Applied and Environmental Microbiology **67**(2): 904-909.
Metcalf, J. S., et al. (2002). "Toxicity of cylindrospermopsin to the brine shrimp *Artemia salina*: comparisons with protein synthesis inhibitors and microcystins." Toxicon **40**(8): 1115-120.
Micheli, L., et al. (2002). "Production of antibodies and development of highly sensitive formats of enzyme immunoassay for saxitoxin analysis." Anal Bioanal Chem **373**(8): 678-84.
Mihali, T. K., et al. (2008). "Characterization of the gene cluster responsible for cylindrospermopsin biosynthesis." Appl Environ Microbiol **74**(3): 716-22.
Mihali, T. K., et al. (2009). "Characterisation of the paralytic shellfish toxin biosynthesis gene clusters in *Anabaena circinalis* AWQC131C and *Aphanizomenon* sp. NH-5." BMC Biochem **10**: 8.
Moffitt, M. C. and B. A. Neilan (2001). "On the presence of peptide synthetase and polyketide synthase genes in the cyanobacterial genus *Nodularia*." FEMS Microbiology Letters **196**(2): 207-214.

Moffitt, M. C. and B. A. Neilan (2004). "Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins." Appl Environ Microbiol **70**(11): 6353-62.
Monis, P. T. and S. Giglio (2006). "Nucleic acid amplification-based techniques for pathogen detection and identification." Infect Genet Evol **6**(1): 2-12.
Monis, P. T., et al. (2005). "Comparison of SYTO9 and SYBR Green I for real-time polymerase chain reaction and investigation of the effect of dye concentration on amplification and DNA melting curve analysis." Anal Biochem **340**(1): 24-34.
Moore, R. E., et al. (1991). "Biosynthesis of Microcystin LR. Origin of the carbons in the Adda and Masp units." J. Am. Chem. Soc. **113**: 5083-5084.
Moss, S. J., et al. (2004). "Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity." Nat Prod Rep **21**(5): 575-93.
Mougous, J. D., et al. (2002). "Sulfotransferases and sulfatases in mycobacteria." Chem Biol **9**(7): 767-76.
Murphy, P. and C. Thomas (2001). "The synthesis and biological activity of the marine metabolite cylindrospermopsin." Chem. Soc. Rev. **30**: 303-312.
Myllyharju, J. (2003). "Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis." Matrix Biol **22**(1): 15-24.
Nakasugi, K., et al. (2007). "Functional analysis of PilT from the toxic cyanobacterium *Microcystis aeruginosa* PCC 7806." J Bacteriol **189**(5): 1689-97.
Namikoshi, M., et al. (2004). "Biosynthesis and transformation of homoanatoxin-a in the cyanobacterium *Raphidiopsis mediterranea* Skuja and structures of three new homologues." Chem Res Toxicol **17**(12): 1692-6.
Nei, M. and T. Gojobori (1986). "Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions." Mol Biol Evol **3**(5): 418-26.
Neilan, B. A., et al. (1995). "Genetic diversity and phylogeny of toxic cyanobacteria determined by DNA polymorphisms within the phycocyanin locus." Applied and Environmental Microbiology **61**(11): 3875-3883.
Neilan, B. A., et al. (2003). "Phylogeography of the invasive cyanobacterium *Cylindrospermopsis raciborskii*." Mol Ecol **12**(1): 133-40.
Neuwald, A. F., et al. (1999). "AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes." Genome Res **9**(1): 27-43.
NHMRC/NRMMC (2004). Australian Drinking Water Guidelines. Canberra, National Health and Medical Research Council/Natural Resource Management Ministerial Council.
Nishizawa, T., et al. (1999). "Genetic analysis of the peptide synthetase genes for a cyclic heptapeptide microcystin in *Microcystis* spp." Journal of Biochemistry **126**: 520-529.
Nishizawa, T., et al. (2001). "Cyclic heptapeptide microcystin biosynthesis requires the glutamate racemase gene." Microbiology **147**(Pt 5): 1235-1241.
Nishizawa, T., et al. (2000). "Polyketide synthase gene coupled to the peptide synthetase module involved in the biosynthesis of the cyclic heptapeptide microcystin." Journal of Biochemistry (Tokyo) **127**: 779-789.
Nixdorf, B., et al. (2003). "Phytoplankton assemblages and steady state in deep and shallow eutrophic lakes – an approach to differentiate the habitat properties of Oscillatoriales." Hydrobiologia **502**: 111-121.
Nonneman, D. and P. V. Zimba (2002). "A PCR-based test to assess the potential for microcystin occurrence in channel catfish production ponds." Journal of Phycology **38**: 230-233.
Norris, R. L., et al. (1999). "Deoxycylindrospermopsin, an analog of cylindrospermopsin from *Cylindrospermopsis raciborskii*." Environmental Toxicology **14**(1): 163-165.

Nübel, U., et al. (1997). "PCR primers to amplify 16S rRNA genes from cyanobacteria." Applied and Environmental Microbiology **63**(8): 3327-3332.
Ochman, H., et al. (1988). "Genetic applications of an inverse polymerase chain reaction." Genetics **120**(3): 621-3.
Ochman, H., et al. (2000). "Lateral gene transfer and the nature of bacterial innovation." Nature **405**(6784): 299-304.
Ohtake, A., et al. (1989). "Toxicity of *Microcystis* species isolated from natural blooms and purification of the toxin." Appl Environ Microbiol **55**(12): 3202-7.
Ohtani, I., et al. (1992). "Cylindrospermopsin: a potent hepatotoxin from the blue-green alga *Cylindrospermopsis raciborskii*." Journal of the American Chemical Society **114**: 7941-7942.
Okada, K., et al. (1997). "Three-dimensional structure of *Escherichia coli* branched-chain amino acid aminotransferase at 2.5 A resolution." J Biochem **121**(4): 637-41.
Onodera, H., et al. (1997). "Confirmation of anatoxin-a(s), in the cyanobacterium *Anabaena lemmermannii*, as the cause of bird kills in Danish lakes." Toxicon **35**(11): 1645-8.
Onodera, H., et al. (1997). "New saxitoxin analogues from the freshwater filamentous cyanobacterium *Lyngbya wollei*." Natural Toxins **5**: 146-151.
Ouahid, Y., et al. (2005). "Identification of potentially toxic environmental *Microcystis* by individual and multiple PCR amplification of specific microcystin synthetase gene regions." Environ Toxicol **20**(3): 235-42.
Padisák, J. (1997). "*Cylindrospermopsis raciborskii* (Woloszynska) Seenayya and Subba Raju, an expanding, highly adaptive cyanobacterium: worldwide distribution and review of its ecology." Archiv fuer Hydrobiologie **4**: 563-593.
Paerl, H. W., et al. (2001). "Harmful freshwater algal blooms, with an emphasis on cyanobacteria." ScientificWorldJournal **1**: 76-113.
Paerl, H. W. and J. Huisman (2009). "Climate change: a catalyst for global expansion of harmful cyanobacterial blooms." Environmental Microbiology Reports **1**(1): 27-37.
Page, R. D. and E. C. Holmes (1998). Molecular Evolution: A phylogenetic Approach. Oxford, United Kingdom, Blackwell Science Ltd.
Palenik, B. and R. Haselkorn (1992). "Multiple evolutionary origins of prochlorophytes, the chlorophyll-containing prokaryotes." Nature (London) **355**: 265-267.
Palenik, B. and R. Haselkorn (1992). "Multiple evolutionary origins of prochlorophytes, the chlorophyll b-containing prokaryotes." Nature **355**(6357): 265-7.
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.
Pearson, L. A., et al. (2004). "Inactivation of an ABC transporter gene, *mcy*H, results in loss of microcystin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806." Appl Environ Microbiol **70**(11): 6370-8.
Pereira, P., et al. (2000). "Paralytic shellfish toxins in the freshwater cyanobacterium *Aphanizomenon flos-aquae,* isolated from Montargil reservoir, Portugal." Toxicon **38**: 1689-1702.
Pitois, S., et al. (2001). "Sources of the eutrophication problems associated with toxic algae: an overview." Journal of Environmental Health **64**(5): 25-32.
Ploutno, A., et al. (2002). "Three novel protease inhibitors from a natural bloom of the cyanobacterium *Microcystis aeruginosa*." J Nat Prod **65**(7): 973-8.
Pomati, F., et al. (2000). "The freshwater cyanobacterium *Planktothrix* sp. FP1: molecular identification and detection of paralytic shellfish poisoning toxins." Journal of Phycology **36**: 553-562.

Preussel, K., et al. (2006). "First report on cylindrospermopsin producing *Aphanizomenon flos-aquae* (Cyanobacteria) isolated from two German lakes." Toxicon **47**(2): 156-62.
Putman, M., et al. (2000). "Molecular properties of bacterial multidrug transporters." Microbiol Mol Biol Rev **64**(4): 672-93.
Rajaniemi, P., et al. (2005). "Phylogenetic and morphological evaluation of the genera *Anabaena*, *Aphanizomenon*, *Trichormus* and *Nostoc* (Nostocales, Cyanobacteria)." Int J Syst Evol Microbiol **55**(Pt 1): 11-26.
Rantala, A., et al. (2004). "Phylogenetic evidence for the early evolution of microcystin synthesis." Proc Natl Acad Sci U S A **101**(2): 568-73.
Rao, P. V., et al. (2002). "Toxins and bioactive compounds from cyanobacteria and their implications on human health." J Environ Biol **23**(3): 215-24.
Rasmussen, J. P., et al. (2008a). "Rapid preparation of cyanobacterial DNA for real-time PCR analysis." Lett Appl Microbiol **46**(1): 14-9.
Rasmussen, J. P., et al. (2008b). "Development and field testing of a real-time PCR assay for cylindrospermopsin-producing cyanobacteria." J Appl Microbiol **104**(5): 1503-15.
Reynolds, C. S., et al. (2002). "Towards a functional classification of the freshwater phytoplankton." Journal of Plankton Research **24**(5): 417-428.
Rinta-Kanto, J. M., et al. (2005). "Quantification of toxic *Microcystis* spp. during the 2003 and 2004 blooms in western Lake Erie using quantitative real-time PCR." Environ Sci Technol **39**(11): 4198-205.
Rippka, R. (1988). "Recognition and identification of cyanobacteria " Methods in Enzymology **167**: 28-67.
Rouhiainen, L., et al. (2000). "Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in *Anabaena* strain 90." Molecular Microbiology **37**(1): 156-167.
Rouhiainen, L., et al. (2004). "Genes coding for hepatotoxic heptapeptides (microcystins) in the cyanobacterium *Anabaena* strain 90." Appl Environ Microbiol **70**(2): 686-92.
Runnegar, M. T. and I. R. Falconer (1986). "Effect of toxin from the cyanobacterium *Microcystis aeruginosa* on ultrastructural morphology and actin polymerization in isolated hepatocytes." Toxicon **24**(2): 109-15.
Runnegar, M. T., et al. (1995). "Inhibition of reduced glutathione synthesis by cyanobacterial alkaloid cylindrospermopsin in cultured rat hepatocytes." Biochemical Pharmacology **49**(2): 219-225.
Runnegar, M. T., et al. (1993). "Protein phosphatase inhibition and in vivo hepatotoxicity of microcystins." Am J Physiol **265**(2 Pt 1): G224-30.
Saiki, R. K., et al. (1988). "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase." Science **239**(4839): 487-91.
Saiki, R. K., et al. (1985). "Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia." Science **230**(4732): 1350-4.
Saker, M. L. and D. J. Griffiths (2000). "The effect of temperature on growth and cylindrospermopsin content of seven isolates of *Cylindrospermopsis raciborskii* (Nostocales, Cyanophyceae) from water bodies in northern Australia." Phycologia **39**(4): 349-354.
Saker, M. L., et al. (1999). "Cattle mortality attributed to the toxic cyanobacterium *Cylindrospermopsis raciborskii* in an outback region of North Queensland." Environmental Toxicology **14**(1): 179-182.
Satishchandran, C., et al. (1992). "Characterization of the phosphorylated enzyme intermediate formed in the adenosine 5'-phosphosulfate kinase reaction." Biochemistry **31**(47): 11684-8.

Saunders, N. A. (2004). Quantitative Real-Time PCR. Real-Time PCR: An essential Guide. K. Edwards, J. Logan and N. Saunders. Wymondham,UK, Horizon Bioscience**:** 103-125.

Saunders, N. A., Ed. (2004a). An introduction to Real-Time PCR. Real-Time PCR: an essential guide. Wymondham,UK, Horizon Bioscience.

Saunders, N. F., et al. (2003). "Mechanisms of thermal adaptation revealed from the genomes of the Antarctic Archaea *Methanogenium frigidum* and *Methanococcoides burtonii*." Genome Res **13**(7): 1580-8.

Schantz, E. J., et al. (1975). "Letter: The structure of saxitoxin." J Am Chem Soc **97**(5): 1238.

Schatz, D., et al. (2007). "Towards clarification of the biological role of microcystins, a family of cyanobacterial toxins." Environ Microbiol **9**(4): 965-70.

Schembri, M. A., et al. (2001). "Identification of genes implicated in toxin production in the cyanobacterium *Cylindrospermopsis raciborskii*." Environmental Toxicology **16**: 413-421.

Schneider, G. J., et al. (1987). "Purification and characterization of RNA polymerase from the cyanobacterium *Anabaena* 7120." J Biol Chem **262**(30): 14633-9.

Schriek, S., et al. (2007). "Bioinformatic evaluation of L-arginine catabolic pathways in 24 cyanobacteria and transcriptional analysis of genes encoding enzymes of L-arginine catabolism in the cyanobacterium *Synechocystis* sp. PCC 6803." BMC Genomics **8**: 437.

Seibert, C. M. and F. M. Raushel (2005). "Structural and catalytic diversity within the amidohydrolase superfamily." Biochemistry **44**(17): 6383-91.

Seifert, M., et al. (2007). "First evidence for the production of cylindrospermopsin and deoxy-cylindrospermopsin by the freshwater benthic cyanobacterium, *Lyngbya wollei* (Farlow ex Gomont) Speziale and Dyck " Harmful Algae **6**(1): 73-80.

Shalev-Alon, G., et al. (2002). "A novel gene encoding amidinotransferase in the cylindrospermopsin producing cyanobacterium *Aphanizomenon ovalisporum*." FEMS Microbiology Letters **209**: 87-91.

Shalev-Malul, G., et al. (2008). "An AbrB-like protein might be involved in the regulation of cylindrospermopsin production by *Aphanizomenon ovalisporum*." Environ Microbiol **10**(4): 988-99.

Sharp, P. M., et al. (1986). "Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes." Nucleic Acids Res **14**(13): 5125-43.

Shen, B. (2003). "Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms." Current Opinion in Chemical Biology **7**: 285–295.

Siebert, P. D., et al. (1995). "An improved PCR method for walking in uncloned genomic DNA." Nucleic Acids Res **23**(6): 1087-8.

Sitachitta, N., et al. (2000). "Biosynthetic pathway and origin of the chlorinated methyl group in barbamide and dechlorobarbamide, metabolites from the marine cyanobacterium *Lyngbya majuscula*." Tetrahedron **56**: 9130-9133.

Sivonen, K. (2008). "Emerging high throughput analyses of cyanobacterial toxins and toxic cyanobacteria." Adv Exp Med Biol **619**: 539-57.

Sivonen, K., et al. (1992a). "Isolation and characterization of a variety of microcystins from seven strains of the cyanobacterial genus *Anabaena*." Appl Environ Microbiol **58**(8): 2495-500.

Sivonen, K., et al. (1992b). "Three new microcystins, cyclic heptapeptide hepatotoxins, from *Nostoc* sp. strain 152." Chem Res Toxicol **5**(4): 464-9.

Smith, D. S. and D. D. Kitts (1994). "Development of a monoclonal-based enzyme-linked immunoassay for saxitoxin-induced protein." Toxicon **32**(3): 317-23.

Smith, M. J., et al. (2008). "Elucidating the factors influencing the biodegradation of cylindrospermopsin in drinking water sources." Environ Toxicol **23**(3): 413-21.
Spoof, L., et al. (2006). "First observation of cylindrospermopsin in *Anabaena lapponica* isolated from the boreal environment (Finland)." Environ Toxicol **21**(6): 552-60.
Stachelhaus, T., et al. (1999). "The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases." Chem Biol **6**(8): 493-505.
Staunton, J. and K. J. Weissman (2001). "Polyketide biosynthesis: a millennium review." Nat Prod Rep **18**(4): 380-416.
Staunton, J. and B. Wilkinson (1997). "Biosynthesis of erythromycin and rapamycin." Chem Rev **97**(7): 2611-2630.
Steffensen, D. A. (2008). "Economic cost of cyanobacterial blooms." Adv Exp Med Biol **619**: 855-65.
Stewart, I., et al. (2006b). "Cutaneous hypersensitivity reactions to freshwater cyanobacteria--human volunteer studies." BMC Dermatol **6**: 6.
Stewart, I., et al. (2006a). "Primary irritant and delayed-contact hypersensitivity reactions to the freshwater cyanobacterium *Cylindrospermopsis raciborskii* and its associated toxin cylindrospermopsin." BMC Dermatol **6**: 5.
Stewart, I., et al. (2008). "Cyanobacterial poisoning in livestock, wild mammals and birds--an overview." Adv Exp Med Biol **619**: 613-37.
Stirling, D. J. and M. A. Quilliam (2001). "First report of the cyanobacterial toxin cylindrospermopsin in New Zealand." Toxicon **39**(8): 1219-1222.
Stucken, K., et al. (2009). "Toxicity phenotype does not correlate with phylogeny of *Cylindrospermopsis raciborskii* strains." Syst Appl Microbiol **32**(1): 37-48.
Stüken, A. (2008). Cylindrospermopsin and cylindrospermopsin producing cyanobacteria in the area Berlin-Brandenburg. Cottbus, Germany, Brandenburg University of Technology. **PhD Thesis**.
Stüken, A., et al. (2009). "Genetic and morphologic characterization of four putative cylindrospermopsin producing species of the cyanobacterial genera *Anabaena* and *Aphanizomenon*." Journal of Plankton Research **31**(5): 465-480.
Suda, S., et al. (2002). "Taxonomic revision of water-bloom-forming species of oscillatorioid cyanobacteria." Int J Syst Evol Microbiol **52**(Pt 5): 1577-95.
Takahashi, I., et al. (1996). "Restriction barrier composed of an extracellular nuclease and restriction endonuclease in the unicellular cyanobacterium *Microcystis* sp." FEMS Microbiol Lett **145**(1): 107-11.
Tamagnini, P., et al. (2007). "Cyanobacterial hydrogenases: diversity, regulation and applications." FEMS Microbiol Rev **31**(6): 692-720.
Tamas, I., et al. (2000). "Determinative value of a portion of the *nif*H sequence for the genera *Nostoc* and *Anabaena* (cyanobacteria)." Current Microbiology **41**(3): 197-200.
Tamura, K., et al. (2007). "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0." Molecular Biology and Evolution **24**: 1596-1599.
Tamura, K. and S. Kumar (2002). "Evolutionary distance estimation under heterogeneous substitution pattern among lineages." Mol Biol Evol **19**(10): 1727-36.
Tanabe, Y., et al. (2004). "Evidence for recombination in the microcystin synthetase (mcy) genes of toxic cyanobacteria *Microcystis* spp." J Mol Evol **58**(6): 633-41.
Terao, K., et al. (1994). "Electron microscopic studies on experimental poisoning in mice induced by cylindrospermopsin isolated from blue-green alga *Umezakia natans*." Toxicon **32**: 833-843.
Thoden, J. B., et al. (2001). "Molecular structure of dihydroorotase: a paradigm for catalysis through the use of a binuclear metal center." Biochemistry **40**(24): 6989-97.

Thomas, A. D., et al. (1998). "Cyanobacterium *Cylindrospermopsis raciborskii* as a probable cause of death in cattle in northern Queensland." Australian Veterinary Journal **76**(9): 592-594.

Tillett, D. (2000b). Microcystin:Genetics,Biosynthesis and Evolution. School of Microbiology and Immunology. Sydney, University of New South Wales. **PhD:** 286.

Tillett, D., et al. (2000a). "Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC 7806: an integrated peptide-polyketide synthetase system." Chemistry and Biology **7**: 753-764.

Tillett, D., et al. (2001). "Detection of toxigenicity by a probe for the microcystin synthetase A gene (*mcy*A) of the cyanobacterial genus *Microcystis*: comparison of toxicities with 16S rRNA and phycocyanin operon (Phycocyanin Intergenic Spacer) phylogenies." Appl Environ Microbiol **67**(6): 2810-8.

Toledo, G. and B. Palenik (1997). "*Synechococcus* diversity in the California current as seen by RNA polymerase (*rpo*C1) gene sequences of isolated strains." Applied and Environmental Microbiology **63**(11): 4298-4303.

Tooming-Klunderud, A., et al. (2008). "The mosaic structure of the *mcy*ABC operon in *Microcystis*." Microbiology **154**(Pt 7): 1886-99.

Torokne, A., et al. (2004). "Interlaboratory comparison trial on cylindrospermopsin measurement." Anal Biochem **332**(2): 280-4.

Townsend, C. A. (1997). "Structural studies of natural product biosynthetic proteins." Chem Biol **4**(10): 721-30.

Vaitomaa, J., et al. (2003). "Quantitative real-time PCR for determination of microcystin synthetase e copy numbers for *Microcystis* and *Anabaena* in lakes." Appl Environ Microbiol **69**(12): 7289-97.

Valerio, E., et al. (2009). "Molecular identification, typing and traceability of cyanobacteria from freshwater reservoirs." Microbiology **155**(Pt 2): 642-56.

van Apeldoorn, M. E., et al. (2007). "Toxins of cyanobacteria." Mol Nutr Food Res **51**(1): 7-60.

Verdier-Pinard, P., et al. (1999). "Biosynthesis of radiolabeled curacin A and its rapid and apparently irreversible binding to the colchicine site of tubulin." Arch Biochem Biophys **370**(1): 51-8.

Vezie, C., et al. (1998). "Variation of microcystin content of cyanobacterial blooms and isolated strains in Lake Grand-Lieu (France)." Microb Ecol **35**(2): 126-35.

Viaggiu, E., et al. (2004). "Anatoxin-a toxin in the cyanobacterium *Planktothrix rubescens* from a fishing pond in northern Italy." Environ Toxicol **19**(3): 191-7.

von Döhren, H., et al. (1999). "The nonribosomal code." Chemistry & Biology **6**: R273-R279.

Wacklin, P., et al. (2009). "Nomenclatural validation of the genetically revised cyanobacterial genus *Dolichospermum* (Ralfs ex Bornet et Flahault) comb.nova." Fottea **9**(1): 59-64.

Walsh, C. T. (2004). "Polyketide and nonribosomal peptide antibiotics: modularity and versatility." Science **303**(5665): 1805-10.

Waterbury, J. B. (2006). The cyanobacteria-isolation, purification and identification. The Prokaryotes. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer and E. Stacklebrandt. New York, Springer. **4:** 1053-1073.

Weile, J. and C. Knabbe (2009). "Current applications and future trends of molecular diagnostics in clinical bacteriology." Anal Bioanal Chem **394**(3): 731-42.

Welker, M., Lenka Šejnohová, Danka Némethová, Hans von Döhren, Jirí Jarkovský, and Blahoslav Maršálek (2007). "Seasonal shifts in chemotype composition of *Microcystis* sp. communities in the pelagial and the sediment of a shallow reservoir." Limnol. Oceanogr **52**(2): 609-619.

WHO (2004). World Health Organization; Guidelines for Drinking-water Quality. Volume 1. Recommendations. Geneva, WHO Publishing.
Wiedner, C., et al. (2007). "Climate change affects timing and size of populations of an invasive cyanobacterium in temperate regions." Oecologia **152**(3): 473-84.
Wilmotte, A. (1994). Molecular evolution and taxonomy of the cyanobacteria. The molecular biology of cyanobacteria. D. A. Bryant. Dordrecht, The Netherlands, Kluwer Academic Publishers**:** 1-25.
Wilmotte, A. and M. Herdman (2001). Phylogenetic Relationships Among the Cyanobacteria Based on 16S rRNA Sequences. New York, Springer-Verlag.
Wilson, K. M. (1998). Genetic analysis of the toxic cyanobacterium *Cylindrospermopsis raciborskii*. School of Chemical Technology. Adelaide, University of South Australia**:** 84.
Wilson, K. M., et al. (2000). "Molecular characterization of the toxic cyanobacterium *Cylindrospermopsis raciborskii* and design of a species-specific PCR." Appl Environ Microbiol **66**(1): 332-8.
Woese, C. R. (1987). "Bacterial evolution." Microbiol Rev **51**(2): 221-71.
Wood, S. A., et al. (2007b). "First report of the cyanotoxin anatoxin-A from *Aphanizomenon issatschenkoi*(cyanobacteria)." Journal of Phycology **43**(2): 356-365.
Wood, S. A., et al. (2007a). "First report of homoanatoxin-a and associated dog neurotoxicosis in New Zealand." Toxicon **50**(2): 292-301.
Xie, W., et al. (1989). "Cyanobacterial RNA polymerase genes *rpo*C1 and *rpo*C2 correspond to *rpo*C of *Escherichia coli*." Journal of Bacteriology **171**(4): 1967-1973.
Yilmaz, M., et al. (2008). "A comparative study of Florida strains of *Cylindrospermopsis* and *Aphanizomenon* for cylindrospermopsin production." Toxicon **51**(1): 130-9.
Yu, F. Y., et al. (2002). "Development of a sensitive ELISA for the determination of microcystins in algae." J Agric Food Chem **50**(15): 4176-82.
Yvon, M., et al. (2000). "Characterization and role of the branched-chain aminotransferase (BcaT) isolated from *Lactococcus lactis* subsp. cremoris NCDO 763." Appl Environ Microbiol **66**(2): 571-7.
Zhaxybayeva, O., et al. (2006). "Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events." Genome Res **16**(9): 1099-108.
Zhou, L., et al. (2002). "Relationship between microcystin in drinking water and colorectal cancer." Biomed Environ Sci **15**(2): 166-71.

Appendices

Appendix 1 *Cylindrospermopsis raciborskii* AWT205 CYN Biosynthetic Gene Cluster Sequence

```
*** INPUT INFORMATION ***

    FILE  NAME  : CYNJ.SEQ       SEQUENCE : NORMAL   3450 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -   3450

    TRANSLATION REGION :      1 -   3450


*** DNA TRANSLATION ***

   1  Met Val Cys Val Ala Leu Ala Val Lys Ser Leu Arg Gln Gln Glu Cys      16
   1  ATG GTT TGT GTG GCA TTG GCT GTC AAG AGC CTA CGT CAA CAG GAG TGT      48

  17  Asp Leu Ala Leu Ala Gly Gly Val Gln Ile Gln Val Ile Pro Asp Gly      32
  49  GAT TTG GCA TTG GCG GGT GGT GTA CAG ATA CAA GTG ATA CCA GAT GGC      96

  33  Phe Ile Lys Ala Cys Gln Ser Arg Met Leu Ser Pro Asp Gly Arg Cys      48
  97  TTT ATC AAA GCC TGT CAA TCC CGT ATG TTG TCG CCT GAT GGA CGG TGC     144

  49  Lys Thr Phe Asp Phe Gln Ala Asp Gly Tyr Ala Arg Ala Glu Gly Cys      64
 145  AAA ACA TTT GAT TTC CAG GCA GAT GGT TAT GCC CGT GCT GAG GGG TGT     192

  65  Gly Met Val Val Leu Lys Arg Leu Ser Asp Ala Ile Ala Asp Asn Asp      80
 193  GGG ATG GTA GTT CTC AAA CGC CTA TCC GAT GCA ATT GCT GAC AAT GAT     240

  81  Asn Ile Leu Ala Leu Ile Arg Gly Ala Ala Val Asn His Asp Gly Tyr      96
 241  AAT ATC CTG GCC TTG ATT CGT GGT GCC GCA GTC AAT CAT GAT GGC TAC     288

  97  Thr Ser Gly Leu Thr Val Pro Ser Gly Pro Ser Gln Arg Ala Val Ile     112
 289  ACG AGT GGA TTA ACC GTT CCC AGT GGT CCC TCA CAA CGG GCG GTG ATC     336

 113  Gln Gln Ala Leu Ala Asp Ala Gly Ile His Pro Asp Gln Ile Ser Tyr     128
 337  CAA CAG GCA TTA GCG GAT GCT GGA ATA CAC CCG GAT CAA ATT AGC TAT     384

 129  Ile Glu Ala His Gly Thr Gly Thr Ser Leu Gly Asp Pro Ile Glu Met     144
 385  ATT GAG GCA CAT GGC ACA GGT ACA TCC TTA GGC GAT CCT ATT GAA ATG     432

 145  Gly Ala Ile Gly Gln Val Phe Gly Gln Arg Ser Gln Met Leu Phe Val     160
 433  GGT GCG ATT GGG CAA GTC TTT GGT CAA CGC TCA CAG ATG CTT TTC GTC     480

 161  Gly Ser Val Lys Thr Asn Ile Gly His Thr Glu Ala Ala Ala Gly Ile     176
 481  GGT TCG GTC AAG ACG AAT ATT GGT CAT ACT GAG GCT GCT GCT GGT ATT     528

 177  Ala Gly Leu Ile Lys Val Val Leu Ser Met Gln His Gly Glu Ile Pro     192
 529  GCT GGT CTC ATC AAG GTT GTA CTC TCA ATG CAG CAC GGT GAA ATC CCA     576

 193  Ala Asn Leu His Phe Asp Gln Pro Ser Pro Tyr Ile Asn Trp Asp Gln     208
 577  GCA AAC TTA CAC TTC GAC CAG CCA AGT CCT TAT ATT AAC TGG GAT CAA     624

 209  Leu Pro Val Ser Ile Pro Thr Glu Thr Ile Pro Trp Ser Thr Ser Asp     224
 625  TTA CCA GTC AGT ATC CCA ACA GAA ACA ATA CCT TGG TCT ACT AGC GAT     672

 225  Arg Phe Ala Gly Val Ser Ser Phe Gly Phe Ser Gly Thr Asn Ser His     240
 673  CGC TTT GCA GGA GTC AGT AGC TTT GGC TTT AGT GGC ACA AAC TCT CAT     720

 241  Ile Val Leu Glu Ala Ala Pro Asn Ile Glu Gln Pro Thr Asp Asp Ile     256
 721  ATC GTA CTA GAG GCA GCC CCA AAC ATA GAG CAA CCT ACT GAT GAT ATT     768

 257  Asn Gln Thr Pro His Ile Leu Thr Leu Ala Ala Lys Thr Pro Ala Ala     272
 769  AAT CAA ACG CCG CAT ATT TTG ACC TTA GCT GCA AAA ACA CCC GCA GCC     816
```

```
 273  Leu Gln Glu Leu Ala Arg Arg Tyr Ala Thr Gln Ile Glu Thr Ser Pro     288
 817  CTG CAA GAA CTG GCT CGG CGT TAT GCG ACT CAG ATA GAG ACC TCT CCC     864

 289  Asp Val Pro Leu Ala Asp Ile Cys Phe Thr Ala His Ile Gly Arg Lys     304
 865  GAT GTT CCT CTG GCG GAC ATT TGT TTC ACA GCA CAC ATA GGG CGT AAA     912

 305  His Phe Lys His Arg Phe Ala Val Val Thr Glu Ser Lys Glu Gln Leu     320
 913  CAT TTT AAA CAT AGG TTT GCG GTA GTC ACG GAA TCT AAA GAG CAA CTG     960

 321  Arg Leu Gln Leu Asp Ala Phe Ala Gln Ser Gly Gly Val Gly Arg Glu     336
 961  CGT TTG CAA TTG GAT GCA TTT GCA CAA TCA GGG GGT GTG GGG CGA GAA    1008

 337  Val Lys Ser Leu Pro Lys Ile Ala Phe Leu Phe Thr Gly Gln Gly Ser     352
1009  GTC AAA TCG CTA CCA AAG ATA GCC TTT CTT TTT ACA GGT CAA GGC TCA    1056

 353  Gln Tyr Val Gly Met Gly Arg Gln Leu Tyr Glu Asn Gln Pro Thr Phe     368
1057  CAG TAT GTG GGA ATG GGT CGT CAA CTT TAC GAA AAC CAA CCT ACC TTC    1104

 369  Arg Lys Ala Leu Ala His Cys Asp Asp Ile Leu Arg Ala Gly Ala Tyr     384
1105  CGA AAA GCA CTC GCC CAT TGT GAT GAC ATC TTG CGT GCT GGT GCA TAT    1152

 385  Phe Asp Arg Ser Leu Leu Ser Ile Leu Tyr Pro Glu Gly Lys Ser Glu     400
1153  TTC GAC CGA TCA CTA CTT TCG ATT CTC TAC CCA GAG GGA AAA TCA GAA    1200

 401  Ala Ile His Gln Thr Ala Tyr Thr Gln Pro Ala Leu Phe Ala Leu Glu     416
1201  GCC ATT CAC CAA ACC GCT TAT ACT CAG CCC GCG CTT TTT GCT CTT GAG    1248

 417  Tyr Ala Ile Ala Gln Leu Trp His Ser Trp Gly Ile Lys Pro Asp Ile     432
1249  TAT GCG ATC GCT CAG TTG TGG CAC TCC TGG GGT ATC AAA CCA GAT ATC    1296

 433  Val Met Gly His Ser Val Gly Glu Tyr Val Ala Ala Cys Val Ala Gly     448
1297  GTG ATG GGG CAT AGT GTA GGT GAA TAC GTC GCC GCT TGT GTG GCG GGC    1344

 449  Ile Phe Ser Leu Glu Asp Gly Leu Lys Leu Ile Ala Thr Arg Gly Arg     464
1345  ATA TTT TCT TTA GAG GAT GGG CTG AAA CTA ATT GCT ACT CGT GGT CGT    1392

 465  Leu Met Gln Ser Leu Pro Gln Asp Gly Thr Met Val Ser Ser Leu Ala     480
1393  CTG ATG CAA TCC CTA CCT CAA GAC GGA ACG ATG GTT TCT TCT TTG GCA    1440

 481  Ser Glu Ala Arg Ile Gln Glu Ala Ile Thr Pro Tyr Arg Asp Asp Val     496
1441  AGT GAA GCT CGT ATC CAG GAA GCT ATT ACA CCT TAC CGA GAT GAT GTG    1488

 497  Ser Ile Ala Ala Ile Asn Gly Thr Glu Ser Val Val Ile Ser Gly Lys     512
1489  TCA ATC GCA GCG ATA AAT GGG ACA GAA AGC GTG GTT ATC TCT GGC AAA    1536

 513  Arg Thr Ser Val Met Ala Ile Ala Glu Gln Leu Ala Thr Val Gly Ile     528
1537  CGC ACC TCT GTG ATG GCA ATT GCT GAA CAA CTC GCC ACC GTT GGC ATC    1584

 529  Lys Thr Arg Gln Leu Thr Val Ser His Ala Phe His Ser Pro Leu Met     544
1585  AAG ACA CGC CAA CTG ACG GTT TCC CAT GCC TTC CAT TCA CCA CTT ATG    1632

 545  Thr Pro Ile Leu Asp Glu Phe Arg Gln Val Ala Ala Ser Ile Thr Tyr     560
1633  ACA CCC ATC TTG GAT GAG TTC CGC CAG GTG GCA GCC AGT ATC ACC TAT    1680

 561  His Gln Pro Lys Leu Leu Leu Val Ser Asn Val Ser Gly Lys Val Ala     576
1681  CAC CAG CCC AAG TTG CTA CTT GTC TCC AAC GTC TCC GGG AAA GTG GCC    1728

 577  Gly Pro Glu Ile Thr Arg Pro Asp Tyr Trp Val Arg His Val Arg Glu     592
1729  GGC CCT GAA ATC ACC AGA CCA GAT TAC TGG GTA CGC CAT GTC CGT GAG    1776

 593  Ala Val Arg Phe Ala Asp Gly Val Arg Thr Leu Asn Glu Gln Gly Val     608
1777  GCA GTG CGC TTT GCC GAT GGA GTG AGG ACG CTG AAT GAA CAA GGT GTC    1824

 609  Asn Ile Phe Leu Glu Ile Gly Ser Thr Ala Thr Leu Leu Gly Met Ala     624
1825  AAT ATC TTT CTG GAA ATC GGT TCT ACC GCT ACC CTG TTG GGC ATG GCA    1872

 625  Leu Arg Val Asn Glu Glu Asp Ser Asn Ala Ser Lys Gly Thr Ser Ser     640
1873  CTG CGA GTA AAT GAG GAA GAT TCA AAT GCC TCA AAA GGA ACT TCG TCT    1920
```

```
 641  Cys Tyr Leu Pro Ser Leu Arg Glu Ser Gln Lys Asp Cys Gln Gln Met     656
1921  TGC TAC CTG CCC AGT TTA CGG GAA AGC CAG AAG GAT TGT CAG CAG ATG    1968

 657  Phe Thr Ser Leu Gly Glu Leu Tyr Val His Gly Tyr Asp Ile Asp Trp     672
1969  TTC ACT AGT CTG GGT GAG TTG TAC GTA CAT GGA TAT GAT ATT GAT TGG    2016

 673  Gly Ala Phe Asn Arg Gly Tyr Gln Gly Arg Lys Val Ile Leu Pro Thr     688
2017  GGT GCA TTT AAT CGG GGA TAT CAA GGA CGC AAG GTG ATA TTG CCA ACC    2064

 689  Tyr Pro Phe Gln Arg Gln Arg Tyr Trp Leu Pro Asp Pro Lys Leu Ala     704
2065  TAT CCG TTT CAG CGA CAA CGT TAT TGG CTT CCC GAC CCT AAG TTG GCA    2112

 705  Gln Ser Ser Asp Leu Asp Thr Phe Gln Ala Gln Ser Ser Ala Ser Ser     720
2113  CAA AGT TCC GAT TTA GAT ACC TTT CAA GCT CAG AGC AGC GCA TCA TCA    2160

 721  Gln Asn Pro Ser Ala Val Ser Thr Leu Leu Met Glu Tyr Leu Gln Ala     736
2161  CAA AAT CCT AGC GCT GTG TCC ACT TTA CTG ATG GAA TAT TTG CAA GCA    2208

 737  Gly Asp Val Gln Ser Leu Val Gly Leu Leu Asp Asp Glu Arg Lys Leu     752
2209  GGT GAT GTC CAA TCT TTA GTT GGG CTT TTG GAT GAT GAA CGG AAA CTC    2256

 753  Ser Ala Ala Glu Arg Ile Ala Leu Pro Ser Ile Leu Glu Phe Leu Val     768
2257  TCT GCT GCT GAA CGA ATT GCA CTA CCC AGT ATT TTG GAG TTT TTG GTA    2304

 769  Glu Glu Gln Gln Arg Gln Ile Ser Ser Thr Thr Thr Pro Gln Thr Val     784
2305  GAG GAA CAA CAG CGA CAA ATA AGC TCA ACC ACA ACT CCT CAA ACA GTT    2352

 785  Leu Gln Lys Ile Ser Gln Thr Ser His Glu Asp Arg Tyr Glu Ile Leu     800
2353  TTA CAA AAA ATA AGT CAA ACT TCC CAT GAG GAC AGA TAT GAA ATA TTG    2400

 801  Lys Asn Leu Ile Lys Ser Glu Ile Glu Thr Ile Ile Lys Ser Val Pro     816
2401  AAG AAC CTG ATC AAA TCT GAA ATC GAA ACG ATT ATC AAA AGT GTT CCC    2448

 817  Ser Asp Glu Gln Met Phe Ser Asp Leu Gly Ile Asp Ser Leu Met Ala     832
2449  TCC GAT GAA CAA ATG TTT TCT GAC TTA GGA ATT GAT TCC TTG ATG GCG    2496

 833  Ile Glu Leu Arg Asn Lys Leu Arg Ser Ala Ile Gly Leu Glu Leu Pro     848
2497  ATC GAA CTG CGT AAT AAG CTC CGT TCT GCT ATA GGG TTG GAA CTG CCA    2544

 849  Val Ala Ile Val Phe Asp His Pro Thr Ile Lys Gln Leu Thr Asn Phe     864
2545  GTG GCA ATA GTA TTT GAC CAT CCC ACG ATT AAG CAG TTA ACT AAC TTC    2592

 865  Val Leu Asp Arg Ile Val Pro Gln Ala Asp Gln Lys Asp Val Pro Thr     880
2593  GTA CTG GAC AGA ATT GTG CCG CAG GCA GAC CAA AAG GAC GTT CCC ACC    2640

 881  Glu Ser Leu Phe Ala Ser Lys Gln Glu Ile Ser Val Glu Glu Gln Ser     896
2641  GAA TCC TTG TTT GCT TCT AAA CAG GAG ATA TCA GTT GAG GAG CAG TCT    2688

 897  Phe Ala Ile Thr Lys Leu Gly Leu Ser Pro Ala Ser His Ser Leu His     912
2689  TTT GCA ATT ACC AAG CTG GGC TTA TCC CCT GCT TCC CAC TCC CTG CAT    2736

 913  Leu Pro Pro Trp Thr Val Arg Pro Ala Val Met Ala Asp Val Thr Lys     928
2737  CTT CCT CCA TGG ACG GTT AGA CCT GCG GTA ATG GCA GAT GTA ACA AAA    2784

 929  Leu Ser Gln Leu Glu Arg Glu Ala Tyr Gly Trp Ile Gly Glu Gly Ala     944
2785  CTA AGC CAA CTT GAA AGA GAG GCC TAT GGC TGG ATC GGA GAA GGA GCG    2832

 945  Ile Ala Pro Pro His Leu Ile Ala Asp Arg Ile Asn Leu Leu Asn Ser     960
2833  ATC GCC CCG CCC CAT CTC ATT GCC GAT CGC ATC AAT TTA CTC AAC AGT    2880

 961  Gly Asp Met Pro Trp Phe Trp Val Met Glu Arg Ser Gly Glu Leu Gly     976
2881  GGT GAT ATG CCT TGG TTC TGG GTA ATG GAG CGA TCA GGA GAG TTG GGC    2928

 977  Ala Trp Gln Val Leu Gln Pro Thr Ser Val Asp Pro Tyr Thr Tyr Gly     992
2929  GCG TGG CAG GTG CTA CAA CCG ACA TCT GTT GAT CCA TAT ACT TAT GGA    2976

 993  Ser Trp Asp Glu Val Thr Asp Gln Gly Lys Leu Gln Ala Thr Phe Asp    1008
```

```
2977  AGT TGG GAT GAA GTA ACT GAC CAA GGT AAA CTG CAA GCA ACC TTC GAC    3024

1009  Pro Ser Gly Arg Asn Val Tyr Ile Val Ala Gly Gly Ser Ser Asn Leu    1024
3025  CCA AGT GGA CGC AAT GTG TAT ATT GTC GCG GGT GGG TCT AGC AAC CTC    3072

1025  Pro Thr Val Ala Ser His Leu Met Thr Leu Gln Thr Leu Leu Met Leu    1040
3073  CCC ACG GTA GCC AGC CAC CTC ATG ACG CTT CAG ACT TTA TTG ATG CTG    3120

1041  Arg Glu Thr Gly Arg Asp Thr Ile Phe Val Cys Leu Ala Met Pro Gly    1056
3121  CGG GAA ACT GGT CGT GAC ACA ATC TTT GTC TGT CTG GCA ATG CCA GGT    3168

1057  Tyr Ala Lys Tyr His Ser Gln Thr Gly Lys Ser Pro Glu Glu Tyr Ile    1072
3169  TAT GCC AAA TAC CAC AGT CAA ACA GGA AAA TCG CCG GAA GAG TAT ATT    3216

1073  Ala Leu Thr Asp Glu Asp Gly Ile Pro Met Asp Glu Phe Ile Ala Leu    1088
3217  GCG CTG ACT GAC GAG GAT GGT ATC CCA ATG GAC GAG TTT ATT GCA CTT    3264

1089  Ser Val Tyr Asp Trp Pro Val Thr Pro Ser Phe Arg Val Leu Arg Asp    1104
3265  TCT GTC TAC GAC TGG CCT GTT ACC CCA TCG TTT CGT GTT CTG CGA GAC    3312

1105  Gly Tyr Pro Pro Asp Arg Asp Ser Gly Gly His Ala Val Ser Thr Val    1120
3313  GGT TAT CCA CCT GAT CGA GAT TCT GGT GGT CAC GCA GTT AGT ACG GTT    3360

1121  Phe Gln Leu Asn Asp Phe Asp Gly Ala Ile Glu Glu Thr Tyr Arg Arg    1136
3361  TTC CAG CTC AAT GAT TTC GAT GGA GCG ATC GAA GAA ACA TAT CGT CGT    3408

1137  Ile Ile Arg His Ala Asp Val Leu Gly Leu Glu Arg Gly ***            1150
3409  ATT ATC CGC CAT GCC GAT GTC CTT GGT CTC GAA AGA GGC TAA            3450
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : CYNI.SEQ       SEQUENCE : NORMAL   1437 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -   1437

    TRANSLATION REGION :      1 -   1437


*** DNA TRANSLATION ***

   1  Met Asn Lys Lys Gln Val Asp Thr Leu Leu Ile His Ala His Leu Phe      16
   1  ATG AAT AAA AAA CAG GTA GAC ACA TTG TTA ATA CAC GCT CAT CTT TTT      48

  17  Thr Met Gln Gly Asn Gly Leu Gly Tyr Ile Ala Asp Gly Ala Ile Ala      32
  49  ACC ATG CAG GGC AAT GGC CTG GGA TAT ATT GCC GAT GGG GCA ATT GCG      96

  33  Val Gln Gly Ser Gln Ile Val Ala Val Asp Ser Thr Glu Ala Leu Leu      48
  97  GTT CAG GGT AGC CAG ATC GTA GCA GTG GAT TCG ACA GAG GCT TTG CTG     144

  49  Ser His Phe Glu Gly Asn Lys Thr Ile Asn Ala Val Asn Cys Ala Val      64
 145  AGT CAT TTT GAA GGA AAT AAA ACA ATT AAT GCG GTA AAT TGT GCA GTG     192

  65  Leu Pro Gly Leu Ile Asp Ala His Ile His Thr Thr Cys Ala Ile Leu      80
 193  TTG CCT GGA CTA ATT GAT GCT CAT ATA CAT ACG ACT TGT GCT ATT CTG     240

  81  Arg Gly Val Ala Gln Asp Val Thr Asn Trp Leu Met Asp Ala Thr Ile      96
 241  CGT GGA GTG GCA CAG GAT GTA ACC AAT TGG CTA ATG GAC GCG ACA ATT     288

  97  Pro Tyr Ala Leu Gln Met Thr Pro Ala Val Asn Ile Ala Gly Thr Arg     112
 289  CCT TAT GCA CTT CAG ATG ACA CCC GCA GTA AAT ATA GCC GGA ACG CGC     336

 113  Leu Ser Val Leu Glu Gly Leu Lys Ala Gly Thr Thr Thr Phe Gly Asp     128
 337  TTG AGT GTA CTC GAA GGG CTG AAA GCA GGA ACA ACC ACA TTC GGC GAT     384

 129  Ser Glu Thr Pro Tyr Pro Leu Trp Gly Glu Phe Phe Asp Glu Ile Gly     144
 385  TCT GAG ACT CCT TAC CCG CTC TGG GGA GAG TTT TTC GAT GAA ATT GGG     432

 145  Val Arg Ala Ile Leu Ser Pro Ala Phe Asn Ala Phe Pro Leu Glu Trp     160
 433  GTA CGT GCT ATT CTA TCC CCT GCC TTT AAC GCC TTT CCA CTA GAA TGG     480

 161  Ser Ala Trp Lys Glu Gly Asp Leu Tyr Pro Phe Asp Met Lys Ala Gly     176
 481  TCG GCA TGG AAG GAG GGA GAC CTC TAT CCC TTC GAT ATG AAG GCA GGA     528

 177  Arg Arg Gly Met Glu Glu Ala Val Asp Phe Ala Cys Ala Trp Asn Gly     192
 529  CGA CGT GGT ATG GAA GAG GCT GTG GAT TTT GCT TGT GCA TGG AAT GGA     576

 193  Ala Ala Glu Gly Arg Ile Thr Thr Met Leu Gly Leu Gln Ala Ala Asp     208
 577  GCC GCA GAG GGA CGT ATC ACC ACT ATG TTG GGA CTA CAG GCG GCG GAT     624

 209  Met Leu Pro Leu Glu Ile Leu His Ala Ala Lys Glu Ile Ala Gln Arg     224
 625  ATG CTA CCA CTG GAG ATC CTA CAC GCA GCT AAA GAG ATT GCC CAA CGG     672

 225  Glu Gly Leu Met Leu His Ile His Val Ala Gln Gly Asp Arg Glu Thr     240
 673  GAA GGC TTA ATG CTG CAT ATT CAT GTG GCC CAG GGA GAT CGA GAA ACA     720

 241  Lys Gln Ile Val Lys Arg Tyr Gly Lys Arg Pro Ile Ala Phe Leu Ala     256
 721  AAA CAA ATT GTC AAA CGA TAT GGT AAG CGT CCG ATC GCA TTT CTA GCT     768

 257  Glu Ile Gly Tyr Leu Asp Glu Gln Leu Leu Ala Val His Leu Thr Asp     272
 769  GAA ATT GGC TAC TTG GAC GAA CAG TTG CTG GCA GTT CAC CTC ACC GAT     816
```

```
 273  Ala Thr Asp Glu Glu Val Ile Gln Val Ala Lys Ser Gly Ala Gly Met     288
 817  GCC ACA GAT GAA GAA GTG ATA CAA GTA GCC AAA AGT GGT GCT GGC ATG     864

 289  Ala Leu Cys Ser Gly Ala Ile Gly Ile Ile Asp Gly Leu Val Pro Pro     304
 865  GCA CTC TGT TCG GGC GCT ATT GGC ATC ATT GAC GGT CTT GTT CCG CCC     912

 305  Ala His Val Phe Arg Gln Ala Gly Gly Ser Val Ala Leu Gly Ser Asp     320
 913  GCT CAT GTT TTT CGA CAA GCA GGC GGT TCC GTT GCA CTC GGT TCT GAT     960

 321  Gln Ala Cys Gly Asn Asn Cys Cys Asn Ile Phe Asn Glu Met Lys Leu     336
 961  CAA GCC TGT GGC AAC AAC TGT TGT AAC ATC TTC AAT GAA ATG AAG CTG    1008

 337  Thr Ala Leu Phe Asn Lys Ile Lys Tyr His Asp Pro Thr Ile Met Pro     352
1009  ACC GCC TTA TTC AAC AAA ATA AAA TAT CAT GAT CCA ACC ATT ATG CCG    1056

 353  Ala Trp Glu Val Leu Arg Met Ala Thr Ile Glu Gly Ala Gln Ala Ile     368
1057  GCT TGG GAA GTC CTG CGT ATG GCT ACC ATC GAA GGA GCG CAG GCG ATT    1104

 369  Gly Leu Asp His Lys Ile Gly Ser Leu Gln Val Gly Lys Glu Ala Asp     384
1105  GGT TTA GAT CAC AAG ATT GGC TCT CTT CAA GTG GGC AAA GAA GCC GAC    1152

 385  Leu Ile Leu Ile Asp Leu Ser Ser Pro Asn Leu Ser Pro Thr Leu Leu     400
1153  CTG ATC TTA ATA GAC CTC AGT TCC CCT AAC CTC TCG CCC ACC CTG CTC    1200

 401  Asn Pro Ile Arg Asn Leu Val Pro Asn Leu Val Tyr Ala Ala Ser Gly     416
1201  AAC CCT ATT CGT AAC CTT GTA CCT AAC TTG GTG TAT GCT GCT TCA GGA    1248

 417  His Glu Val Lys Ser Val Met Val Ala Gly Lys Leu Leu Val Glu Asp     432
1249  CAT GAA GTT AAA AGC GTC ATG GTG GCG GGA AAA CTT TTA GTG GAA GAC    1296

 433  Tyr Gln Val Leu Thr Val Asp Glu Ser Ala Ile Leu Ala Glu Ala Gln     448
1297  TAC CAA GTC CTC ACG GTA GAT GAG TCC GCT ATT CTC GCT GAA GCG CAA    1344

 449  Val Gln Ala Gln Gln Leu Cys Gln Arg Val Thr Ala Asp Pro Ile His     464
1345  GTA CAA GCT CAA CAA CTC TGC CAA CGT GTG ACC GCT GAC CCC ATT CAC    1392

 465  Lys Lys Met Val Leu Met Glu Ala Met Ala Lys Gly Lys Leu ***         479
1393  AAA AAG ATG GTG TTA ATG GAA GCG ATG GCT AAG GGT AAA TTA TAG        1437
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : CYNH.SEQ       SEQUENCE : NORMAL    858 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -    858

    TRANSLATION REGION :      1 -    858


*** DNA TRANSLATION ***

   1  Met Asn Gln Thr Trp Arg Gly Lys Pro Met Thr Ile Tyr Glu Asn Lys      16
   1  ATG AAT CAA ACC TGG AGG GGC AAA CCA ATG ACC ATA TAT GAA AAT AAG      48

  17  Leu Ser Ser Tyr Gln Lys Asn Gln Asp Ala Ile Ile Ser Ala Lys Glu      32
  49  TTG AGT AGT TAT CAA AAA AAT CAA GAT GCC ATA ATA TCT GCA AAA GAA      96

  33  Leu Glu Glu Trp His Leu Ile Gly Leu Leu Asp His Ser Ile Asp Ala      48
  97  CTC GAA GAA TGG CAT TTA ATT GGA CTT CTA GAC CAT TCA ATA GAT GCG     144

  49  Val Ile Val Pro Asn Tyr Phe Leu Glu Gln Glu Cys Met Thr Ile Ser      64
 145  GTA ATA GTA CCG AAT TAT TTT CTT GAG CAA GAG TGT ATG ACA ATT TCA     192

  65  Glu Arg Ile Lys Lys Ser Lys Tyr Phe Ser Ala Tyr Pro Gly His Pro      80
 193  GAG AGA ATA AAA AAG AGT AAA TAT TTT AGC GCT TAT CCC GGT CAT CCA     240

  81  Ser Val Ser Ser Leu Gly Gln Glu Leu Tyr Glu Cys Glu Ser Glu Leu      96
 241  TCA GTA AGT AGC TTG GGA CAA GAG TTG TAT GAA TGC GAA AGT GAG CTT     288

  97  Glu Leu Ala Lys Tyr Gln Glu Asp Ala Pro Thr Leu Ile Lys Glu Met     112
 289  GAA TTA GCA AAG TAT CAA GAA GAC GCA CCC ACA TTG ATT AAA GAA ATG     336

 113  Arg Arg Leu Val His Pro Tyr Ile Ser Pro Ile Asp Arg Leu Arg Val     128
 337  CGG AGG CTG GTA CAT CCG TAC ATA AGT CCA ATT GAT AGA CTT AGG GTT     384

 129  Glu Val Asp Asp Ile Trp Ser Tyr Gly Cys Asn Leu Ala Lys Leu Gly     144
 385  GAA GTT GAT GAT ATT TGG AGT TAT GGC TGT AAT TTA GCA AAA CTT GGT     432

 145  Asp Lys Lys Leu Phe Ala Gly Ile Val Arg Glu Phe Lys Glu Asp Asn     160
 433  GAT AAA AAA CTG TTT GCG GGT ATC GTT AGA GAG TTT AAA GAA GAT AAC     480

 161  Pro Gly Ala Pro His Cys Asp Val Met Ala Trp Gly Phe Leu Glu Tyr     176
 481  CCT GGC GCA CCA CAT TGT GAC GTA ATG GCA TGG GGT TTT CTC GAA TAT     528

 177  Tyr Lys Asp Lys Pro Asn Ile Ile Asn Gln Ile Ala Ala Asn Val Tyr     192
 529  TAT AAA GAT AAA CCA AAT ATC ATA AAT CAA ATC GCA GCA AAT GTA TAT     576

 193  Leu Lys Thr Ser Ala Ser Gly Gly Glu Ile Val Leu Trp Asp Glu Trp     208
 577  TTA AAA ACG TCT GCA TCA GGA GGA GAA ATA GTG CTT TGG GAT GAA TGG     624

 209  Pro Thr Gln Ser Glu Tyr Ile Ala Tyr Lys Thr Asp Asp Pro Ala Ser     224
 625  CCA ACT CAA AGC GAA TAT ATA GCA TAC AAA ACA GAT GAT CCA GCT AGT     672

 225  Phe Gly Leu Asp Ser Lys Lys Ile Ala Gln Pro Lys Leu Glu Ile Gln     240
 673  TTC GGT CTT GAT AGC AAA AAG ATC GCA CAA CCA AAA CTT GAG ATC CAA     720

 241  Pro Asn Gln Gly Asp Leu Ile Leu Phe Asn Ser Met Arg Ile His Ala     256
 721  CCG AAC CAG GGA GAT TTA ATT CTA TTC AAT TCC ATG AGA ATT CAT GCG     768

 257  Val Lys Lys Ile Glu Thr Gly Val Arg Met Thr Trp Gly Cys Leu Ile     272
 769  GTG AAA AAG ATA GAA ACT GGT GTA CGT ATG ACA TGG GGA TGT TTG ATT     816
```

```
273  Gly Tyr Ser Gly Thr Asp Lys Pro Leu Val Ile Trp Thr ***          286
817  GGA TAC TCT GGA ACT GAT AAA CCG CTT GTT ATT TGG ACT TAA          858
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : CYNL.SEQ       SEQUENCE : NORMAL   1398 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -   1398

    TRANSLATION REGION :      1 -   1398


*** DNA TRANSLATION ***

   1  Met Ser Leu Thr Phe Pro Ser Gln Tyr Asp Phe Ile Pro Arg Tyr Phe      16
   1  ATG AGC CTG ACA TTT CCA AGC CAG TAT GAC TTT ATA CCC CGC TAC TTC      48

  17  Arg Leu Ala Phe Thr Asn Val Leu Ser Asn Ile Ile Val Pro Leu Ser      32
  49  CGA CTA GCT TTT ACC AAC GTC CTA TCG AAT ATT ATA GTA CCA CTG AGT      96

  33  Asn Leu Val Ser Ile Met Phe Leu Gly His Leu Ser Glu Ile His Tyr      48
  97  AAT TTA GTC AGT ATC ATG TTC TTA GGT CAT CTC TCG GAA ATC CAC TAC     144

  49  Leu Ala Gly Val Ala Leu Ala Gly Asn Leu Leu Asn Phe Leu Tyr Phe      64
 145  CTA GCT GGA GTA GCC CTT GCT GGA AAC CTG CTT AAC TTT CTC TAC TTT     192

  65  Val Leu Ser Phe Leu Arg Met Gly Thr Thr Ala Leu Thr Ala Gln Ala      80
 193  GTT TTA TCG TTT TTA CGA ATG GGT ACC ACT GCG TTA ACA GCA CAA GCA     240

  81  Val Gly Arg Asp Asp Arg Glu Gly Val Leu Leu Ala Gly Leu Leu Asn      96
 241  GTC GGA CGA GAT GAC CGA GAG GGC GTA CTG TTG GCG GGA CTG CTT AAT     288

  97  Gly Leu Ile Ala Leu Val Leu Gly Val Ala Ile Ile Leu Leu Gln Tyr     112
 289  GGT TTA ATT GCT CTG GTG CTG GGG GTG GCG ATC ATA CTG TTA CAG TAC     336

 113  Pro Leu Gly Val Leu Gly Phe Ser Ile Leu Asp Val Ala Pro Glu Val     128
 337  CCG TTA GGG GTG TTG GGG TTT TCT ATA CTA GAT GTC GCC CCA GAG GTG     384

 129  Lys Ser Ser Ala Gln Ala Tyr Phe Asn Thr Gln Ser Trp Gly Ala Pro     144
 385  AAG TCT TCA GCT CAA GCG TAT TTT AAC ACC CAG AGT TGG GGA GCG CCT     432

 145  Ala Ile Leu Leu Asn Phe Val Leu Leu Gly Trp Phe Leu Gly Arg Glu     160
 433  GCC ATT TTG CTC AAC TTT GTC CTG CTT GGC TGG TTT CTG GGA CGG GAA     480

 161  Lys Asn Gly Leu Val Val Leu Leu Ser Val Val Gly Asn Ala Ala Asn     176
 481  AAA AAT GGC TTA GTT GTG TTG TTG TCG GTT GTG GGC AAT GCT GCC AAT     528

 177  Ile Ala Leu Asp Tyr Leu Phe Ile Ile His Leu Asp Trp Ser Ser Met     192
 529  ATC GCA CTC GAC TAC TTG TTT ATT ATC CAC TTA GAC TGG TCG AGC ATG     576

 193  Gly Ala Gly Val Ala Ser Ala Ala Ser Gln Tyr Leu Thr Leu Leu Val     208
 577  GGG GCA GGA GTA GCT TCT GCT GCA AGT CAA TAT CTC ACC CTA TTG GTG     624

 209  Gly Leu Phe Phe Val Cys Lys Glu Ile Pro Trp Gln Glu Val Lys Glu     224
 625  GGA TTG TTT TTT GTC TGC AAA GAA ATC CCG TGG CAA GAG GTA AAA GAA     672

 225  Leu Ala Gly Lys Ile Trp Asp Ile Lys Ala Ile Arg Ser Thr Leu Thr     240
 673  TTA GCT GGA AAA ATT TGG GAC ATA AAA GCT ATA CGC TCC ACC TTA ACT     720

 241  Leu Asn Gly Asn Ile Phe Val Ser Asn Leu Val Cys Met Phe Thr Asn     256
 721  CTC AAT GGA AAC ATC TTT GTC AGC AAT TTA GTT TGT ATG TTT ACC AAC     768

 257  Leu Thr Phe Asn Tyr Glu Gly Ala Gln Met Gly Thr Ile Ile Tyr Ala     272
 769  TTA ACG TTT AAC TAC GAA GGG GCA CAA ATG GGG ACA ATT ATC TAC GCT     816
```

```
 273  Gln Asn Ala Leu Leu Leu Gln Ile Trp Ala Leu Ser Ile Tyr Ile Val     288
 817  CAA AAT GCC TTG CTC TTA CAG ATA TGG GCT TTG AGC ATC TAC ATT GTT     864

 289  Glu Gly Leu Gly Phe Cys Thr Glu Thr Leu Ala Gly Asn Phe Lys Gly     304
 865  GAA GGA TTG GGA TTC TGT ACA GAA ACT CTG GCA GGA AAT TTT AAA GGG     912

 305  Lys Gly Ala Thr Glu Gln Leu Ala Pro Leu Ala Lys Val Ser Val Gly     320
 913  AAA GGA GCC ACA GAA CAG TTA GCA CCT CTG GCA AAA GTT TCT GTG GGA     960

 321  Met Ala Leu Leu Val Ser Leu Val Phe Gly Gly Ala Cys Trp Leu Phe     336
 961  ATG GCT CTG CTA GTG AGT CTG GTT TTC GGC GGA GCG TGT TGG CTA TTT    1008

 337  Pro Asp Asp Val Phe Gly Leu Leu Thr Asn His Ala Glu Ile Ile Glu     352
1009  CCC GAT GAT GTT TTT GGG TTG TTA ACC AAC CAC GCA GAA ATT ATT GAG    1056

 353  Asn Ile Asp Ile Phe Ile Pro Trp Leu Val Leu Val Leu Val Cys Asn     368
1057  AAT ATT GAT ATT TTT ATT CCC TGG TTG GTA CTT GTC CTA GTA TGC AAT    1104

 369  Ser Val Cys Cys Ile Leu Cys Gly Tyr Phe Leu Gly Leu Ala Gln Gly     384
1105  TCA GTA TGT TGC ATA CTA TGT GGG TAC TTC TTG GGT TTG GCT CAA GGG    1152

 385  His Thr Leu Arg Asn Val Asn Leu Ile Ala Phe Gly Val Gly Phe Leu     400
1153  CAT ACT CTC CGC AAT GTT AAC TTA ATC GCG TTT GGT GTA GGA TTT TTA    1200

 401  Pro Ala Asp Val Ala Ala Val Lys Phe Glu Ser Asn His Ile Leu Trp     416
1201  CCT GCG GAT GTA GCT GCT GTT AAA TTT GAG AGT AAC CAT ATC TTG TGG    1248

 417  Leu Ala Phe Phe Leu Phe Asn Ala Ile Gly Met Leu Met Phe Gly Ala     432
1249  TTG GCT TTC TTT TTG TTC AAC GCC ATT GGA ATG CTG ATG TTT GGG GCA    1296

 433  Gln Leu Pro Arg Thr Phe Glu Ser Val Val Glu Asn Asp Ser Val Ser     448
1297  CAG TTA CCT AGG ACA TTT GAG AGT GTA GTT GAG AAT GAT AGT GTA TCT    1344

 449  Ile Pro Ala Leu Glu Ala Ser Arg Ala Leu Thr Gln Met Glu Thr Leu     464
1345  ATT CCA GCT TTA GAA GCT TCT CGT GCC TTG ACT CAA ATG GAA ACG CTA    1392

 465  His ***                                                           466
1393  CAT TAA                                                           1398
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : ORF3.SEQ       SEQUENCE : NORMAL    750 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -    750

    TRANSLATION REGION :      1 -    750


*** DNA TRANSLATION ***

   1  Met Leu Asn Leu Asp Arg Ile Leu Asn Gln Glu Arg Leu Leu Arg Glu      16
   1  ATG TTG AAC TTA GAC CGC ATC CTG AAT CAA GAG CGA CTG CTA CGA GAA      48

  17  Met Thr Gly Leu Asn Arg Gln Ala Phe Asn Glu Leu Leu Ser Gln Phe      32
  49  ATG ACT GGA CTT AAC CGC CAA GCA TTC AAC GAG CTG TTA TCT CAG TTT      96

  33  Ala Asp Thr Tyr Glu Arg Thr Val Phe Asn Ser Leu Ala Asn Arg Lys      48
  97  GCT GAT ACC TAT GAA CGC ACC GTG TTC AAC TCC TTA GCA AAC CGC AAA     144

  49  Arg Ala Pro Gly Gly Gly Arg Lys Pro Thr Leu Arg Ser Ile Glu Glu      64
 145  CGT GCG CCC GGG GGC GGA CGC AAG CCT ACA CTC AGA AGT ATA GAG GAA     192

  65  Lys Leu Phe Tyr Ile Leu Leu Tyr Cys Lys Cys Tyr Pro Thr Phe Asp      80
 193  AAA CTA TTT TAT ATC CTG CTG TAC TGC AAA TGT TAT CCG ACG TTT GAC     240

  81  Leu Leu Ser Val Leu Phe Asn Phe Asp Arg Ser Cys Ala His Asp Trp      96
 241  TTG CTG AGT GTG TTG TTC AAC TTT GAC CGC TCC TGT GCT CAT GAT TGG     288

  97  Val His Arg Leu Leu Ser Val Leu Glu Thr Thr Leu Gly Glu Lys Gln     112
 289  GTA CAT CGA CTA CTG TCT GTG CTA GAA ACC ACT TTA GGA GAA AAG CAA     336

 113  Val Leu Pro Ala Arg Lys Leu Arg Ser Met Glu Glu Phe Thr Lys Arg     128
 337  GTT TTG CCA GCA CGC AAA CTC AGG AGC ATG GAG GAA TTC ACC AAA AGG     384

 129  Phe Pro Asp Val Lys Glu Val Ile Val Asp Gly Thr Glu Arg Pro Val     144
 385  TTT CCA GAT GTG AAG GAG GTG ATT GTG GAT GGT ACG GAG CGT CCA GTC     432

 145  Gln Arg Pro Gln Asn Arg Glu Arg Gln Lys Glu Tyr Tyr Ser Gly Lys     160
 433  CAG CGT CCT CAA AAC CGA GAA CGC CAA AAA GAG TAT TAC TCT GGC AAG     480

 161  Lys Lys Arg His Thr Cys Lys Gln Ile Thr Val Ser Thr Arg Glu Lys     176
 481  AAA AAG CGG CAT ACA TGC AAG CAG ATT ACA GTC AGC ACA AGG GAG AAA     528

 177  Arg Val Ile Ile Arg Thr Glu Thr Arg Ala Gly Lys Val His Asp Lys     192
 529  CGA GTG ATT ATT CGG ACG GAA ACC AGA GCA GGT AAA GTG CAT GAC AAA     576

 193  Arg Leu Leu His Glu Ser Glu Ile Val Gln Tyr Ile Pro Asp Glu Val     208
 577  CGG CTA CTC CAT GAA TCA GAG ATA GTG CAA TAC ATT CCT GAT GAA GTA     624

 209  Ala Ile Glu Gly Asp Leu Gly Phe His Gly Leu Glu Lys Glu Phe Val     224
 625  GCA ATA GAG GGA GAT TTG GGT TTT CAT GGG TTG GAG AAA GAA TTT GTC     672

 225  Asn Val His Leu Pro His Lys Lys Pro Lys Gly Ile Glu Ala Arg Arg     240
 673  AAT GTC CAT TTA CCA CAC AAG AAA CCG AAA GGT ATC GAA GCA AGG AGG     720

 241  His Gly Gly Gly Met Gly Gln Phe Leu ***                           250
 721  CAT GGC GGC GGG ATG GGT CAG TTT TTA TAA                           750
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : CYNF.SEQ       SEQUENCE : NORMAL   1431 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -   1431

    TRANSLATION REGION :      1 -   1431


*** DNA TRANSLATION ***

   1  Met Asn Leu Ile Thr Thr Lys Lys Gln Val Asp Thr Leu Val Ile His      16
   1  ATG AAT CTT ATA ACA ACA AAA AAA CAG GTA GAT ACA TTA GTG ATA CAC      48

  17  Ala His Leu Phe Thr Met Gln Gly Asn Gly Val Gly Tyr Ile Ala Asp      32
  49  GCT CAT CTT TTT ACC ATG CAG GGA AAT GGT GTG GGA TAT ATT GCA GAT      96

  33  Gly Ala Leu Ala Val Glu Gly Ser Arg Ile Val Ala Val Asp Ser Thr      48
  97  GGG GCA CTT GCG GTT GAG GGT AGC CGT ATT GTA GCA GTT GAT TCG ACG     144

  49  Glu Ala Leu Leu Ser His Phe Glu Gly Arg Lys Val Ile Glu Ser Ala      64
 145  GAG GCG TTG CTG AGT CAT TTT GAG GGC AGA AAG GTT ATT GAG TCC GCG     192

  65  Asn Cys Ala Val Leu Pro Gly Leu Ile Asn Ala His Val Asp Thr Ser      80
 193  AAT TGT GCC GTC TTG CCT GGG CTG ATT AAT GCT CAC GTA GAC ACA AGT     240

  81  Leu Val Leu Met Arg Gly Ala Ala Gln Asp Val Thr Asn Trp Leu Met      96
 241  TTG GTG CTG ATG CGT GGG GCG GCG CAA GAT GTA ACT AAT TGG CTA ATG     288

  97  Asp Ala Thr Met Pro Tyr Phe Ala His Met Thr Pro Val Ala Ser Met     112
 289  GAC GCG ACC ATG CCT TAT TTT GCT CAC ATG ACA CCC GTG GCG AGT ATG     336

 113  Ala Ala Thr Arg Leu Arg Val Val Glu Glu Leu Lys Ala Gly Thr Thr     128
 337  GCT GCA ACA CGC TTA AGG GTG GTA GAA GAG TTG AAA GCA GGC ACA ACA     384

 129  Thr Phe Cys Asp Asn Lys Ile Ile Ser Pro Leu Trp Gly Glu Phe Phe     144
 385  ACA TTC TGT GAC AAT AAA ATT ATT AGC CCC CTG TGG GGC GAA TTT TTC     432

 145  Asp Glu Ile Gly Val Arg Ala Ser Leu Ala Pro Met Phe Asp Ala Leu     160
 433  GAT GAA ATT GGT GTA CGG GCT AGT TTA GCT CCT ATG TTC GAT GCA CTC     480

 161  Pro Leu Glu Met Pro Pro Leu Gln Asp Gly Glu Leu Tyr Pro Phe Asp     176
 481  CCA CTG GAG ATG CCA CCG CTT CAA GAC GGG GAG CTT TAT CCC TTC GAT     528

 177  Ile Lys Ala Gly Arg Arg Ala Met Ala Glu Ala Val Asp Phe Ala Cys     192
 529  ATC AAG GCG GGA CGG CGG GCG ATG GCA GAG GCT GTG GAT TTT GCC TGT     576

 193  Gly Trp Asn Gly Ala Ala Glu Gly Arg Ile Thr Thr Met Leu Gly Met     208
 577  GGG TGG AAT GGG GCA GCA GAG GGG CGT ATC ACT ACC ATG TTA GGA ATG     624

 209  Tyr Ser Pro Asp Met Met Pro Leu Glu Met Leu Arg Ala Ala Lys Glu     224
 625  TAT TCG CCA GAT ATG ATG CCG CTT GAG ATG CTA CGC GCA GCC AAA GAG     672

 225  Ile Ala Gln Arg Glu Gly Leu Met Leu His Phe His Val Ala Gln Gly     240
 673  ATT GCT CAA CGG GAA GGC TTA ATG CTG CAT TTT CAT GTA GCG CAG GGA     720

 241  Asp Arg Glu Thr Glu Gln Ile Val Lys Arg Tyr Gly Lys Arg Pro Ile     256
 721  GAT CGG GAA ACA GAG CAA ATC GTT AAA CGA TAT GGT AAG CGT CCG ATC     768

 257  Ala Phe Leu Ala Glu Ile Gly Tyr Leu Asp Glu Gln Leu Leu Ala Val     272
 769  GCA TTT CTA GCT GAG ATT GGC TAC TTG GAC GAA CAG TTG CTG GCA GTT     816
```

```
 273  His Leu Thr Asp Ala Thr Asp Glu Glu Val Ile Gln Val Ala Lys Ser     288
 817  CAC CTC ACC GAT GCC ACC GAT GAA GAG GTG ATA CAA GTA GCC AAA AGT     864

 289  Gly Ala Gly Met Val Leu Cys Ser Gly Met Ile Gly Thr Ile Asp Gly     304
 865  GGC GCT GGC ATG GTA CTC TGT TCG GGA ATG ATT GGC ACT ATT GAC GGT     912

 305  Ile Val Pro Pro Ala His Val Phe Arg Gln Ala Gly Gly Pro Val Ala     320
 913  ATC GTG CCG CCC GCT CAT GTG TTT CGG CAA GCA GGC GGA CCC GTT GCG     960

 321  Leu Gly Ser Ser Tyr Asn Asn Ile Phe His Glu Met Lys Leu Thr Ala     336
 961  CTA GGC AGC AGC TAC AAT AAT ATT TTC CAT GAG ATG AAG CTG ACC GCC    1008

 337  Leu Phe Asn Lys Ile Lys Tyr His Asp Pro Thr Ile Met Pro Ala Trp     352
1009  TTA TTC AAC AAA ATA AAA TAT CAC GAT CCA ACC ATT ATG CCG GCT TGG    1056

 353  Glu Val Leu Arg Met Ala Thr Ile Glu Gly Ala Arg Ala Ile Gly Leu     368
1057  GAA GTC CTG CGT ATG GCT ACC ATC GAA GGA GCG CGG GCG ATT GGT TTA    1104

 369  Asp His Lys Ile Gly Ser Leu Glu Val Gly Lys Glu Ala Asp Leu Ile     384
1105  GAT CAC AAG ATT GGC TCT CTT GAA GTT GGC AAA GAA GCC GAC CTG ATC    1152

 385  Leu Ile Asp Leu Ser Thr Pro Asn Leu Ser Pro Thr Leu Leu Asn Pro     400
1153  TTA ATA GAC CTC AGC ACC CCT AAC CTC TCA CCC ACT CTG CTT AAC CCC    1200

 401  Ile Arg Asn Leu Val Pro Asn Phe Val Tyr Ala Ala Ser Gly His Glu     416
1201  ATT CGT AAC CTT GTA CCT AAT TTC GTG TAC GCT GCT TCA GGA CAT GAA    1248

 417  Val Lys Ser Val Met Val Ala Gly Lys Leu Leu Leu Glu Asp Tyr Gln     432
1249  GTT AAA AGT GTC ATG GTG GCG GGA AAA CTG TTA TTG GAA GAC TAC CAA    1296

 433  Val Leu Thr Val Asp Glu Ser Ala Ile Ile Ala Glu Ala Gln Leu Gln     448
1297  GTC CTC ACA GTA GAT GAG TCT GCT ATC ATT GCT GAA GCA CAA TTG CAA    1344

 449  Ala Gln Gln Ile Ser Gln Cys Val Ala Ser Asp Pro Ile His Lys Lys     464
1345  GCC CAA CAG ATT TCT CAA TGC GTA GCA TCT GAC CCT ATC CAC AAA AAA    1392

 465  Met Val Leu Met Ala Ala Met Ala Arg Gly Gln Leu ***                 477
1393  ATG GTG CTG ATG GCG GCG ATG GCA AGG GGC CAA TTG TAG                 1431
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : CYNE.SEQ       SEQUENCE : NORMAL    780 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -    780

    TRANSLATION REGION :      1 -    780


*** DNA TRANSLATION ***

   1  Met Gln Glu Lys Arg Ile Ala Met Trp Ser Val Pro Arg Ser Leu Gly      16
   1  ATG CAA GAA AAA CGA ATC GCA ATG TGG TCT GTG CCA CGA AGT TTG GGT      48

  17  Thr Val Leu Leu Gln Ala Trp Ser Ser Arg Pro Asp Thr Val Val Phe      32
  49  ACA GTG CTG CTA CAA GCC TGG TCG AGT CGG CCA GAT ACC GTA GTC TTT      96

  33  Asp Glu Leu Leu Ser Phe Pro Tyr Leu Phe Ile Lys Gly Lys Asp Met      48
  97  GAT GAA CTT CTC TCC TTT CCC TAT CTC TTT ATC AAA GGG AAA GAT ATG     144

  49  Gly Phe Thr Trp Thr Asp Leu Asp Ser Ser Gln Met Pro His Ala Asp      64
 145  GGC TTT ACT TGG ACA GAC CTT GAT TCT AGC CAA ATG CCC CAC GCA GAT     192

  65  Trp Arg Ser Val Ile Asp Leu Leu Lys Ala Pro Leu Pro Glu Gly Lys      80
 193  TGG CGA TCC GTC ATC GAT CTG TTA AAG GCT CCC CTG CCT GAA GGG AAA     240

  81  Ser Ile Ile Asp Leu Leu Lys Ala Pro Leu Pro Glu Gly Lys Ser Ile      96
 241  TCA ATC ATC GAT CTG TTA AAG GCT CCC CTG CCT GAA GGG AAA TCA ATT     288

  97  Cys Tyr Gln Lys His Gln Ala Tyr His Leu Ile Glu Glu Thr Met Gly     112
 289  TGC TAT CAG AAG CAT CAA GCG TAT CAT TTA ATC GAA GAG ACC ATG GGG     336

 113  Ile Glu Trp Ile Leu Pro Phe Ser Asn Cys Phe Leu Ile Arg Gln Pro     128
 337  ATT GAG TGG ATA TTG CCC TTC AGC AAC TGC TTT CTG ATT CGC CAA CCC     384

 129  Lys Glu Met Leu Leu Ser Phe Arg Lys Ile Val Pro His Phe Thr Phe     144
 385  AAA GAA ATG CTC TTA TCT TTT CGT AAG ATT GTG CCA CAT TTT ACC TTT     432

 145  Glu Glu Thr Gly Trp Ile Glu Leu Lys Arg Leu Phe Asp Tyr Val His     160
 433  GAA GAA ACA GGC TGG ATC GAA TTA AAA CGG CTG TTT GAC TAT GTA CAT     480

 161  Gln Thr Ser Gly Val Ile Pro Pro Val Ile Asp Ala His Asp Leu Leu     176
 481  CAA ACG AGC GGA GTA ATC CCG CCT GTC ATA GAT GCA CAC GAC TTG CTG     528

 177  Asn Asp Pro Arg Arg Met Leu Ser Lys Leu Cys Gln Val Val Gly Val     192
 529  AAC GAT CCG CGG AGA ATG CTC TCC AAG CTT TGT CAG GTT GTA GGG GTT     576

 193  Glu Phe Thr Glu Thr Met Leu Ser Trp Pro Pro Met Glu Val Glu Leu     208
 577  GAG TTT ACC GAG ACA ATG CTC AGT TGG CCC CCC ATG GAG GTC GAG TTG     624

 209  Asn Glu Lys Leu Ala Pro Trp Tyr Ser Thr Val Ala Ser Ser Thr His     224
 625  AAC GAA AAA CTA GCC CCT TGG TAC AGC ACC GTA GCA AGT TCT ACG CAT     672

 225  Phe His Ser Tyr Gln Asn Lys Asn Glu Ser Leu Pro Leu Tyr Leu Val     240
 673  TTT CAC TCG TAT CAG AAT AAA AAT GAG TCG TTG CCG CTA TAT CTT GTC     720

 241  Asp Ile Cys Lys Arg Cys Asp Glu Ile Tyr Gln Glu Leu Tyr Gln Phe     256
 721  GAT ATT TGT AAA CGC TGC GAT GAA ATA TAT CAG GAA TTA TAT CAA TTT     768

 257  Arg Leu Tyr ***                                                  260
 769  CGA CTT TAT TAG                                                  780
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : CYNC.SEQ       SEQUENCE : NORMAL   1176 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -   1176

    TRANSLATION REGION :      1 -   1176


*** DNA TRANSLATION ***

   1  Met Gln Thr Arg Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Lys      16
   1  ATG CAA ACA AGA ATT GTA AAT AGC TGG AAT GAG TGG GAT GAA CTA AAG      48

  17  Glu Met Val Val Gly Ile Ala Asp Gly Ala Tyr Phe Glu Pro Thr Glu      32
  49  GAG ATG GTT GTC GGG ATT GCA GAT GGT GCT TAT TTT GAA CCA ACT GAG      96

  33  Pro Gly Asn Arg Pro Ala Leu Arg Asp Lys Asn Ile Ala Lys Met Phe      48
  97  CCA GGT AAC CGC CCT GCT TTA CGC GAT AAG AAC ATT GCC AAA ATG TTC     144

  49  Ser Phe Pro Arg Gly Pro Lys Lys Gln Glu Val Thr Glu Lys Ala Asn      64
 145  TCT TTT CCC AGG GGT CCG AAA AAG CAA GAG GTA ACA GAG AAA GCT AAT     192

  65  Glu Glu Leu Asn Gly Leu Val Ala Leu Leu Glu Ser Gln Gly Val Thr      80
 193  GAG GAG TTG AAT GGG CTG GTA GCG CTT CTA GAA TCA CAG GGC GTA ACT     240

  81  Val Arg Arg Pro Glu Lys His Asn Phe Gly Leu Ser Val Lys Thr Pro      96
 241  GTA CGC CGC CCA GAG AAA CAT AAC TTT GGC CTG TCT GTG AAG ACA CCA     288

  97  Phe Phe Glu Val Glu Asn Gln Tyr Cys Ala Val Cys Pro Arg Asp Val     112
 289  TTC TTT GAG GTA GAG AAT CAA TAT TGT GCG GTC TGC CCA CGT GAT GTT     336

 113  Met Ile Thr Phe Gly Asn Glu Ile Leu Glu Ala Thr Met Ser Arg Arg     128
 337  ATG ATC ACC TTT GGG AAC GAA ATT CTC GAA GCA ACT ATG TCA CGG CGG     384

 129  Ser Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr     144
 385  TCA CGC TTC TTT GAG TAT TTA CCC TAT CGC AAA CTA GTC TAT GAA TAT     432

 145  Trp His Lys Asp Pro Asp Met Ile Trp Asn Ala Ala Pro Lys Pro Thr     160
 433  TGG CAT AAA GAT CCA GAT ATG ATC TGG AAT GCT GCG CCT AAA CCG ACT     480

 161  Met Gln Asn Ala Met Tyr Arg Glu Asp Phe Trp Glu Cys Pro Met Glu     176
 481  ATG CAA AAT GCC ATG TAC CGC GAA GAT TTC TGG GAG TGT CCG ATG GAA     528

 177  Asp Arg Phe Glu Ser Met His Asp Phe Glu Phe Cys Val Thr Gln Asp     192
 529  GAT CGA TTT GAG AGT ATG CAT GAT TTT GAG TTC TGC GTC ACC CAG GAT     576

 193  Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Phe Gly Arg Asp Ile     208
 577  GAG GTG ATT TTT GAC GCA GCA GAC TGT AGC CGC TTT GGC CGT GAT ATT     624

 209  Phe Val Gln Glu Ser Met Thr Thr Asn Arg Ala Gly Ile Arg Trp Leu     224
 625  TTT GTG CAG GAG TCA ATG ACG ACT AAT CGT GCA GGG ATT CGC TGG CTC     672

 225  Lys Arg His Leu Glu Pro Arg Arg Phe Arg Val His Asp Ile His Phe     240
 673  AAA CGG CAT TTA GAG CCG CGT CGC TTC CGC GTG CAT GAT ATT CAC TTC     720

 241  Pro Leu Asp Ile Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu     256
 721  CCA CTA GAT ATT TTC CCA TCC CAC ATT GAT TGT ACT TTT GTC CCC TTA     768

 257  Ala Pro Gly Val Val Leu Val Asn Pro Asp Arg Pro Ile Lys Glu Gly     272
```

```
 769  GCA CCT GGG GTT GTG TTA GTG AAT CCA GAT CGC CCC ATC AAA GAG GGT     816

 273  Glu Glu Lys Leu Phe Met Asp Asn Gly Trp Gln Phe Ile Glu Ala Pro     288
 817  GAA GAG AAA CTC TTC ATG GAT AAC GGT TGG CAA TTC ATC GAA GCA CCC     864

 289  Leu Pro Thr Ser Thr Asp Asp Glu Met Pro Met Phe Cys Gln Ser Ser     304
 865  CTC CCC ACT TCC ACC GAC GAT GAG ATG CCT ATG TTC TGC CAG TCC AGT     912

 305  Lys Trp Leu Ala Met Asn Val Leu Ser Ile Ser Pro Lys Lys Val Ile     320
 913  AAG TGG TTG GCG ATG AAT GTG TTA AGC ATT TCC CCC AAG AAG GTC ATC     960

 321  Cys Glu Glu Gln Glu His Pro Leu His Glu Leu Leu Asp Lys His Gly     336
 961  TGT GAA GAG CAA GAG CAT CCG CTT CAT GAG TTG CTA GAT AAA CAC GGC    1008

 337  Phe Glu Val Tyr Pro Ile Pro Phe Arg Asn Val Phe Glu Phe Gly Gly     352
1009  TTT GAG GTC TAT CCA ATT CCC TTT CGC AAT GTC TTT GAG TTT GGC GGT    1056

 353  Ser Leu His Cys Ala Thr Trp Asp Ile His Arg Thr Gly Thr Cys Glu     368
1057  TCG CTC CAT TGT GCC ACC TGG GAT ATC CAT CGC ACG GGA ACC TGT GAG    1104

 369  Asp Tyr Phe Pro Lys Leu Asn Tyr Thr Pro Val Thr Ala Ser Thr Asn     384
1105  GAT TAC TTC CCT AAA CTA AAC TAT ACG CCG GTA ACT GCA TCA ACC AAT    1152

 385  Gly Val Ser Arg Phe Ile Ile ***                                     392
1153  GGC GTT TCT CGC TTC ATC ATT TAG                                    1176
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : CYND.SEQ       SEQUENCE : NORMAL   8754 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -   8754

    TRANSLATION REGION :      1 -   8754


*** DNA TRANSLATION ***

   1  Met Gln Lys Arg Glu Ser Pro Gln Ile Leu Phe Asp Gly Asn Gly Thr      16
   1  ATG CAA AAG AGA GAA AGC CCA CAG ATA CTA TTT GAT GGG AAT GGA ACA      48

  17  Gln Ser Glu Phe Pro Asp Ser Cys Ile His His Leu Phe Glu Asp Gln      32
  49  CAA TCT GAG TTT CCA GAT AGT TGC ATT CAC CAC TTG TTC GAG GAT CAA      96

  33  Ala Ala Lys Arg Pro Asp Ala Ile Ala Leu Ile Asp Gly Glu Gln Ser      48
  97  GCC GCA AAG CGA CCG GAT GCG ATC GCT CTC ATT GAC GGT GAG CAA TCC     144

  49  Leu Thr Tyr Gly Glu Leu Asn Val Arg Ala Asn His Leu Ala Gln His      64
 145  CTT ACC TAC GGG GAA CTA AAT GTA CGC GCT AAC CAC CTA GCC CAG CAT     192

  65  Leu Leu Ser Leu Gly Cys Gln Pro Asp Asp Leu Leu Ala Ile Cys Ile      80
 193  CTC TTG TCC CTA GGC TGT CAA CCC GAT GAC CTC CTC GCC ATC TGC ATC     240

  81  Glu Arg Ser Ala Glu Leu Phe Ile Gly Leu Leu Gly Ile Leu Lys Ala      96
 241  GAG CGT TCG GCA GAA CTC TTT ATT GGT TTG TTG GGT ATC CTA AAA GCC     288

  97  Gly Cys Ala Tyr Val Pro Leu Asp Val Gly Tyr Pro Gly Asp Arg Ile     112
 289  GGA TGT GCT TAT GTG CCT TTG GAT GTA GGC TAT CCT GGC GAT CGC ATA     336

 113  Glu Tyr Met Leu Arg Asp Ser Asp Ala Arg Ile Leu Leu Thr Ser Thr     128
 337  GAG TAT ATG TTG CGG GAC TCG GAT GCG CGT ATT TTA CTA ACC TCA ACG     384

 129  Asp Val Ala Lys Lys Leu Ala Leu Thr Ile Pro Ala Leu Gln Glu Cys     144
 385  GAT GTC GCT AAG AAA CTT GCC TTA ACC ATA CCT GCA TTG CAA GAG TGC     432

 145  Gln Thr Val Tyr Leu Asp Gln Glu Ile Phe Glu Tyr Asp Phe His Phe     160
 433  CAA ACC GTC TAT TTA GAT CAA GAG ATA TTT GAG TAT GAT TTT CAT TTT     480

 161  Leu Ala Ile Ala Lys Leu Leu His Asn Gln Tyr Leu Arg Leu Leu His     176
 481  TTA GCG ATA GCT AAA CTA TTA CAT AAC CAA TAC TTG AGA TTA TTA CAT     528

 177  Phe Tyr Phe Tyr Thr Leu Ile Gln Gln Cys Gln Ala Thr Ser Val Ser     192
 529  TTT TAT TTT TAT ACC TTG ATT CAG CAA TGC CAG GCA ACT TCG GTT TCC     576

 193  Gln Gly Ile Gln Thr Gln Val Leu Pro Asn Asn Leu Ala Tyr Cys Ile     208
 577  CAA GGG ATT CAG ACA CAG GTT CTC CCC AAT AAT CTC GCT TAC TGC ATT     624

 209  Tyr Thr Ser Gly Ser Thr Gly Asn Pro Lys Gly Ile Leu Met Glu His     224
 625  TAC ACC TCT GGC TCT ACC GGA AAT CCC AAA GGG ATC TTG ATG GAA CAT     672

 225  Arg Ser Leu Val Asn Met Leu Trp Trp His Gln Gln Thr Arg Pro Ser     240
 673  CGC TCA CTG GTG AAT ATG CTT TGG TGG CAT CAG CAA ACG CGG CCT TCG     720

 241  Val Gln Gly Val Arg Thr Leu Gln Phe Cys Ala Val Ser Phe Asp Phe     256
 721  GTT CAG GGT GTT AGG ACG CTG CAA TTT TGT GCA GTC AGC TTT GAC TTT     768

 257  Ser Cys His Glu Ile Phe Ser Thr Leu Cys Leu Gly Gly Ile Leu Val     272
 769  TCC TGC CAT GAA ATT TTT TCT ACC CTC TGT CTT GGC GGG ATA TTG GTC     816
```

```
 273  Leu Val Pro Glu Ala Val Arg Gln Asn Pro Phe Ala Leu Ala Glu Phe     288
 817  TTG GTG CCA GAG GCA GTG CGC CAA AAT CCC TTT GCA TTG GCT GAG TTC     864

 289  Ile Ser Gln Gln Lys Ile Glu Lys Leu Phe Leu Pro Val Ile Ala Leu     304
 865  ATC AGT CAA CAG AAA ATT GAA AAA TTG TTT CTT CCC GTT ATA GCA TTA     912

 305  Leu Gln Leu Ala Glu Ala Val Asn Gly Asn Lys Ser Thr Ser Leu Ala     320
 913  CTA CAG TTG GCC GAA GCT GTA AAT GGG AAT AAA AGC ACC TCC CTC GCG     960

 321  Leu Cys Glu Val Ile Thr Thr Gly Glu Gln Met Gln Ile Thr Pro Ala     336
 961  CTT TGC GAA GTT ATC ACT ACC GGG GAG CAG ATG CAG ATC ACA CCT GCT    1008

 337  Val Ala Asn Leu Phe Gln Lys Thr Gly Ala Met Leu His Asn His Tyr     352
1009  GTC GCC AAC CTC TTT CAG AAA ACC GGG GCG ATG TTG CAT AAT CAC TAC    1056

 353  Gly Ala Thr Glu Phe Gln Asp Ala Thr Thr His Thr Leu Lys Gly Asn     368
1057  GGG GCA ACA GAA TTT CAA GAT GCC ACC ACT CAT ACC CTC AAG GGC AAT    1104

 369  Pro Glu Gly Trp Pro Thr Leu Val Pro Val Gly Arg Pro Leu His Asn     384
1105  CCA GAG GGC TGG CCA ACA CTG GTG CCA GTG GGT CGT CCA CTG CAC AAT    1152

 385  Val Gln Val Tyr Ile Leu Asp Glu Ala Gln Gln Pro Val Pro Leu Gly     400
1153  GTT CAA GTG TAT ATT CTG GAT GAG GCA CAG CAA CCT GTA CCT CTT GGT    1200

 401  Gly Glu Gly Glu Phe Cys Ile Gly Gly Ile Gly Leu Ala Arg Gly Tyr     416
1201  GGA GAG GGT GAA TTC TGT ATT GGT GGT ATT GGA CTG GCT CGT GGC TAT    1248

 417  His Asn Leu Pro Asp Leu Thr Asn Glu Lys Phe Ile Pro Asn Pro Phe     432
1249  CAC AAT TTG CCT GAC CTA ACG AAT GAA AAA TTT ATT CCC AAT CCA TTT    1296

 433  Gly Ala Asn Glu Asn Ala Lys Lys Leu Tyr Arg Thr Gly Asp Leu Ala     448
1297  GGG GCT AAT GAG AAC GCT AAA AAA CTC TAC CGC ACA GGG GAC TTG GCA    1344

 449  Arg Tyr Leu Pro Asp Gly Thr Ile Glu His Leu Gly Arg Ile Asp His     464
1345  CGC TAC CTA CCC GAC GGC ACG ATT GAG CAT TTA GGA CGG ATA GAC CAC    1392

 465  Gln Val Lys Ile Arg Gly Phe Arg Val Glu Leu Gly Glu Ile Glu Ser     480
1393  CAG GTT AAG ATC CGA GGT TTC CGC GTG GAA TTG GGG GAA ATT GAG TCC    1440

 481  Val Leu Ala Ser His Gln Ala Val Arg Glu Cys Ala Val Val Ala Arg     496
1441  GTG CTG GCA AGT CAC CAA GCT GTG CGT GAA TGT GCC GTT GTG GCA CGG    1488

 497  Glu Ile Ala Gly His Thr Gln Leu Val Gly Tyr Ile Ile Ala Lys Asp     512
1489  GAG ATT GCA GGT CAT ACA CAG TTG GTA GGG TAT ATC ATA GCA AAG GAT    1536

 513  Thr Leu Asn Leu Ser Phe Asp Lys Leu Glu Pro Ile Leu Arg Gln Tyr     528
1537  ACA CTT AAT CTC AGT TTC GAC AAA CTT GAA CCT ATC CTG CGT CAA TAT    1584

 529  Ser Glu Ala Val Leu Pro Glu Tyr Met Ile Pro Thr Arg Phe Ile Asn     544
1585  TCG GAA GCG GTG CTG CCA GAA TAC ATG ATA CCC ACT CGG TTC ATC AAT    1632

 545  Ile Ser Asn Met Pro Leu Thr Pro Ser Gly Lys Leu Asp Arg Arg Ala     560
1633  ATC AGT AAT ATG CCG TTG ACT CCC AGT GGT AAA CTT GAC CGC AGG GCA    1680

 561  Leu Pro Asp Pro Lys Gly Asp Arg Pro Ala Leu Ser Thr Pro Leu Val     576
1681  TTA CCT GAT CCC AAA GGC GAT CGC CCT GCA TTG TCT ACC CCA CTT GTC    1728

 577  Lys Pro Arg Thr Gln Thr Glu Lys Arg Leu Ala Glu Ile Trp Gly Ser     592
1729  AAG CCT CGT ACC CAG ACA GAG AAA CGT TTA GCA GAG ATT TGG GGC AGT    1776

 593  Tyr Leu Ala Val Asp Ile Val Gly Thr His Asp Asn Phe Phe Asp Leu     608
1777  TAT CTT GCT GTA GAT ATT GTG GGA ACC CAC GAC AAT TTC TTT GAT CTA    1824

 609  Gly Gly Thr Ser Leu Leu Leu Thr Gln Ala His Lys Phe Leu Cys Glu     624
1825  GGC GGT ACG TCA CTG CTA TTG ACT CAA GCG CAC AAA TTC CTG TGC GAG    1872

 625  Thr Phe Asn Ile Asn Leu Ser Ala Val Ser Leu Phe Gln Tyr Pro Thr     640
1873  ACC TTT AAT ATT AAT TTG TCC GCT GTC TCA CTC TTT CAA TAT CCC ACA    1920
```

```
 641  Ile Gln Thr Leu Ala Gln Tyr Ile Asp Cys Gln Gly Asp Thr Thr Ser     656
1921  ATT CAG ACA TTG GCA CAA TAT ATT GAT TGC CAA GGA GAC ACA ACC TCA    1968

 657  Ser Asp Thr Ala Ser Arg His Lys Lys Val Arg Lys Lys Gln Ser Gly     672
1969  AGC GAT ACA GCA TCC AGG CAC AAG AAA GTA CGT AAA AAG CAG TCC GGT    2016

 673  Asp Ser Asn Asp Ile Ala Ile Ile Ser Val Ala Gly Arg Phe Pro Gly     688
2017  GAC AGC AAC GAT ATT GCC ATC ATC AGT GTG GCA GGT CGC TTT CCG GGT    2064

 689  Ala Glu Thr Ile Glu Gln Phe Trp His Asn Leu Cys Asn Gly Val Glu     704
2065  GCT GAA ACG ATT GAG CAG TTC TGG CAT AAT CTC TGT AAT GGT GTT GAA    2112

 705  Ser Ile Thr Leu Phe Ser Asp Asp Glu Leu Glu Gln Thr Leu Pro Glu     720
2113  TCC ATC ACC CTT TTT AGT GAT GAT GAG CTA GAG CAG ACT TTG CCT GAG    2160

 721  Leu Phe Asn Asn Pro Ala Tyr Val Lys Ala Gly Ala Val Leu Glu Gly     736
2161  TTA TTT AAT AAT CCC GCT TAT GTC AAA GCA GGT GCG GTG CTA GAA GGC    2208

 737  Val Glu Leu Phe Asp Ala Thr Phe Phe Gly Tyr Ser Pro Lys Glu Ala     752
2209  GTT GAA TTA TTT GAT GCT ACC TTT TTT GGC TAC AGC CCC AAA GAA GCT    2256

 753  Ala Val Thr Asp Pro Gln Gln Arg Ile Leu Leu Glu Cys Ala Trp Glu     768
2257  GCG GTG ACA GAC CCT CAG CAA CGG ATT TTG CTA GAG TGT GCC TGG GAA    2304

 769  Ala Phe Glu Arg Ala Gly Tyr Asn Pro Glu Thr Tyr Pro Glu Pro Val     784
2305  GCA TTT GAA CGG GCT GGC TAC AAC CCC GAA ACC TAT CCA GAA CCA GTT    2352

 785  Gly Val Tyr Ala Gly Ser Ser Leu Ser Thr Tyr Leu Leu Asn Asn Ile     800
2353  GGT GTT TAT GCT GGT TCA AGC CTG AGT ACC TAT CTG CTT AAC AAT ATT    2400

 801  Gly Ser Ala Leu Gly Ile Ile Thr Glu Gln Pro Phe Ile Glu Thr Asp     816
2401  GGC TCT GCT TTA GGC ATA ATT ACC GAG CAA CCC TTT ATT GAA ACG GAT    2448

 817  Met Glu Gln Phe Gln Ala Lys Ile Gly Asn Asp Arg Ser Tyr Leu Ala     832
2449  ATG GAG CAG TTT CAG GCT AAA ATT GGC AAT GAC CGG AGC TAT CTT GCT    2496

 833  Thr Arg Ile Ser Tyr Lys Leu Asn Leu Lys Gly Pro Ser Val Asn Val     848
2497  ACA CGC ATC TCT TAC AAG CTG AAT CTC AAG GGT CCA AGC GTC AAT GTG    2544

 849  Gln Thr Ala Cys Ser Thr Ser Leu Val Ala Val His Met Ala Cys Gln     864
2545  CAG ACC GCC TGC TCA ACC TCG TTA GTT GCG GTT CAC ATG GCC TGT CAG    2592

 865  Ser Leu Ile Ser Gly Glu Cys Gln Met Ala Leu Ala Gly Gly Ile Ser     880
2593  AGT CTC ATT AGT GGA GAG TGT CAA ATG GCT TTA GCC GGT GGT ATT TCT    2640

 881  Val Val Val Pro Gln Lys Gly Gly Tyr Leu Tyr Glu Glu Gly Met Val     896
2641  GTG GTT GTA CCA CAG AAG GGG GGC TAT CTC TAC GAA GAA GGC ATG GTT    2688

 897  Arg Ser Gln Asp Gly His Cys Arg Ala Phe Asp Ala Glu Ala Gln Gly     912
2689  CGT TCC CAG GAT GGT CAT TGT CGC GCC TTT GAT GCC GAA GCC CAA GGG    2736

 913  Thr Ile Phe Gly Asn Gly Gly Gly Leu Val Leu Leu Lys Arg Leu Gln     928
2737  ACT ATA TTT GGC AAT GGC GGC GGC TTG GTT TTG CTT AAA CGG TTG CAG    2784

 929  Asp Ala Leu Asp Asp Asn Asp Asn Ile Met Ala Val Ile Lys Ala Thr     944
2785  GAT GCA CTG GAC GAT AAC GAC AAC ATT ATG GCA GTC ATC AAA GCC ACA    2832

 945  Ala Ile Asn Asn Asp Gly Ala Leu Lys Met Gly Tyr Thr Ala Pro Ser     960
2833  GCC ATC AAC AAC GAC GGT GCG CTC AAG ATG GGC TAC ACA GCA CCG AGC    2880

 961  Val Asp Gly Gln Ala Asp Val Ile Ser Glu Ala Ile Ala Ile Ala Asp     976
2881  GTG GAT GGG CAA GCT GAT GTA ATT AGC GAG GCG ATT GCT ATC GCT GAC    2928

 977  Ile Asp Ala Ser Thr Ile Gly Tyr Val Glu Ala His Gly Thr Ala Thr     992
2929  ATA GAT GCA AGC ACC ATT GGC TAT GTA GAA GCT CAT GGC ACA GCC ACC    2976

 993  Gln Leu Gly Asp Pro Ile Glu Val Ala Gly Leu Ala Arg Ala Phe Gln    1008
```

```
2977  CAA TTG GGT GAT CCG ATT GAA GTA GCA GGG TTA GCA AGG GCA TTT CAG    3024

1009  Arg Ser Thr Asp Ser Val Leu Gly Lys Gln Gln Cys Ala Ile Gly Ser    1024
3025  CGT AGT ACG GAC AGC GTC CTT GGT AAA CAA CAA TGC GCT ATT GGA TCA    3072

1025  Val Lys Thr Asn Ile Gly His Leu Asp Glu Ala Ala Gly Ile Ala Gly    1040
3073  GTT AAA ACT AAT ATT GGC CAC TTA GAT GAG GCG GCA GGC ATT GCC GGA    3120

1041  Leu Ile Lys Ala Ala Leu Ala Leu Gln Tyr Gly Gln Ile Pro Pro Ser    1056
3121  CTG ATA AAG GCT GCT CTA GCT CTA CAA TAT GGA CAG ATT CCA CCG AGC    3168

1057  Leu His Tyr Ala Asn Pro Asn Pro Arg Ile Asp Phe Asp Ala Thr Pro    1072
3169  TTG CAC TAT GCC AAT CCT AAT CCA CGG ATT GAT TTT GAC GCA ACC CCA    3216

1073  Phe Phe Val Asn Thr Glu Leu Arg Glu Trp Ser Arg Asn Gly Tyr Pro    1088
3217  TTT TTT GTC AAC ACA GAA CTA CGC GAA TGG TCA AGG AAT GGT TAT CCT    3264

1089  Arg Arg Ala Gly Val Ser Ser Phe Gly Val Gly Gly Thr Asn Ser His    1104
3265  CGG CGG GCG GGG GTG AGT TCT TTT GGT GTG GGT GGA ACT AAC AGC CAT    3312

1105  Ile Val Leu Glu Glu Ser Pro Val Lys Gln Pro Thr Leu Phe Ser Ser    1120
3313  ATT GTG CTG GAG GAG TCG CCT GTA AAG CAA CCC ACA TTG TTC TCT TCT    3360

1121  Leu Pro Glu Arg Ser His His Leu Leu Thr Leu Ser Ala His Thr Gln    1136
3361  TTG CCA GAA CGC AGT CAT CAT CTG CTG ACG CTT TCT GCC CAT ACA CAA    3408

1137  Glu Ala Leu His Glu Leu Val Gln Arg Tyr Ile Gln His Asn Glu Thr    1152
3409  GAG GCT TTG CAT GAG TTG GTG CAA CGC TAC ATC CAA CAT AAC GAG ACA    3456

1153  His Leu Asp Ile Asn Leu Gly Asp Leu Cys Phe Thr Ala Asn Thr Gly    1168
3457  CAC CTT GAT ATT AAC TTA GGC GAC CTC TGT TTC ACA GCC AAT ACG GGA    3504

1169  Arg Lys His Phe Glu His Arg Leu Ala Val Val Ala Glu Ser Ile Pro    1184
3505  CGC AAG CAT TTT GAG CAT CGC CTA GCG GTT GTA GCC GAA TCA ATC CCT    3552

1185  Gly Leu Gln Ala Gln Leu Glu Thr Ala Gln Thr Ala Ile Ser Ala Gln    1200
3553  GGC TTA CAG GCA CAA CTG GAA ACT GCA CAG ACT GCG ATT TCA GCA CAG    3600

1201  Lys Lys Asn Ala Pro Pro Thr Ile Ala Phe Leu Phe Thr Gly Gln Gly    1216
3601  AAA AAA AAT GCC CCG CCG ACG ATC GCA TTC CTG TTT ACA GGT CAA GGC    3648

1217  Ser Gln Tyr Ile Asn Met Gly Arg Thr Leu Tyr Asp Thr Glu Ser Thr    1232
3649  TCA CAA TAC ATT AAC ATG GGG CGC ACC CTC TAC GAT ACT GAA TCA ACA    3696

1233  Phe Arg Ala Ala Leu Asp Arg Cys Glu Thr Ile Leu Gln Asn Leu Gly    1248
3697  TTC CGT GCA GCC CTT GAC CGA TGT GAA ACC ATT CTC CAA AAT TTA GGG    3744

1249  Ile Glu Ser Ile Leu Ser Val Ile Phe Gly Ser Ser Glu His Gly Leu    1264
3745  ATC GAG TCC ATT CTC TCC GTT ATT TTT GGT TCA TCT GAG CAT GGA CTC    3792

1265  Ser Leu Asp Asp Thr Ala Tyr Thr Gln Pro Ala Leu Phe Ala Ile Glu    1280
3793  TCA TTA GAT GAC ACA GCC TAT ACC CAG CCC GCA CTC TTT GCC ATC GAA    3840

1281  Tyr Ala Leu Tyr Gln Leu Trp Lys Ser Trp Gly Ile Gln Pro Ser Val    1296
3841  TAC GCG CTC TAT CAA TTA TGG AAG TCG TGG GGC ATC CAG CCC TCA GTG    3888

1297  Val Ile Gly His Ser Val Gly Glu Tyr Val Ser Ala Cys Val Ala Gly    1312
3889  GTG ATA GGT CAT AGT GTA GGT GAA TAT GTG TCC GCT TGT GTG GCG GGA    3936

1313  Val Phe Ser Leu Glu Asp Gly Leu Lys Leu Ile Ala Glu Arg Gly Arg    1328
3937  GTC TTT AGC TTA GAG GAT GGG TTG AAA CTG ATT GCA GAA CGA GGA CGA    3984

1329  Leu Ile Gln Ala Leu Pro Arg Asp Gly Ser Met Val Ser Val Met Ala    1344
3985  CTG ATA CAG GCA CTT CCT CGT GAT GGG AGC ATG GTT TCC GTG ATG GCA    4032

1345  Ser Glu Lys Arg Ile Ala Asp Ile Ile Leu Pro Tyr Gly Gly Gln Val    1360
4033  AGC GAG AAG CGT ATT GCA GAT ATC ATT TTA CCT TAT GGG GGA CAG GTA    4080
```

```
1361  Gly Ile Ala Ala Ile Asn Gly Pro Gln Ser Val Val Ile Ser Gly Gln    1376
4081  GGG ATC GCC GCG ATT AAT GGC CCA CAA AGT GTT GTA ATT TCT GGG CAA    4128

1377  Gln Gln Ala Ile Asp Ala Ile Cys Ala Ile Leu Glu Thr Glu Gly Ile    1392
4129  CAG CAA GCG ATT GAT GCT ATT TGT GCC ATC TTG GAA ACT GAG GGC ATC    4176

1393  Lys Ser Lys Lys Leu Asn Val Ser His Ala Phe His Ser Pro Leu Val    1408
4177  AAA AGC AAG AAG CTA AAC GTC TCC CAT GCC TTC CAC TCG CCG CTA GTG    4224

1409  Glu Ala Met Leu Asp Ser Phe Leu Gln Val Ala Gln Glu Val Thr Tyr    1424
4225  GAA GCA ATG TTA GAC TCT TTC TTG CAG GTT GCA CAA GAG GTC ACT TAC    4272

1425  Ser Gln Pro Gln Ile Lys Leu Ile Ser Asn Val Thr Gly Thr Leu Ala    1440
4273  TCG CAA CCT CAA ATC AAG CTT ATC TCT AAT GTA ACG GGA ACA TTG GCA    4320

1441  Ser His Glu Ser Cys Pro Asp Glu Leu Pro Ile Thr Thr Ala Glu Tyr    1456
4321  AGC CAT GAA TCT TGT CCC GAT GAA CTT CCG ATC ACC ACC GCA GAG TAT    4368

1457  Trp Val Arg His Val Arg Gln Pro Val Arg Phe Ala Ala Gly Met Glu    1472
4369  TGG GTA CGT CAT GTG CGA CAG CCC GTC CGG TTT GCG GCG GGA ATG GAG    4416

1473  Ser Leu Glu Gly Gln Gly Val Asn Val Phe Ile Glu Ile Gly Pro Lys    1488
4417  AGC CTT GAG GGT CAA GGG GTA AAC GTA TTT ATA GAA ATC GGT CCT AAA    4464

1489  Pro Val Leu Leu Gly Met Gly Arg Asp Cys Leu Pro Glu Gln Glu Gly    1504
4465  CCT GTT CTT TTA GGC ATG GGA CGC GAC TGC TTG CCT GAA CAA GAG GGA    4512

1505  Leu Trp Leu Pro Ser Leu Arg Pro Lys Gln Asp Asp Trp Gln Gln Val    1520
4513  CTT TGG TTG CCT AGT TTG CGC CCA AAA CAG GAT GAT TGG CAA CAG GTG    4560

1521  Leu Ser Ser Leu Arg Asp Leu Tyr Leu Ala Gly Val Thr Val Asp Trp    1536
4561  TTA AGT AGT TTG CGT GAT CTA TAC TTA GCA GGT GTA ACC GTA GAT TGG    4608

1537  Ser Ser Phe Asp Gln Gly Tyr Ala Arg Arg Arg Val Pro Leu Pro Thr    1552
4609  AGC AGT TTC GAT CAG GGG TAT GCT CGT CGC CGT GTG CCA CTA CCG ACT    4656

1553  Tyr Pro Trp Gln Arg Glu Arg His Trp Val Glu Pro Ile Ile Arg Gln    1568
4657  TAT CCT TGG CAG CGA GAG CGG CAT TGG GTA GAG CCA ATT ATT CGT CAA    4704

1569  Arg Gln Ser Val Leu Gln Ala Thr Asn Thr Thr Lys Leu Thr Arg Asn    1584
4705  CGG CAA TCA GTA TTA CAA GCC ACA AAT ACC ACC AAG CTA ACT CGT AAC    4752

1585  Ala Ser Val Ala Gln His Pro Leu Leu Gly Gln Arg Leu His Leu Ser    1600
4753  GCC AGC GTG GCG CAG CAT CCT CTG CTT GGT CAA CGG CTG CAT TTG TCG    4800

1601  Arg Thr Gln Glu Ile Tyr Phe Gln Thr Phe Ile His Ser Asp Phe Pro    1616
4801  CGG ACT CAA GAG ATT TAC TTT CAA ACC TTC ATC CAC TCC GAC TTC CCA    4848

1617  Ile Trp Val Ala Asp His Lys Val Phe Gly Asn Val Ile Ile Pro Gly    1632
4849  ATA TGG GTT GCT GAT CAT AAA GTA TTT GGA AAT GTC ATC ATT CCG GGT    4896

1633  Val Ala Tyr Phe Glu Met Ala Leu Ala Ala Gly Lys Ala Leu Lys Pro    1648
4897  GTC GCC TAT TTT GAG ATG GCA CTG GCA GCA GGG AAG GCA CTT AAA CCA    4944

1649  Asp Ser Ile Phe Trp Leu Glu Asp Val Ser Ile Ala Gln Ala Leu Ile    1664
4945  GAC AGT ATA TTT TGG CTC GAA GAT GTA TCC ATC GCC CAA GCA CTG ATT    4992

1665  Ile Pro Asp Glu Gly Gln Thr Val Gln Ile Val Leu Ser Pro Gln Glu    1680
4993  ATT CCC GAT GAA GGG CAA ACT GTG CAA ATA GTA TTA AGC CCA CAG GAA    5040

1681  Glu Ser Ala Tyr Phe Phe Glu Ile Leu Ser Leu Glu Lys Glu Asn Ser    1696
5041  GAG TCA GCT TAT TTT TTT GAA ATC CTC TCT TTA GAA AAA GAA AAC TCT    5088

1697  Trp Val Leu His Ala Ser Gly Lys Leu Val Ala Gln Glu Gln Val Leu    1712
5089  TGG GTG CTT CAT GCC TCT GGT AAG CTA GTC GCC CAA GAG CAA GTG CTA    5136

1713  Glu Thr Glu Pro Ile Asp Leu Ile Ala Leu Gln Ala His Cys Ser Glu    1728
5137  GAA ACC GAG CCA ATT GAC TTG ATT GCG TTA CAG GCA CAT TGT TCC GAA    5184
```

```
1729  Glu Val Ser Val Asp Val Leu Tyr Gln Glu Glu Met Ala Arg Arg Leu    1744
5185  GAA GTG TCA GTA GAT GTG CTA TAT CAG GAA GAA ATG GCG CGC CGG CTG    5232

1745  Asp Met Gly Pro Met Met Arg Gly Val Lys Gln Leu Trp Arg Tyr Pro    1760
5233  GAT ATG GGT CCA ATG ATG CGT GGG GTG AAG CAG CTT TGG CGT TAT CCG    5280

1761  Leu Ser Phe Ala Lys Ser His Asp Ala Ile Ala Leu Ala Lys Val Ser    1776
5281  CTC TCC TTT GCC AAA AGT CAT GAT GCG ATC GCA CTC GCC AAG GTC AGC    5328

1777  Leu Pro Glu Ile Leu Leu His Glu Ser Asn Ala Tyr Gln Phe His Pro    1792
5329  TTG CCA GAA ATC TTG CTT CAT GAG TCC AAT GCC TAC CAA TTC CAT CCT    5376

1793  Val Ile Leu Asp Ala Gly Leu Gln Met Ile Thr Val Ser Tyr Pro Glu    1808
5377  GTA ATC TTG GAT GCG GGG CTG CAA ATG ATA ACG GTC TCT TAT CCT GAA    5424

1809  Ala Asn Gln Gly Gln Thr Tyr Val Pro Val Gly Ile Glu Gly Leu Gln    1824
5425  GCA AAC CAA GGC CAG ACT TAT GTA CCT GTT GGT ATA GAG GGT CTA CAA    5472

1825  Val Tyr Gly Arg Pro Ser Ser Glu Leu Trp Cys Arg Ala Gln Tyr Arg    1840
5473  GTC TAT GGT CGT CCC AGT TCA GAA CTT TGG TGT CGC GCC CAA TAT CGG    5520

1841  Pro Pro Leu Asp Thr Asp Gln Arg Gln Gly Ile Asp Leu Leu Pro Lys    1856
5521  CCT CCT TTG GAT ACA GAT CAA AGG CAG GGT ATT GAT TTG CTG CCA AAG    5568

1857  Lys Leu Ile Ala Asp Leu His Leu Phe Asp Thr Gln Gly Arg Val Val    1872
5569  AAA TTG ATT GCA GAC TTG CAT CTA TTT GAT ACC CAG GGT CGT GTG GTT    5616

1873  Ala Ile Met Phe Gly Val Gln Ser Val Leu Val Gly Arg Glu Ala Met    1888
5617  GCC ATC ATG TTT GGT GTG CAA TCT GTC CTT GTG GGA CGG GAA GCA ATG    5664

1889  Leu Arg Ser Gln Asp Thr Trp Arg Asn Trp Leu Tyr Gln Val Leu Trp    1904
5665  TTG CGA TCG CAA GAT ACT TGG CGA AAT TGG CTT TAT CAA GTC CTG TGG    5712

1905  Lys Pro Gln Ala Cys Phe Gly Leu Leu Pro Asn Tyr Leu Pro Thr Pro    1920
5713  AAA CCT CAA GCC TGT TTT GGA CTT TTA CCG AAT TAC CTG CCA ACC CCA    5760

1921  Asp Lys Ile Arg Lys Arg Leu Glu Thr Lys Leu Ala Thr Leu Ile Ile    1936
5761  GAT AAG ATT CGG AAA CGC CTG GAA ACA AAG TTA GCG ACA TTG ATC ATC    5808

1937  Glu Ala Asn Leu Ala Thr Tyr Ala Ile Ala Tyr Thr Gln Leu Glu Arg    1952
5809  GAA GCT AAT TTG GCG ACT TAT GCG ATC GCC TAT ACC CAA CTG GAA AGG    5856

1953  Leu Ser Leu Ala Tyr Val Val Ala Ala Phe Arg Gln Met Gly Trp Leu    1968
5857  TTA AGT CTA GCT TAC GTT GTG GCG GCT TTC CGA CAA ATG GGC TGG CTG    5904

1969  Phe Gln Pro Gly Glu Arg Phe Ser Thr Ala Gln Lys Val Ser Ala Leu    1984
5905  TTT CAA CCC GGT GAG CGT TTT TCC ACC GCC CAG AAG GTA TCA GCG TTA    5952

1985  Gly Ile Val Asp Gln His Arg Gln Leu Phe Ala Arg Leu Leu Asp Ile    2000
5953  GGA ATC GTT GAT CAA CAT CGG CAA CTA TTC GCT CGT TTG CTC GAC ATT    6000

2001  Leu Ala Glu Ala Asp Ile Leu Arg Ser Glu Asn Leu Met Thr Ile Trp    2016
6001  CTA GCC GAA GCA GAC ATA CTC CGC AGC GAA AAC TTG ATG ACG ATA TGG    6048

2017  Glu Val Ile Ser Tyr Pro Glu Thr Ile Asp Ile Gln Val Leu Leu Asp    2032
6049  GAA GTC ATT TCA TAC CCG GAA ACG ATT GAT ATA CAG GTA CTT CTT GAC    6096

2033  Asp Leu Glu Ala Lys Glu Ala Glu Ala Glu Val Thr Leu Val Ser Arg    2048
6097  GAC CTC GAA GCC AAA GAA GCA GAA GCC GAA GTC ACA CTG GTT TCC CGT    6144

2049  Cys Ser Ala Lys Leu Ala Glu Val Leu Gln Gly Lys Cys Asp Pro Ile    2064
6145  TGC AGT GCA AAA TTG GCC GAA GTA TTA CAA GGA AAA TGT GAC CCC ATA    6192

2065  Gln Leu Leu Phe Pro Ala Gly Asp Thr Thr Thr Leu Ser Lys Leu Tyr    2080
6193  CAG TTG CTC TTT CCC GCA GGG GAC ACA ACA ACG TTA AGC AAA CTC TAT    6240

2081  Arg Glu Ala Pro Val Leu Gly Val Thr Asn Thr Leu Val Gln Glu Ala    2096
```

```
6241  CGT GAA GCC CCA GTT TTG GGT GTT ACT AAT ACT CTA GTC CAA GAA GCG    6288

2097  Leu Leu Ser Ala Leu Glu Gln Leu Pro Pro Glu Arg Gly Trp Arg Ile    2112
6289  CTT CTT TCC GCC CTG GAG CAG TTG CCG CCG GAA CGT GGT TGG CGA ATT    6336

2113  Leu Glu Ile Gly Ala Gly Thr Gly Gly Thr Thr Ala Tyr Leu Leu Pro    2128
6337  TTA GAG ATT GGT GCT GGA ACA GGT GGA ACC ACA GCC TAC TTG TTA CCG    6384

2129  His Leu Pro Gly Asp Gln Thr Lys Tyr Val Phe Thr Asp Ile Ser Ala    2144
6385  CAT CTG CCT GGG GAT CAG ACA AAA TAT GTC TTT ACC GAT ATT AGT GCC    6432

2145  Phe Phe Leu Ala Lys Ala Glu Glu Arg Phe Lys Asp Tyr Pro Phe Val    2160
6433  TTT TTT CTT GCC AAA GCG GAA GAG CGT TTT AAA GAT TAC CCG TTT GTA    6480

2161  Arg Tyr Gln Val Leu Asp Ile Glu Gln Ala Pro Gln Ala Gln Gly Phe    2176
6481  CGT TAT CAG GTA TTA GAT ATC GAA CAA GCA CCA CAG GCG CAA GGA TTT    6528

2177  Glu Pro Gln Ile Tyr Asp Leu Ile Val Ala Ala Asp Val Leu His Ala    2192
6529  GAA CCC CAA ATA TAC GAT TTA ATC GTA GCA GCG GAT GTC TTG CAT GCT    6576

2193  Thr Ser Asp Leu Arg Gln Thr Leu Val His Ile Arg Gln Leu Leu Ala    2208
6577  ACT AGT GAC CTG CGT CAA ACT CTT GTA CAT ATC CGG CAA TTA TTA GCG    6624

2209  Pro Gly Gly Met Leu Ile Leu Met Glu Asp Ser Glu Pro Ala Arg Trp    2224
6625  CCG GGC GGG ATG TTG ATC CTG ATG GAA GAC AGC GAA CCC GCA CGC TGG    6672

2225  Ala Asp Leu Thr Phe Gly Leu Thr Glu Gly Trp Trp Lys Phe Thr Asp    2240
6673  GCT GAT TTA ACC TTT GGC TTA ACA GAA GGC TGG TGG AAG TTT ACA GAC    6720

2241  His Asp Leu Arg Pro Asn His Pro Leu Leu Ser Pro Glu Gln Trp Gln    2256
6721  CAT GAC TTA CGC CCC AAC CAT CCG CTA TTG TCT CCT GAG CAG TGG CAA    6768

2257  Ile Leu Leu Ser Glu Met Gly Phe Ser Gln Thr Thr Ala Leu Trp Pro    2272
6769  ATC TTG TTG TCA GAA ATG GGA TTT AGT CAA ACA ACC GCC TTA TGG CCA    6816

2273  Lys Ile Asp Ser Pro His Lys Leu Pro Arg Glu Ala Val Ile Val Ala    2288
6817  AAA ATA GAT AGC CCC CAT AAA TTG CCA CGG GAG GCG GTG ATT GTG GCG    6864

2289  Arg Asn Glu Pro Ala Ile Arg Lys Pro Arg Arg Trp Leu Ile Leu Ala    2304
6865  CGT AAT GAA CCA GCC ATC AGA AAA CCC CGA AGA TGG CTG ATC TTG GCT    6912

2305  Asp Glu Glu Ile Gly Gly Leu Leu Ala Lys Gln Leu Arg Glu Glu Gly    2320
6913  GAC GAG GAG ATT GGT GGA CTA CTA GCC AAA CAG CTA CGT GAA GAA GGA    6960

2321  Glu Asp Cys Ile Leu Leu Leu Pro Gly Glu Lys Tyr Thr Glu Arg Asp    2336
6961  GAA GAT TGT ATA CTC CTC TTG CCA GGG GAA AAG TAC ACA GAG AGA GAT    7008

2337  Ser Gln Thr Phe Thr Ile Asn Pro Gly Asp Ile Glu Glu Trp Gln Gln    2352
7009  TCA CAA ACG TTT ACA ATC AAT CCT GGA GAT ATT GAA GAG TGG CAA CAG    7056

2353  Leu Leu Asn Arg Val Pro Asn Ile Gln Glu Ile Val His Cys Trp Ser    2368
7057  TTA TTG AAC CGA GTA CCG AAC ATA CAA GAA ATT GTA CAT TGT TGG AGT    7104

2369  Met Val Ser Thr Asp Leu Asp Arg Ala Thr Ile Phe Ser Cys Ser Ser    2384
7105  ATG GTT TCC ACT GAC TTA GAT AGA GCC ACT ATT TTC AGT TGC AGC AGT    7152

2385  Thr Leu His Leu Val Gln Ala Leu Ala Asn Tyr Pro Lys Asn Pro Arg    2400
7153  ACG CTG CAT TTA GTT CAA GCA TTA GCA AAC TAT CCA AAA AAC CCT CGC    7200

2401  Leu Ser Leu Val Thr Leu Gly Ala Gln Ala Val Asn Glu His His Val    2416
7201  TTG TCA CTT GTC ACC CTA GGC GCA CAA GCC GTT AAC GAA CAT CAT GTT    7248

2417  Gln Asn Val Val Gly Ala Ala Leu Trp Gly Met Gly Lys Val Ile Ala    2432
7249  CAA AAT GTA GTT GGA GCA GCC CTC TGG GGC ATG GGA AAG GTA ATT GCA    7296

2433  Leu Glu His Pro Glu Leu Gln Val Ala Gln Met Asp Leu Asp Pro Asn    2448
7297  CTC GAA CAC CCA GAG CTA CAA GTA GCA CAA ATG GAT TTA GAC CCG AAT    7344
```

```
2449  Gly Lys Val Lys Ala Gln Val Glu Val Leu Arg Asp Glu Leu Leu Ala    2464
7345  GGG AAG GTT AAG GCG CAA GTA GAA GTG CTT AGG GAT GAA CTT CTC GCC    7392

2465  Arg Lys Asp Pro Ala Ser Ala Met Ser Val Pro Asp Leu Gln Thr Arg    2480
7393  AGA AAA GAC CCT GCA TCA GCA ATG TCT GTG CCT GAT CTG CAA ACA CGA    7440

2481  Pro His Glu Lys Gln Ile Ala Phe Arg Glu Gln Thr Arg Tyr Val Ala    2496
7441  CCT CAT GAA AAG CAA ATA GCC TTT CGT GAG CAA ACA CGT TAT GTG GCA    7488

2497  Arg Leu Ser Pro Leu Asp Arg Pro Asn Pro Gly Glu Lys Gly Thr Gln    2512
7489  AGA CTT TCG CCC TTA GAC CGC CCC AAT CCT GGA GAG AAA GGC ACA CAA    7536

2513  Glu Ala Leu Thr Phe Arg Asp Asp Gly Ser Tyr Leu Ile Ala Gly Gly    2528
7537  GAG GCT CTT ACC TTC CGT GAT GAT GGC AGC TAT CTG ATT GCT GGT GGT    7584

2529  Leu Gly Gly Leu Gly Leu Val Val Ala Arg Phe Leu Val Thr Asn Gly    2544
7585  TTA GGC GGA CTG GGG TTA GTG GTG GCT CGT TTT CTG GTT ACA AAT GGG    7632

2545  Ala Lys Tyr Leu Val Leu Val Gly Arg Arg Gly Ala Arg Glu Glu Gln    2560
7633  GCT AAA TAC CTT GTG CTA GTC GGA CGA CGT GGT GCG AGG GAG GAA CAG    7680

2561  Gln Ala Gln Leu Ser Glu Leu Glu Gln Leu Gly Ala Ser Val Lys Val    2576
7681  CAA GCT CAA TTA AGC GAA CTA GAG CAA CTC GGA GCT TCC GTG AAA GTT    7728

2577  Leu Gln Ala Asp Ile Ala Asp Ala Glu Gln Leu Ala Gln Ala Leu Ser    2592
7729  TTA CAA GCC GAT ATT GCT GAT GCA GAA CAA CTA GCC CAA GCA CTT TCA    7776

2593  Ala Val Thr Tyr Pro Pro Leu Arg Gly Val Ile His Ala Ala Gly Thr    2608
7777  GCA GTA ACC TAC CCA CCA TTA CGG GGT GTT ATT CAT GCG GCA GGT ACA    7824

2609  Leu Asn Asp Gly Ile Leu Gln Gln Gln Ser Trp Gln Ala Phe Lys Glu    2624
7825  TTG AAC GAT GGG ATT CTA CAG CAG CAA AGT TGG CAA GCC TTT AAA GAA    7872

2625  Val Met Asn Pro Lys Val Ala Gly Ala Trp Asn Leu His Ile Leu Thr    2640
7873  GTG ATG AAT CCC AAG GTA GCA GGT GCG TGG AAC CTA CAT ATA CTG ACA    7920

2641  Lys Asn Gln Pro Leu Asp Phe Phe Val Leu Phe Ser Ser Ala Thr Ser    2656
7921  AAA AAT CAG CCT TTA GAC TTC TTT GTC CTG TTC TCC TCC GCC ACC TCT    7968

2657  Leu Leu Gly Asn Ala Gly Gln Ala Asn His Ala Ala Ala Asn Ala Phe    2672
7969  TTG TTA GGT AAC GCT GGA CAA GCC AAT CAC GCC GCC GCA AAT GCT TTC    8016

2673  Leu Asp Gly Leu Ala Ser Tyr Arg Arg His Leu Gly Leu Pro Ser Leu    2688
8017  CTT GAT GGG TTA GCC TCC TAT CGT CGT CAC TTA GGA CTA CCG AGC CTC    8064

2689  Ser Ile Asn Trp Gly Thr Trp Ser Glu Val Gly Ile Ala Ala Arg Leu    2704
8065  TCG ATT AAT TGG GGG ACA TGG AGC GAA GTG GGA ATT GCG GCT CGA CTT    8112

2705  Glu Leu Asp Lys Leu Ser Ser Lys Gln Gly Glu Gly Thr Ile Thr Leu    2720
8113  GAA CTA GAT AAG TTG TCC AGC AAA CAG GGA GAG GGA ACC ATT ACG CTA    8160

2721  Gly Gln Gly Leu Gln Ile Leu Glu Gln Leu Leu Lys Asp Glu Asn Gly    2736
8161  GGA CAG GGC TTA CAA ATT CTT GAG CAG TTG CTC AAA GAC GAG AAT GGG    8208

2737  Val Tyr Gln Val Gly Val Met Pro Ile Asn Trp Thr Gln Phe Leu Ala    2752
8209  GTG TAT CAA GTG GGT GTC ATG CCT ATC AAC TGG ACA CAA TTC TTA GCA    8256

2753  Arg Gln Leu Thr Pro Gln Pro Phe Phe Ser Asp Ala Met Lys Ser Ile    2768
8257  AGG CAA TTG ACT CCG CAG CCG TTC TTC AGC GAT GCC ATG AAG AGT ATT    8304

2769  Asp Thr Ser Val Gly Lys Leu Thr Leu Gln Glu Arg Asp Ser Cys Pro    2784
8305  GAC ACC TCT GTA GGT AAA CTA ACC TTG CAG GAG CGG GAC TCT TGC CCC    8352

2785  Gln Gly Tyr Gly His Asn Ile Arg Glu Gln Leu Glu Asn Ala Pro Pro    2800
8353  CAA GGT TAC GGG CAT AAT ATT CGA GAG CAA TTA GAG AAC GCT CCG CCC    8400

2801  Lys Glu Gly Leu Thr Leu Leu Gln Ala His Val Arg Glu Gln Val Ser    2816
8401  AAA GAG GGT CTG ACT CTC TTG CAG GCT CAT GTT CGG GAG CAG GTT TCC    8448
```

```
2817  Gln Val Leu Gly Ile Asp Thr Lys Thr Leu Leu Ala Glu Gln Asp Val    2832
8449  CAA GTT TTG GGG ATA GAC ACG AAG ACA TTA TTG GCA GAA CAA GAC GTG    8496

2833  Gly Phe Phe Thr Leu Gly Met Asp Ser Leu Thr Ser Val Glu Leu Arg    2848
8497  GGT TTC TTT ACC CTG GGG ATG GAT TCG CTG ACC TCT GTC GAG TTA AGA    8544

2849  Asn Arg Leu Gln Ala Ser Leu Gly Cys Ser Leu Ser Ser Thr Leu Ala    2864
8545  AAC AGG TTA CAA GCC AGT TTG GGC TGC TCT CTT TCT TCC ACT TTG GCT    8592

2865  Phe Asp Tyr Pro Thr Gln Gln Ala Leu Val Asn Tyr Leu Ala Asn Glu    2880
8593  TTT GAC TAT CCA ACA CAA CAG GCT CTT GTG AAT TAT CTT GCC AAT GAA    8640

2881  Leu Leu Gly Thr Pro Glu Gln Leu Gln Glu Pro Glu Ser Asp Glu Glu    2896
8641  TTG CTG GGA ACC CCT GAG CAG CTA CAA GAG CCT GAA TCT GAT GAA GAA    8688

2897  Asp Gln Ile Ser Ser Met Asp Asp Ile Val Gln Leu Leu Ser Ala Lys    2912
8689  GAT CAG ATA TCG TCA ATG GAT GAC ATC GTG CAG TTG CTG TCC GCG AAA    8736

2913  Leu Glu Met Glu Ile ***                                           2918
8737  CTA GAG ATG GAA ATT TAA                                           8754
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : CYNA.SEQ       SEQUENCE : NORMAL   5667 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -   5667

    TRANSLATION REGION :      1 -   5667


*** DNA TRANSLATION ***

   1  Met Asp Glu Lys Leu Arg Thr Tyr Glu Arg Leu Ile Lys Gln Ser Tyr      16
   1  ATG GAT GAA AAA CTA AGA ACA TAC GAA CGA TTA ATC AAG CAA TCC TAT      48

  17  His Lys Ile Glu Ala Leu Glu Ala Glu Val Asn Arg Leu Lys Gln Thr      32
  49  CAC AAG ATA GAG GCT CTG GAA GCT GAA GTT AAC AGG TTG AAG CAA ACC      96

  33  Gln Cys Glu Pro Ile Ala Ile Val Gly Met Gly Cys Arg Phe Pro Gly      48
  97  CAA TGT GAA CCT ATC GCC ATC GTC GGC ATG GGC TGT CGT TTT CCT GGT     144

  49  Ala Asn Ser Pro Glu Ala Phe Trp Gln Leu Leu Cys Asp Gly Val Asp      64
 145  GCG AAT AGT CCA GAA GCG TTT TGG CAG TTG TTG TGT GAT GGG GTT GAT     192

  65  Ala Ile Arg Glu Ile Pro Lys Asn Arg Trp Val Val Asp Ala Tyr Ile      80
 193  GCT ATT CGT GAG ATA CCA AAA AAT CGA TGG GTT GTT GAT GCC TAC ATA     240

  81  Asp Glu Asn Leu Asp Arg Ala Asp Lys Thr Ser Met Arg Phe Gly Gly      96
 241  GAT GAA AAT TTG GAC CGC GCA GAC AAG ACA TCA ATG CGA TTT GGC GGG     288

  97  Phe Val Glu Gln Leu Glu Lys Phe Asp Ala Gln Phe Phe Gly Ile Ser     112
 289  TTT GTC GAG CAA CTT GAG AAG TTT GAT GCC CAA TTC TTT GGC ATA TCA     336

 113  Pro Arg Glu Ala Val Ser Leu Asp Pro Gln Gln Arg Leu Leu Leu Glu     128
 337  CCG CGA GAA GCG GTT TCT CTT GAC CCT CAG CAA CGT TTG TTA TTA GAA     384

 129  Val Ser Trp Glu Ala Leu Glu Asn Ala Ala Val Ile Pro Pro Ser Ala     144
 385  GTA AGT TGG GAA GCA CTG GAA AAT GCA GCG GTG ATA CCA CCT TCG GCA     432

 145  Thr Gly Val Phe Val Gly Ile Ser Asn Leu Asp Tyr Arg Glu Thr Leu     160
 433  ACG GGC GTA TTC GTC GGT ATT AGT AAC CTT GAT TAT CGT GAA ACG CTC     480

 161  Leu Lys Gln Gly Ala Ile Gly Thr Tyr Phe Ala Ser Gly Asn Ala His     176
 481  TTG AAG CAA GGA GCA ATT GGT ACT TAT TTT GCT TCG GGT AAT GCC CAT     528

 177  Ser Thr Ala Ser Gly Arg Leu Ser Tyr Phe Leu Gly Leu Thr Gly Pro     192
 529  AGC ACA GCC AGT GGT CGC TTG TCT TAC TTT CTC GGT CTG ACA GGC CCC     576

 193  Cys Leu Ser Ile Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Val His     208
 577  TGT CTC TCG ATA GAT ACA GCT TGT TCT TCG TCG TTG GTC GCT GTA CAT     624

 209  Gln Ser Leu Ile Ser Leu Arg Gln Arg Glu Cys Asp Leu Ala Leu Val     224
 625  CAG TCA CTG ATA AGT CTG CGT CAG CGA GAA TGT GAC TTA GCG TTG GTT     672

 225  Gly Gly Val His Arg Leu Ile Ala Pro Glu Glu Ser Val Ser Leu Ala     240
 673  GGG GGA GTC CAT CGG CTG ATA GCC CCA GAG GAA AGT GTC TCG TTA GCA     720

 241  Lys Ala His Met Leu Ser Pro Asp Gly Arg Cys Lys Val Phe Asp Ala     256
 721  AAA GCC CAT ATG TTA TCT CCC GAT GGT CGT TGC AAA GTC TTT GAT GCG     768

 257  Ser Ala Asn Gly Tyr Val Arg Ala Glu Gly Cys Gly Met Ile Val Leu     272
 769  TCG GCA AAC GGG TAT GTC CGA GCC GAA GGA TGT GGC ATG ATA GTC CTC     816
```

```
 273  Lys Arg Leu Ser Asp Ala Gln Ala Asp Gly Asp Lys Ile Leu Ala Leu     288
 817  AAA CGA TTA TCG GAC GCG CAA GCT GAT GGG GAT AAA ATC TTG GCG TTG     864

 289  Ile Arg Gly Ser Ala Ile Asn Gln Asp Gly Arg Thr Ser Gly Leu Thr     304
 865  ATT CGC GGG TCA GCC ATA AAT CAA GAC GGT CGC ACG AGT GGC TTG ACC     912

 305  Val Pro Asn Gly Pro Gln Gln Ala Asp Val Ile Arg Gln Ala Leu Ala     320
 913  GTT CCA AAT GGT CCC CAA CAA GCC GAC GTG ATT CGC CAA GCC CTC GCC     960

 321  Asn Ser Gly Ile Arg Pro Glu Gln Val Asn Tyr Val Glu Ala His Gly     336
 961  AAT AGT GGC ATA AGA CCA GAA CAA GTT AAC TAT GTA GAA GCT CAT GGC    1008

 337  Thr Gly Thr Ser Leu Gly Asp Pro Ile Glu Val Gly Ala Leu Gly Thr     352
1009  ACA GGG ACT TCC CTA GGA GAC CCG ATT GAG GTC GGC GCG TTG GGA ACG    1056

 353  Ile Phe Asn Gln Arg Ser Gln Pro Leu Ile Ile Gly Ser Val Lys Thr     368
1057  ATC TTT AAT CAA CGC TCC CAA CCT TTA ATT ATT GGT TCA GTT AAA ACA    1104

 369  Asn Ile Gly His Leu Glu Ala Ala Ala Gly Ile Ala Gly Leu Ile Lys     384
1105  AAT ATT GGG CAT CTA GAA GCA GCA GCA GGG ATT GCT GGA CTG ATT AAA    1152

 385  Val Val Leu Ala Met Gln His Gly Glu Ile Pro Pro Asn Leu His Phe     400
1153  GTC GTC CTT GCC ATG CAG CAT GGA GAA ATT CCA CCT AAT TTA CAC TTT    1200

 401  His Gln Pro Asn Pro Arg Ile Asn Trp Asp Lys Leu Pro Ile Arg Ile     416
1201  CAC CAG CCC AAT CCT CGC ATT AAC TGG GAT AAA TTG CCA ATC AGG ATC    1248

 417  Pro Thr Glu Arg Thr Ala Trp Pro Thr Gly Asp Arg Ile Ala Gly Ile     432
1249  CCC ACA GAA CGA ACA GCT TGG CCT ACT GGC GAT CGC ATC GCA GGG ATA    1296

 433  Ser Ser Phe Gly Phe Ser Gly Thr Asn Ser His Val Val Leu Glu Glu     448
1297  AGT TCT TTC GGC TTT AGT GGC ACT AAT TCT CAT GTC GTG TTA GAG GAA    1344

 449  Ala Pro Lys Ile Glu Pro Ser Thr Leu Glu Ile His Ser Lys Gln Tyr     464
1345  GCC CCA AAA ATA GAG CCG TCT ACT TTA GAG ATT CAT TCA AAG CAG TAT    1392

 465  Val Phe Thr Leu Ser Ala Ala Thr Pro Gln Ala Leu Gln Glu Leu Thr     480
1393  GTT TTT ACC TTA TCA GCA GCG ACA CCT CAA GCA CTA CAA GAA CTT ACT    1440

 481  Gln Arg Tyr Val Thr Tyr Leu Thr Glu His Leu Gln Glu Ser Leu Ala     496
1441  CAG CGT TAT GTA ACT TAT CTC ACT GAA CAC TTA CAA GAG AGT CTG GCG    1488

 497  Asp Ile Cys Phe Thr Ala Asn Thr Gly Arg Lys His Phe Arg His Arg     512
1489  GAT ATT TGC TTT ACA GCC AAC ACA GGG CGC AAA CAC TTT AGA CAT CGC    1536

 513  Phe Ala Val Val Ala Glu Ser Lys Thr Gln Leu Arg Gln Gln Leu Glu     528
1537  TTT GCA GTA GTA GCA GAG TCT AAA ACC CAG TTG CGC CAA CAA TTG GAA    1584

 529  Thr Phe Ala Gln Ser Gly Glu Gly Gln Gly Lys Arg Thr Ser Leu Ser     544
1585  ACG TTT GCC CAA TCG GGA GAG GGG CAG GGG AAG AGG ACA TCT CTC TCA    1632

 545  Lys Ile Ala Phe Leu Phe Thr Gly Gln Gly Ser Gln Tyr Val Gly Met     560
1633  AAA ATA GCT TTT CTC TTT ACA GGT CAA GGC TCA CAG TAT GTG GGG ATG    1680

 561  Gly Gln Glu Leu Tyr Glu Ser Gln Pro Thr Phe Arg Gln Thr Ile Asp     576
1681  GGG CAA GAA CTT TAT GAG AGC CAA CCC ACC TTC CGG CAA ACC ATT GAC    1728

 577  Arg Cys Asp Glu Ile Leu Arg Ser Leu Leu Gly Lys Ser Ile Leu Ser     592
1729  CGA TGT GAT GAG ATT CTT CGT TCA CTG TTG GGC AAA TCA ATC CTC TCA    1776

 593  Ile Leu Tyr Pro Ser Gln Gln Met Gly Leu Glu Thr Pro Ser Gln Ile     608
1777  ATA CTC TAT CCC AGC CAA CAA ATG GGA TTG GAA ACG CCA TCC CAA ATT    1824

 609  Asp Glu Thr Ala Tyr Thr Gln Pro Thr Leu Phe Ser Leu Glu Tyr Ala     624
1825  GAT GAA ACC GCC TAT ACT CAA CCC ACT CTT TTT TCT CTT GAA TAT GCA    1872

 625  Leu Ala Gln Leu Trp Arg Ser Trp Gly Ile Glu Pro Asp Val Val Met     640
1873  CTG GCG CAG TTG TGG CGC TCC TGG GGT ATT GAG CCT GAT GTG GTG ATG    1920
```

```
 641  Gly His Ser Val Gly Glu Tyr Val Ala Ala Cys Val Ala Gly Val Phe     656
1921  GGG CAT AGT GTG GGA GAA TAT GTG GCC GCT TGT GTG GCG GGT GTC TTT    1968

 657  Ser Leu Glu Asp Gly Leu Lys Leu Ile Ala Glu Arg Gly Arg Leu Met     672
1969  TCT TTA GAG GAT GGA CTC AAA CTA ATT GCT GAA AGA GGC CGT CTG ATG    2016

 673  Gln Glu Leu Pro Pro Asp Gly Ala Met Val Ser Val Met Ala Asn Lys     688
2017  CAA GAA TTG CCT CCC GAT GGG GCG ATG GTT TCA GTT ATG GCC AAT AAA    2064

 689  Ser Arg Ile Glu Gln Ala Ile Gln Ser Val Ser Arg Glu Val Ser Ile     704
2065  TCG CGC ATA GAG CAA GCA ATT CAA TCT GTC AGC CGA GAG GTT TCT ATT    2112

 705  Ala Ala Ile Asn Gly Pro Glu Ser Val Val Ile Ser Gly Lys Arg Glu     720
2113  GCG GCC ATC AAT GGA CCT GAG AGT GTG GTT ATC TCT GGT AAA AGG GAG    2160

 721  Ile Leu Gln Gln Ile Thr Glu His Leu Val Ala Glu Gly Ile Lys Thr     736
2161  ATA TTA CAA CAG ATT ACC GAA CAT CTG GTT GCC GAA GGC ATT AAG ACA    2208

 737  Arg Gln Leu Lys Val Ser His Ala Phe His Ser Pro Leu Met Glu Pro     752
2209  CGC CAA CTG AAG GTC TCT CAT GCC TTT CAC TCA CCA TTG ATG GAG CCA    2256

 753  Ile Leu Gly Gln Phe Arg Arg Val Ala Asn Thr Ile Thr Tyr Arg Pro     768
2257  ATA TTA GGT CAG TTC CGC CGA GTT GCC AAT ACC ATC ACC TAT CGG CCA    2304

 769  Pro Gln Ile Asn Leu Val Ser Asn Val Thr Gly Gly Gln Val Tyr Lys     784
2305  CCG CAA ATT AAC CTT GTC TCA AAT GTC ACA GGC GGA CAG GTG TAT AAA    2352

 785  Glu Ile Ala Thr Pro Asp Tyr Trp Val Arg His Leu Gln Glu Thr Val     800
2353  GAA ATC GCT ACT CCC GAT TAT TGG GTG AGA CAT CTG CAA GAG ACT GTC    2400

 801  Arg Phe Ala Asp Gly Val Lys Val Leu His Glu Gln Asn Val Asn Phe     816
2401  CGT TTT GCG GAT GGG GTT AAG GTG TTA CAT GAA CAG AAT GTC AAT TTC    2448

 817  Met Leu Glu Ile Gly Pro Lys Pro Thr Leu Leu Gly Met Val Glu Leu     832
2449  ATG CTC GAA ATT GGT CCC AAA CCC ACA CTG CTG GGC ATG GTT GAG TTA    2496

 833  Gln Ser Ser Glu Asn Pro Phe Ser Met Pro Met Met Met Pro Ser Leu     848
2497  CAA AGT TCT GAG AAT CCA TTT TCT ATG CCA ATG ATG ATG CCC AGT TTG    2544

 849  Arg Gln Asn Arg Ser Asp Trp Gln Gln Met Leu Glu Ser Leu Ser Gln     864
2545  CGT CAG AAT CGT AGC GAC TGG CAG CAG ATG TTG GAG AGC TTG AGT CAA    2592

 865  Leu Tyr Val His Gly Val Glu Ile Asp Trp Ile Gly Phe Asn Lys Asp     880
2593  CTC TAT GTT CAT GGT GTT GAG ATT GAC TGG ATC GGT TTT AAT AAA GAC    2640

 881  Tyr Val Arg His Lys Val Val Leu Pro Thr Tyr Pro Trp Gln Lys Glu     896
2641  TAT GTG CGA CAT AAA GTT GTC CTG CCG ACA TAC CCA TGG CAG AAG GAG    2688

 897  Arg Tyr Trp Val Glu Leu Asp Gln Gln Lys His Ala Ala Lys Asn Leu     912
2689  CGT TAC TGG GTA GAA TTG GAT CAA CAG AAG CAC GCC GCT AAA AAT CTA    2736

 913  His Pro Leu Leu Asp Arg Cys Met Lys Leu Pro Arg His Asn Glu Thr     928
2737  CAT CCT CTA CTG GAC AGG TGC ATG AAG CTG CCT CGT CAT AAC GAA ACA    2784

 929  Ile Phe Glu Lys Glu Phe Ser Leu Glu Thr Leu Pro Phe Leu Ala Asp     944
2785  ATT TTT GAG AAA GAA TTT AGT CTA GAG ACA TTG CCC TTT CTT GCT GAC    2832

 945  Tyr Arg Ile Tyr Gly Ser Val Val Ser Pro Gly Ala Ser Tyr Leu Ser     960
2833  TAT CGC ATT TAT GGT TCA GTT GTG TCG CCA GGT GCA AGT TAT CTA TCA    2880

 961  Met Ile Leu Ser Ile Ala Glu Ser Tyr Ala Asn Gly His Leu Asn Gly     976
2881  ATG ATA CTA AGT ATT GCC GAG TCG TAT GCA AAT GGT CAT TTG AAT GGA    2928

 977  Gly Asn Ser Ala Lys Gln Thr Thr Tyr Leu Leu Lys Asp Val Thr Phe     992
2929  GGG AAT AGT GCA AAG CAA ACC ACT TAT TTA CTA AAG GAT GTC ACA TTC    2976

 993  Pro Val Pro Leu Val Ile Ser Asp Glu Ala Asn Tyr Met Val Gln Val    1008
```

```
2977  CCA GTA CCT CTT GTG ATC TCT GAT GAG GCA AAT TAC ATG GTG CAA GTT    3024

1009  Ala Cys Ser Leu Ser Cys Ala Ala Pro His Asn Arg Gly Asp Glu Thr    1024
3025  GCT TGT TCT CTC TCT TGT GCT GCG CCA CAC AAT CGT GGC GAC GAG ACG    3072

1025  Gln Phe Glu Leu Phe Ser Phe Ala Glu Asn Val Pro Glu Ser Ser Ser    1040
3073  CAG TTT GAA TTG TTC AGT TTT GCT GAG AAT GTA CCT GAA AGT AGC AGT    3120

1041  Ile Asn Ala Asp Phe Gln Thr Pro Ile Ile His Ala Lys Gly Gln Phe    1056
3121  ATA AAT GCT GAT TTT CAG ACA CCC ATT ATT CAT GCA AAA GGG CAA TTT    3168

1057  Lys Leu Glu Asp Thr Ala Pro Pro Lys Val Glu Leu Glu Glu Leu Gln    1072
3169  AAG CTT GAA GAT ACA GCA CCT CCT AAA GTG GAG CTA GAA GAA CTA CAA    3216

1073  Ala Gly Cys Pro Gln Glu Ile Asp Leu Asn Leu Phe Tyr Gln Thr Phe    1088
3217  GCG GGT TGT CCC CAA GAA ATT GAT CTC AAC CTT TTC TAT CAA ACA TTC    3264

1089  Thr Asp Lys Gly Phe Val Phe Gly Ser Arg Phe Arg Trp Leu Glu Gln    1104
3265  ACA GAC AAA GGT TTT GTT TTT GGA TCT CGT TTT CGC TGG TTA GAA CAA    3312

1105  Ile Trp Val Gly Asp Gly Glu Ala Leu Ala Arg Leu Arg Gln Pro Glu    1120
3313  ATC TGG GTG GGC GAT GGA GAA GCA TTG GCG CGT CTG CGA CAA CCG GAA    3360

1121  Ser Ile Glu Ser Phe Lys Gly Tyr Val Ile His Pro Gly Leu Leu Asp    1136
3361  AGT ATT GAA TCG TTT AAA GGA TAT GTG ATT CAT CCC GGT TTG TTG GAT    3408

1137  Ala Cys Thr Gln Val Pro Phe Ala Ile Ser Ser Asp Asp Glu Asn Arg    1152
3409  GCC TGT ACA CAA GTC CCA TTT GCA ATT TCG TCT GAC GAT GAA AAT AGG    3456

1153  Gln Ser Glu Thr Thr Met Pro Phe Ala Leu Asn Glu Leu Arg Cys Tyr    1168
3457  CAA TCA GAA ACG ACA ATG CCC TTT GCG CTG AAT GAA TTA CGT TGT TAT    3504

1169  Gln Pro Ala Asn Gly Gln Met Trp Trp Val His Ala Thr Glu Lys Asp    1184
3505  CAG CCT GCA AAC GGA CAA ATG TGG TGG GTT CAT GCA ACA GAA AAA GAT    3552

1185  Arg Tyr Thr Trp Asp Val Ser Leu Phe Asp Glu Ser Gly Gln Val Ile    1200
3553  AGA TAT ACA TGG GAT GTT TCT CTG TTT GAT GAG AGC GGG CAA GTT ATT    3600

1201  Ala Glu Phe Ile Gly Leu Glu Val Arg Ala Ala Met Pro Glu Gly Leu    1216
3601  GCG GAA TTT ATA GGT TTA GAA GTT CGT GCT GCT ATG CCC GAA GGC TTA    3648

1217  Leu Arg Ala Asp Phe Trp His Asn Trp Leu Tyr Thr Val Asn Trp Arg    1232
3649  CTA AGG GCA GAC TTT TGG CAT AAC TGG CTC TAT ACA GTG AAT TGG CGA    3696

1233  Ser Gln Pro Leu Gln Ile Pro Glu Val Leu Asp Ile Asn Lys Thr Gly    1248
3697  TCG CAA CCT CTA CAA ATC CCA GAG GTG CTG GAT ATT AAT AAG ACA GGT    3744

1249  Ala Glu Thr Trp Leu Leu Phe Ala Gln Pro Glu Gly Ile Gly Ala Asp    1264
3745  GCA GAA ACA TGG CTT CTT TTT GCA CAA CCA GAG GGA ATA GGA GCG GAC    3792

1265  Leu Ala Glu Tyr Leu Gln Ser Gln Gly Lys His Cys Val Phe Val Val    1280
3793  TTA GCC GAA TAT TTG CAG AGC CAA GGA AAG CAC TGT GTT TTT GTA GTG    3840

1281  Pro Gly Ser Glu Tyr Thr Val Thr Glu Gln His Ile Gly Arg Thr Gly    1296
3841  CCT GGG AGT GAG TAT ACA GTG ACC GAG CAA CAC ATT GGA CGC ACT GGA    3888

1297  His Leu Asp Val Thr Lys Leu Thr Lys Ile Val Thr Ile Asn Pro Ala    1312
3889  CAT CTT GAT GTG ACG AAA CTG ACA AAA ATT GTC ACG ATC AAT CCT GCT    3936

1313  Ser Pro His Asp Tyr Lys Tyr Phe Leu Glu Thr Leu Thr Asp Ile Arg    1328
3937  TCT CCT CAT GAC TAT AAA TAT TTT TTA GAA ACT CTG ACG GAC ATT AGA    3984

1329  Leu Pro Cys Glu His Ile Leu Tyr Leu Trp Asn Arg Tyr Asp Leu Thr    1344
3985  TTA CCT TGT GAA CAT ATA CTC TAT TTA TGG AAT CGT TAT GAT TTA ACA    4032

1345  Asn Thr Ser Asn His Arg Thr Glu Leu Thr Val Pro Asp Ile Val Leu    1360
4033  AAT ACT TCT AAT CAT CGG ACA GAA TTG ACT GTA CCA GAT ATA GTC TTA    4080
```

```
1361  Asn Leu Cys Thr Ser Leu Thr Tyr Leu Val Gln Ala Leu Ser His Met    1376
4081  AAC TTA TGT ACT AGT CTT ACT TAT TTG GTA CAA GCC CTT AGC CAC ATG    4128

1377  Gly Phe Ser Pro Lys Leu Trp Leu Ile Thr Gln Asn Ser Gln Ala Val    1392
4129  GGT TTT TCC CCG AAA TTA TGG CTA ATT ACA CAA AAT AGT CAA GCG GTT    4176

1393  Gly Ser Asp Leu Ala Asn Leu Glu Ile Glu Gln Ser Pro Leu Trp Ala    1408
4177  GGT AGT GAC TTA GCG AAT TTA GAA ATC GAA CAA TCC CCA TTA TGG GCA    4224

1409  Leu Gly Arg Ser Ile Arg Ala Glu His Pro Glu Phe Asp Cys Arg Cys    1424
4225  TTG GGT CGA AGC ATC CGC GCC GAA CAC CCT GAA TTT GAT TGC CGT TGT    4272

1425  Leu Asp Phe Asp Thr Leu Ser Asn Ile Ala Pro Leu Leu Leu Lys Glu    1440
4273  TTA GAT TTT GAC ACG CTC TCA AAT ATC GCA CCA CTC TTG TTG AAA GAG    4320

1441  Met Gln Ala Ile Asp Tyr Glu Ser Gln Ile Ala Tyr Arg Gln Gly Thr    1456
4321  ATG CAA GCT ATA GAC TAT GAA TCT CAA ATT GCT TAC CGA CAA GGA ACG    4368

1457  Arg Tyr Val Ala Arg Leu Ile Arg Asn Gln Ser Glu Cys His Ala Pro    1472
4369  CGC TAT GTT GCA CGA CTA ATT CGT AAT CAA TCA GAA TGT CAC GCA CCG    4416

1473  Ile Gln Thr Gly Ile Arg Pro Asp Gly Ser Tyr Leu Ile Thr Gly Gly    1488
4417  ATT CAA ACA GGA ATC CGT CCT GAT GGC AGC TAT TTG ATT ACA GGT GGA    4464

1489  Leu Gly Gly Leu Gly Leu Gln Val Ala Leu Ala Leu Ala Asp Ala Gly    1504
4465  TTA GGC GGT CTA GGA TTG CAG GTA GCA CTC GCC CTT GCG GAC GCT GGA    4512

1505  Ala Arg His Leu Ile Leu Asn Ser Arg Arg Gly Thr Val Ser Lys Glu    1520
4513  GCA AGA CAC TTG ATC CTC AAT AGT CGC CGT GGT ACG GTC TCC AAA GAA    4560

1521  Ala Gln Leu Ile Ile Asp Arg Leu Arg Gln Glu Asp Val Arg Val Asp    1536
4561  GCC CAG TTA ATT ATT GAC CGA CTA CGC CAA GAG GAT GTT AGG GTT GAT    4608

1537  Leu Ile Ala Ala Asp Val Ser Asp Ala Ala Asp Ser Glu Arg Leu Leu    1552
4609  TTG ATT GCG GCA GAT GTC TCT GAT GCG GCA GAT AGC GAA CGA CTC TTA    4656

1553  Val Glu Ser Gln Arg Lys Thr Ser Leu Arg Gly Ile Val His Val Ala    1568
4657  GTA GAA AGT CAG CGC AAG ACC TCT CTT CGA GGG ATT GTC CAT GTT GCG    4704

1569  Gly Val Leu Asp Asp Gly Ile Leu Leu Gln Gln Asn Gln Glu Arg Phe    1584
4705  GGA GTC TTG GAT GAT GGC ATC CTG CTC CAA CAA AAT CAA GAG CGT TTT    4752

1585  Glu Lys Val Met Ala Ala Lys Val Arg Gly Ala Trp His Leu Asp Gln    1600
4753  GAA AAA GTG ATG GCG GCT AAG GTA CGC GGA GCT TGG CAT CTG GAC CAA    4800

1601  Gln Ser Gln Thr Leu Asp Leu Asp Phe Phe Val Ala Phe Ser Ser Val    1616
4801  CAG AGC CAA ACC CTC GAT TTA GAT TTC TTT GTT GCG TTC TCA TCT GTT    4848

1617  Ala Ser Leu Ile Glu Glu Pro Gly Gln Ala Asn Tyr Ala Ala Ala Asn    1632
4849  GCG TCG CTC ATA GAA GAA CCA GGA CAA GCC AAT TAC GCC GCA GCG AAT    4896

1633  Ala Phe Leu Asp Ser Leu Met Tyr Tyr Arg His Ile Lys Gly Ser Asn    1648
4897  GCG TTT TTG GAT TCA TTA ATG TAT TAT CGT CAC ATA AAG GGA TCT AAT    4944

1649  Ser Leu Ser Ile Asn Trp Gly Ala Trp Ala Glu Val Gly Met Ala Ala    1664
4945  AGC TTG AGT ATC AAC TGG GGG GCT TGG GCA GAA GTC GGC ATG GCA GCC    4992

1665  Asn Leu Ser Trp Glu Gln Arg Gly Ile Ala Ala Ile Ser Pro Lys Gln    1680
4993  AAT TTA TCA TGG GAA CAA CGG GGA ATC GCG GCA ATT TCT CCA AAG CAA    5040

1681  Gly Arg His Ile Leu Val Gln Leu Ile Gln Lys Leu Asn Gln His Thr    1696
5041  GGG AGG CAT ATT CTC GTC CAA CTT ATT CAA AAA CTT AAT CAG CAT ACA    5088

1697  Ile Pro Gln Val Ala Val Gln Pro Thr Asn Trp Ala Glu Tyr Leu Ser    1712
5089  ATC CCC CAA GTT GCT GTA CAA CCG ACC AAT TGG GCT GAA TAT CTA TCC    5136

1713  His Asp Gly Val Asn Met Pro Phe Tyr Glu Tyr Phe Thr His His Leu    1728
5137  CAT GAT GGC GTG AAT ATG CCA TTC TAT GAA TAT TTT ACA CAC CAC TTG    5184
```

```
1729  Arg Asn Glu Lys Glu Ala Lys Leu Arg Gln Thr Ala Gly Ser Thr Ser    1744
5185  CGT AAC GAA AAA GAA GCC AAA TTG CGG CAA ACA GCA GGC AGC ACC TCA    5232

1745  Glu Glu Val Ser Leu Arg Gln Gln Leu Gln Thr Leu Ser Glu Lys Asp    1760
5233  GAG GAA GTC AGT CTG CGG CAA CAG CTT CAA ACA CTC TCA GAG AAA GAC    5280

1761  Arg Asp Ala Leu Leu Met Glu His Leu Gln Lys Thr Ala Ile Arg Val    1776
5281  CGG GAT GCC CTT TTG ATG GAA CAT CTT CAA AAA ACT GCG ATC AGA GTT    5328

1777  Leu Gly Leu Ala Ser Asn Gln Lys Ile Asp Pro Tyr Gln Gly Leu Met    1792
5329  CTC GGT TTG GCA TCT AAT CAA AAA ATT GAT CCC TAT CAG GGA TTG ATG    5376

1793  Asn Met Gly Leu Asp Ser Leu Met Ala Val Glu Phe Arg Asn His Leu    1808
5377  AAT ATG GGA CTA GAC TCT TTG ATG GCG GTT GAA TTT CGG AAT CAC TTG    5424

1809  Ile Arg Ser Leu Glu Arg Pro Leu Pro Ala Thr Leu Leu Phe Asn Cys    1824
5425  ATA CGT AGT TTA GAA CGC CCT CTG CCA GCC ACT CTG CTC TTT AAT TGC    5472

1825  Pro Thr Leu Asp Ser Leu His Asp Tyr Leu Val Ala Lys Met Phe Asp    1840
5473  CCA ACA CTT GAT TCA TTG CAT GAT TAC CTA GTC GCA AAA ATG TTT GAT    5520

1841  Asp Ala Pro Gln Lys Ala Glu Gln Met Ala Gln Pro Thr Thr Leu Thr    1856
5521  GAT GCC CCT CAG AAG GCA GAG CAA ATG GCA CAA CCA ACA ACA CTG ACA    5568

1857  Ala His Ser Ile Ser Ile Glu Ser Lys Ile Asp Asp Asn Glu Ser Val    1872
5569  GCA CAC AGC ATA TCA ATA GAA TCC AAA ATA GAT GAT AAC GAA AGC GTG    5616

1873  Asp Asp Ile Ala Gln Met Leu Ala Gln Ala Leu Asn Ile Ala Phe Glu    1888
5617  GAT GAC ATT GCA CAA ATG CTG GCA CAA GCA CTC AAT ATC GCC TTT GAG    5664

1889  ***                                                               1889
5665  TAG                                                               5667
```

*** INPUT INFORMATION ***

FILE NAME : *CYNB*.SEQ SEQUENCE : NORMAL 5004 BP

CODON TABLE : UNIV.TCN

SEQUENCE REGION : 1 - 5004

TRANSLATION REGION : 1 - 5004

*** DNA TRANSLATION ***

```
   1  Met Ser Gln Pro Asn Tyr Gly Ile Leu Met Lys Asn Ala Leu Asn Glu      16
   1  ATG AGT CAG CCC AAT TAT GGC ATT TTG ATG AAA AAT GCG TTG AAC GAA      48

  17  Ile Asn Ser Leu Arg Ser Gln Leu Ala Ala Val Glu Ala Gln Lys Asn      32
  49  ATA AAT AGC CTA CGA TCG CAA CTA GCT GCG GTA GAA GCC CAA AAA AAT      96

  33  Glu Ser Ile Ala Ile Val Gly Met Ser Cys Arg Phe Pro Gly Gly Ala      48
  97  GAG TCT ATT GCC ATT GTT GGT ATG AGT TGC CGT TTT CCA GGC GGT GCA     144

  49  Thr Thr Pro Glu Arg Phe Trp Val Leu Leu Arg Glu Gly Ile Ser Ala      64
 145  ACT ACT CCA GAG CGT TTT TGG GTA TTA CTG CGC GAG GGT ATA TCA GCC     192

  65  Ile Thr Glu Ile Pro Ala Asp Arg Trp Asp Val Asp Lys Tyr Tyr Asp      80
 193  ATT ACA GAA ATC CCT GCT GAT CGC TGG GAT GTT GAT AAA TAT TAT GAT     240

  81  Ala Asp Pro Thr Ser Ser Gly Lys Met His Thr Arg Tyr Gly Gly Phe      96
 241  GCT GAC CCC ACA TCG TCC GGT AAA ATG CAT ACT CGT TAC GGC GGT TTT     288

  97  Leu Asn Glu Val Asp Thr Phe Glu Pro Ser Phe Phe Asn Ile Ala Ala     112
 289  CTG AAT GAA GTT GAT ACA TTT GAG CCA TCA TTC TTT AAT ATT GCT GCC     336

 113  Arg Glu Ala Val Ser Met Asp Pro Gln Gln Arg Leu Leu Leu Glu Val     128
 337  CGT GAA GCC GTT AGC ATG GAT CCA CAG CAA CGC TTG CTA CTT GAA GTC     384

 129  Ser Trp Glu Ala Leu Glu Ser Gly Asn Ile Val Pro Ala Thr Leu Phe     144
 385  AGT TGG GAA GCT CTG GAA TCC GGT AAT ATT GTT CCT GCA ACT CTT TTT     432

 145  Asp Ser Ser Thr Gly Val Phe Ile Gly Ile Gly Gly Ser Asn Tyr Lys     160
 433  GAT AGT TCC ACT GGT GTA TTT ATC GGT ATT GGT GGT AGC AAC TAC AAA     480

 161  Ser Leu Met Ile Glu Asn Arg Ser Arg Ile Gly Lys Thr Asp Leu Tyr     176
 481  TCT TTA ATG ATC GAA AAC AGG AGT CGG ATC GGG AAA ACC GAT TTG TAT     528

 177  Glu Leu Ser Gly Thr Asp Val Ser Val Ala Ala Gly Arg Ile Ser Tyr     192
 529  GAG TTA AGT GGC ACT GAT GTG AGT GTT GCT GCC GGC AGG ATA TCC TAT     576

 193  Val Leu Gly Leu Met Gly Pro Ser Phe Val Ile Asp Thr Ala Cys Ser     208
 577  GTC CTG GGT TTG ATG GGT CCC AGT TTT GTG ATT GAT ACA GCT TGT TCA     624

 209  Ser Ser Leu Val Ser Val His Gln Ala Cys Gln Ser Leu Arg Gln Arg     224
 625  TCT TCT TTG GTC TCA GTT CAT CAA GCC TGT CAG AGT CTG CGT CAG AGA     672

 225  Glu Cys Asp Leu Ala Leu Ala Gly Gly Val Gly Leu Leu Ile Asp Pro     240
 673  GAA TGT GAT CTA GCA CTA GCT GGT GGA GTC GGT TTA CTC ATT GAT CCA     720

 241  Asp Glu Met Ile Gly Leu Ser Gln Gly Gly Met Leu Ala Pro Asp Gly     256
 721  GAT GAG ATG ATT GGT CTT TCT CAA GGG GGG ATG CTG GCA CCT GAT GGT     768

 257  Ser Cys Lys Thr Phe Asp Ala Asn Ala Asn Gly Tyr Val Arg Gly Glu     272
 769  AGT TGT AAA ACA TTT GAT GCC AAT GCA AAT GGC TAT GTG CGA GGC GAA     816
```

```
 273  Gly Cys Gly Met Ile Val Leu Lys Arg Leu Ser Asp Ala Thr Ala Asp     288
 817  GGT TGT GGG ATG ATT GTT CTA AAA CGT CTC TCG GAT GCA ACA GCC GAT     864

 289  Gly Asp Asn Ile Leu Ala Ile Ile Arg Gly Ser Met Val Asn His Asp     304
 865  GGG GAT AAT ATT CTT GCC ATC ATT CGT GGG TCT ATG GTT AAT CAT GAT     912

 305  Gly His Ser Ser Gly Leu Thr Ala Pro Arg Gly Pro Ala Gln Val Ser     320
 913  GGT CAT AGC AGT GGT TTA ACT GCT CCA AGA GGC CCC GCA CAA GTC TCT     960

 321  Val Ile Lys Gln Ala Leu Asp Arg Ala Gly Ile Ala Pro Asp Ala Val     336
 961  GTC ATT AAG CAA GCC TTA GAT AGA GCA GGT ATT GCA CCG GAT GCC GTA    1008

 337  Ser Tyr Leu Glu Ala His Gly Thr Gly Thr Pro Leu Gly Asp Pro Ile     352
1009  AGT TAT TTA GAA GCC CAT GGT ACA GGC ACA CCC CTT GGT GAT CCT ATC    1056

 353  Glu Met Asp Ser Leu Asn Glu Val Phe Gly Arg Arg Thr Glu Pro Leu     368
1057  GAG ATG GAT TCA TTG AAC GAA GTG TTT GGT CGG AGA ACA GAA CCA CTT    1104

 369  Trp Val Gly Ser Val Lys Thr Asn Ile Gly His Leu Glu Ala Ala Ser     384
1105  TGG GTC GGC TCA GTT AAG ACA AAT ATT GGT CAT TTA GAA GCC GCG TCC    1152

 385  Gly Ile Ala Gly Leu Ile Lys Val Val Leu Met Leu Lys Asn Lys Gln     400
1153  GGT ATT GCA GGG CTG ATT AAG GTT GTC TTG ATG CTA AAA AAC AAG CAG    1200

 401  Ile Pro Pro His Leu His Phe Lys Thr Pro Asn Pro Tyr Ile Asp Trp     416
1201  ATT CCT CCT CAC TTG CAT TTC AAG ACA CCA AAT CCA TAT ATT GAT TGG    1248

 417  Lys Asn Leu Pro Val Glu Ile Pro Thr Thr Leu His Ala Trp Asp Asp     432
1249  AAA AAT CTC CCG GTC GAA ATT CCG ACC ACC CTT CAT GCT TGG GAT GAC    1296

 433  Lys Thr Leu Lys Asp Arg Lys Arg Ile Ala Gly Val Ser Ser Phe Ser     448
1297  AAG ACA TTG AAG GAC AGA AAG CGA ATT GCA GGG GTT AGT TCT TTT AGT    1344

 449  Phe Ser Gly Thr Asn Ala His Ile Val Leu Ser Glu Ala Pro Ser Ser     464
1345  TTC AGT GGT ACT AAC GCC CAC ATT GTA TTA TCT GAA GCC CCA TCT AGC    1392

 465  Glu Leu Ile Ser Asn His Ala Ala Val Glu Arg Pro Trp His Leu Leu     480
1393  GAA CTA ATT AGT AAT CAT GCG GCA GTG GAA AGA CCA TGG CAC TTG TTA    1440

 481  Thr Leu Ser Ala Lys Asn Glu Glu Ala Leu Ala Asn Leu Val Gly Leu     496
1441  ACC CTT AGT GCT AAG AAT GAG GAA GCG TTG GCT AAC TTG GTT GGG CTT    1488

 497  Tyr Gln Ser Phe Ile Ser Thr Thr Asp Ala Ser Leu Ala Asp Ile Cys     512
1489  TAT CAG TCA TTT ATT TCT ACT ACT GAT GCA AGT CTT GCC GAT ATA TGC    1536

 513  Tyr Thr Ala Asn Thr Ala Arg Thr His Phe Ser His Arg Leu Ala Leu     528
1537  TAC ACT GCT AAT ACG GCA CGA ACC CAT TTT TCT CAT CGC CTT GCT CTA    1584

 529  Ser Ala Thr Ser His Ile Gln Ile Glu Ala Leu Leu Ala Ala Tyr Lys     544
1585  TCG GCT ACT TCA CAC ATC CAA ATA GAG GCT CTT TTA GCC GCT TAT AAG    1632

 545  Glu Gly Ser Val Ser Leu Ser Ile Asn Gln Gly Cys Val Leu Ser Asn     560
1633  GAA GGG TCG GTG AGT TTG AGC ATC AAT CAA GGT TGT GTC CTT TCC AAC    1680

 561  Ser Arg Ala Pro Lys Val Ala Phe Leu Phe Thr Gly Gln Gly Ser Gln     576
1681  AGT CGT GCG CCG AAG GTC GCT TTT CTC TTT ACA GGT CAA GGT TCG CAA    1728

 577  Tyr Val Gln Met Ala Gly Glu Leu Tyr Glu Thr Gln Pro Thr Phe Arg     592
1729  TAT GTG CAA ATG GCT GGA GAA CTT TAT GAG ACC CAG CCT ACT TTC CGT    1776

 593  Asn Cys Leu Asp Arg Cys Ala Glu Ile Leu Gln Ser Ile Phe Ser Ser     608
1777  AAT TGC TTA GAT CGC TGT GCC GAA ATC TTG CAA TCC ATC TTT TCA TCG    1824

 609  Arg Asn Ser Pro Trp Gly Asn Pro Leu Leu Ser Val Leu Tyr Pro Asn     624
1825  AGA AAC AGC CCT TGG GGA AAC CCA CTG CTT TCG GTA TTA TAT CCA AAC    1872

 625  His Glu Ser Lys Glu Ile Asp Gln Thr Ala Tyr Thr Gln Pro Ala Leu     640
1873  CAT GAG TCA AAG GAA ATT GAC CAG ACG GCT TAT ACC CAA CCT GCC CTT    1920
```

```
 641  Phe Ala Val Glu Tyr Ala Leu Ala Gln Met Trp Arg Ser Trp Gly Ile     656
1921  TTT GCT GTA GAA TAT GCC CTA GCA CAG ATG TGG CGG TCG TGG GGA ATC    1968

 657  Glu Pro Asp Ile Val Met Gly His Ser Ile Gly Glu Tyr Val Ala Ala     672
1969  GAG CCA GAT ATC GTA ATG GGT CAT AGC ATA GGT GAA TAT GTG GCA GCT    2016

 673  Cys Val Ala Gly Ile Phe Ser Leu Glu Asp Gly Leu Lys Leu Ala Ala     688
2017  TGT GTG GCG GGG ATC TTT TCT CTG GAG GAT GGT CTC AAA CTT GCT GCC    2064

 689  Glu Arg Gly Arg Leu Met Gln Ala Leu Pro Gln Asn Gly Glu Met Val     704
2065  GAA AGA GGC CGT TTG ATG CAG GCG CTA CCA CAA AAT GGC GAG ATG GTT    2112

 705  Ala Ile Ser Ala Ser Leu Glu Glu Val Lys Pro Ala Ile Gln Ser Asp     720
2113  GCT ATA TCG GCC TCC CTT GAG GAA GTT AAG CCG GCT ATT CAA TCT GAC    2160

 721  Gln Arg Val Val Ile Ala Ala Val Asn Gly Pro Arg Ser Val Val Ile     736
2161  CAG CGA GTT GTG ATA GCG GCG GTA AAT GGA CCA CGA AGT GTC GTC ATT    2208

 737  Ser Gly Asp Arg Gln Ala Val Gln Val Phe Thr Asn Thr Leu Glu Asp     752
2209  TCG GGC GAT CGC CAA GCT GTG CAA GTC TTC ACC AAC ACC CTA GAA GAT    2256

 753  Gln Gly Ile Arg Cys Lys Arg Leu Ser Val Ser His Ala Phe His Ser     768
2257  CAA GGA ATC CGG TGC AAG AGA CTG TCT GTT TCA CAC GCT TTC CAC TCT    2304

 769  Pro Leu Met Lys Pro Met Glu Gln Glu Phe Ala Gln Val Ala Arg Glu     784
2305  CCA TTG ATG AAA CCA ATG GAG CAG GAG TTC GCA CAG GTG GCC AGG GAA    2352

 785  Ile Asn Tyr Ser Pro Pro Lys Ile Ala Leu Val Ser Asn Leu Thr Gly     800
2353  ATC AAC TAT AGT CCT CCA AAA ATA GCT CTT GTC AGT AAT CTA ACC GGC    2400

 801  Asp Leu Ile Ser Pro Glu Ser Ser Leu Glu Glu Gly Val Ile Ala Ser     816
2401  GAC TTG ATT TCA CCT GAG TCT TCC CTG GAG GAA GGA GTG ATC GCT TCC    2448

 817  Pro Gly Tyr Trp Val Asn His Leu Cys Asn Pro Val Leu Phe Ala Asp     832
2449  CCT GGT TAC TGG GTA AAT CAT TTA TGC AAT CCT GTC TTG TTC GCT GAT    2496

 833  Gly Ile Ala Thr Met Gln Ala Gln Asp Val Gln Val Phe Leu Glu Val     848
2497  GGT ATT GCA ACT ATG CAA GCG CAG GAT GTC CAA GTC TTC CTT GAA GTT    2544

 849  Gly Pro Lys Pro Thr Leu Ser Gly Leu Val Gln Gln Tyr Phe Asp Glu     864
2545  GGA CCA AAA CCG ACC TTA TCA GGA CTA GTG CAA CAA TAT TTT GAC GAG    2592

 865  Val Ala His Ser Asp Arg Pro Val Thr Ile Pro Thr Leu Arg Pro Lys     880
2593  GTT GCC CAT AGC GAT CGC CCT GTC ACC ATT CCC ACC TTG CGC CCC AAG    2640

 881  Gln Pro Asn Trp Gln Thr Leu Leu Glu Ser Leu Gly Gln Leu Tyr Ala     896
2641  CAA CCC AAC TGG CAG ACA CTA TTG GAG AGT TTG GGA CAA CTG TAT GCG    2688

 897  Leu Gly Val Gln Val Asn Trp Ala Gly Phe Asp Arg Asp Tyr Thr Arg     912
2689  CTT GGT GTC CAG GTA AAT TGG GCG GGC TTT GAT AGA GAT TAC ACC AGA    2736

 913  Arg Lys Val Ser Leu Pro Thr Tyr Ala Trp Lys Arg Gln Arg Tyr Trp     928
2737  CGC AAA GTA AGC CTA CCC ACC TAT GCT TGG AAG CGT CAA CGT TAT TGG    2784

 929  Leu Glu Lys Gln Ser Ala Pro Arg Leu Glu Thr Thr Gln Val Arg Pro     944
2785  CTA GAG AAA CAG TCC GCT CCA CGT TTA GAA ACA ACA CAA GTT CGT CCC    2832

 945  Ala Thr Ala Ile Val Glu His Leu Glu Gln Gly Asn Val Pro Lys Ile     960
2833  GCA ACT GCC ATT GTA GAG CAT CTT GAA CAA GGC AAT GTG CCG AAA ATC    2880

 961  Val Asp Leu Leu Ala Ala Thr Asp Val Leu Ser Gly Glu Ala Arg Lys     976
2881  GTG GAC TTG TTA GCG GCG ACG GAT GTA CTT TCA GGC GAA GCA CGG AAA    2928

 977  Leu Leu Pro Ser Ile Ile Glu Leu Leu Val Ala Lys His Arg Glu Glu     992
2929  TTG CTA CCC AGC ATC ATT GAA CTA TTG GTT GCA AAA CAT CGT GAG GAA    2976

 993  Ala Thr Gln Lys Pro Ile Cys Asp Trp Leu Tyr Glu Val Val Trp Gln    1008
```

```
2977  GCG ACA CAG AAG CCC ATC TGC GAT TGG CTT TAT GAA GTG GTT TGG CAA    3024

1009  Pro Gln Leu Leu Thr Leu Ser Thr Leu Pro Ala Val Glu Thr Glu Gly    1024
3025  CCC CAG TTG CTG ACC CTA TCT ACC TTA CCT GCT GTG GAA ACA GAG GGT    3072

1025  Arg Gln Trp Leu Ile Phe Ala Asp Ala Ser Gly His Gly Glu Ala Leu    1040
3073  AGA CAA TGG CTC ATC TTC GCC GAT GCT AGT GGA CAC GGT GAA GCA CTT    3120

1041  Ala Ala Gln Leu Arg Gln Gln Gly Asp Ile Ile Thr Leu Val Tyr Ala    1056
3121  GCG GCT CAA TTA CGT CAG CAA GGG GAT ATA ATT ACG CTT GTC TAT GCT    3168

1057  Gly Leu Lys Tyr His Ser Ala Asn Asn Lys Gln Asn Thr Gly Gly Asp    1072
3169  GGT CTA AAA TAT CAC TCG GCT AAT AAT AAA CAA AAT ACC GGG GGG GAC    3216

1073  Ile Pro Tyr Phe Gln Ile Asp Pro Ile Gln Arg Glu Asp Tyr Glu Arg    1088
3217  ATC CCA TAT TTT CAG ATT GAT CCG ATC CAA AGG GAG GAT TAT GAA AGG    3264

1089  Leu Phe Ala Ala Leu Pro Pro Leu Tyr Gly Ile Val His Leu Trp Ser    1104
3265  TTG TTT GCT GCT TTG CCT CCA CTG TAT GGT ATT GTT CAT CTT TGG AGT    3312

1105  Leu Asp Ile Leu Ser Leu Asp Lys Val Ser Asn Leu Ile Glu Asn Val    1120
3313  TTA GAT ATA CTT AGC TTG GAC AAA GTA TCT AAC CTA ATT GAA AAT GTA    3360

1121  Gln Leu Gly Ser Gly Thr Leu Leu Asn Leu Ile Gln Thr Val Leu Gln    1136
3361  CAA TTA GGT AGT GGC ACG CTA TTA AAT TTA ATA CAG ACA GTC TTG CAA    3408

1137  Leu Glu Thr Pro Thr Pro Ser Leu Trp Leu Val Thr Lys Asn Ala Gln    1152
3409  CTT GAA ACG CCC ACC CCT AGC TTG TGG CTC GTG ACA AAG AAC GCG CAA    3456

1153  Ala Val Arg Lys Asn Asp Ser Leu Val Gly Val Leu Gln Ser Pro Leu    1168
3457  GCT GTG CGT AAA AAC GAT AGC CTA GTC GGA GTG CTT CAG TCA CCC TTA    3504

1169  Trp Gly Met Gly Lys Val Ile Ala Leu Glu His Pro Glu Leu Asn Cys    1184
3505  TGG GGT ATG GGT AAG GTG ATA GCC TTA GAA CAC CCT GAA CTC AAC TGT    3552

1185  Val Ser Ile Asp Leu Asp Gly Glu Gly Leu Pro Asp Glu Gln Ala Lys    1200
3553  GTA TCA ATC GAC CTT GAT GGT GAA GGG CTT CCA GAT GAA CAA GCC AAG    3600

1201  Phe Leu Ala Ala Glu Leu Arg Ala Ala Ser Glu Phe Arg His Thr Thr    1216
3601  TTT CTG GCG GCT GAA CTC CGC GCC GCC TCC GAG TTC AGA CAT ACC ACC    3648

1217  Ile Pro His Glu Ser Gln Val Ala Trp Arg Asn Arg Thr Arg Tyr Val    1232
3649  ATT CCC CAC GAA AGT CAA GTT GCT TGG CGT AAT AGG ACT CGC TAT GTG    3696

1233  Ser Arg Phe Lys Gly Tyr Gln Lys His Pro Ala Thr Ser Ser Lys Met    1248
3697  TCA CGG TTC AAA GGT TAT CAG AAG CAT CCC GCG ACC TCA TCA AAA ATG    3744

1249  Pro Ile Arg Pro Asp Ala Thr Tyr Leu Ile Thr Gly Gly Phe Gly Gly    1264
3745  CCT ATT CGA CCA GAT GCC ACT TAT TTG ATC ACG GGC GGC TTT GGT GGT    3792

1265  Leu Gly Leu Leu Val Ala Arg Trp Met Val Glu Gln Gly Ala Thr His    1280
3793  TTG GGC TTG CTT GTG GCT CGT TGG ATG GTT GAA CAG GGG GCT ACC CAT    3840

1281  Leu Phe Leu Met Gly Arg Ser Gln Pro Lys Pro Ala Ala Gln Lys Gln    1296
3841  CTA TTT CTG ATG GGA CGC AGC CAA CCC AAA CCA GCC GCC CAA AAA CAA    3888

1297  Leu Gln Glu Ile Ala Ala Leu Gly Ala Thr Val Thr Val Val Gln Ala    1312
3889  CTG CAA GAG ATA GCC GCG CTG GGT GCA ACA GTG ACG GTG GTG CAA GCC    3936

1313  Asp Val Gly Ile Arg Ser Gln Val Ala Asn Val Leu Ala Gln Ile Asp    1328
3937  GAT GTT GGC ATC CGC TCC CAA GTA GCC AAT GTG TTG GCA CAG ATT GAT    3984

1329  Lys Ala Tyr Pro Leu Ala Gly Ile Ile His Thr Ala Gly Val Leu Asp    1344
3985  AAG GCA TAT CCT TTG GCT GGT ATT ATT CAT ACT GCC GGT GTA TTA GAC    4032

1345  Asp Gly Ile Leu Leu Gln Gln Asn Trp Ala Arg Phe Ser Lys Val Phe    1360
4033  GAC GGA ATC TTA TTG CAG CAA AAT TGG GCG CGT TTT AGC AAG GTG TTC    4080
```

```
1361  Ala Pro Lys Leu Glu Gly Ala Trp His Leu His Thr Leu Thr Glu Glu    1376
4081  GCC CCC AAA CTA GAG GGA GCT TGG CAT CTA CAT ACA CTG ACT GAA GAG    4128

1377  Met Pro Leu Asp Phe Phe Ile Cys Phe Ser Ser Thr Ala Gly Leu Leu    1392
4129  ATG CCG CTT GAT TTC TTT ATT TGT TTT TCC TCA ACA GCA GGA TTG CTG    4176

1393  Gly Ser Gly Gly Gln Ala Asn Tyr Ala Ala Ala Asn Ala Phe Leu Asp    1408
4177  GGC AGT GGT GGA CAA GCT AAC TAT GCT GCT GCC AAT GCC TTT TTA GAT    4224

1409  Ala Phe Ala His His Arg Arg Ile Gln Gly Leu Pro Ala Leu Ser Ile    1424
4225  GCC TTT GCC CAT CAT CGG CGA ATA CAA GGC TTG CCA GCT CTC TCG ATT    4272

1425  Asn Trp Asp Ala Trp Ser Gln Val Gly Met Thr Val Arg Leu Gln Gln    1440
4273  AAC TGG GAC GCT TGG TCT CAA GTG GGA ATG ACG GTA CGT CTC CAA CAA    4320

1441  Ala Ser Ser Gln Ser Thr Thr Val Gly Gln Asp Ile Ser Thr Leu Glu    1456
4321  GCT TCT TCA CAA AGC ACC ACA GTT GGG CAA GAT ATT AGC ACT TTG GAA    4368

1457  Ile Ser Pro Glu Gln Gly Leu Gln Ile Phe Ala Tyr Leu Leu Gln Gln    1472
4369  ATT TCA CCA GAA CAG GGA TTG CAA ATC TTT GCC TAT CTT CTG CAA CAA    4416

1473  Pro Ser Ala Gln Ile Ala Ala Ile Ser Thr Asp Gly Leu Arg Lys Met    1488
4417  CCA TCC GCC CAA ATA GCG GCC ATT TCT ACC GAT GGG CTT CGC AAG ATG    4464

1489  Tyr Asp Thr Ser Ser Ala Phe Phe Ala Leu Leu Asp Leu Asp Arg Ser    1504
4465  TAC GAC ACA AGC TCG GCC TTT TTT GCT TTA CTT GAT CTT GAC AGG TCT    4512

1505  Ser Ser Thr Thr Gln Glu Gln Ser Thr Leu Ser His Glu Val Gly Leu    1520
4513  TCC TCC ACT ACC CAG GAG CAA TCT ACA CTT TCT CAT GAA GTT GGC CTT    4560

1521  Thr Leu Leu Glu Gln Leu Gln Gln Ala Arg Pro Lys Glu Arg Glu Lys    1536
4561  ACC TTA CTC GAA CAA TTG CAG CAA GCT CGG CCA AAA GAG CGA GAG AAA    4608

1537  Met Leu Leu Arg His Leu Gln Thr Gln Val Ala Ala Val Leu Arg Ser    1552
4609  ATG TTA CTG CGC CAT CTA CAG ACC CAA GTT GCT GCG GTC TTG CGT AGT    4656

1553  Pro Glu Leu Pro Ala Val His Gln Pro Phe Thr Asp Leu Gly Met Asp    1568
4657  CCC GAA CTG CCC GCA GTT CAT CAA CCC TTC ACT GAC TTG GGG ATG GAT    4704

1569  Ser Leu Met Ser Leu Glu Leu Met Arg Arg Leu Glu Glu Ser Leu Gly    1584
4705  TCG TTG ATG TCA CTT GAA TTG ATG CGG CGT TTG GAA GAA AGT CTG GGG    4752

1585  Ile Gln Met Pro Ala Thr Leu Ala Phe Asp Tyr Pro Met Val Asp Arg    1600
4753  ATT CAG ATG CCT GCA ACG CTT GCA TTC GAT TAT CCT ATG GTA GAC CGT    4800

1601  Leu Ala Lys Phe Ile Leu Thr Gln Ile Cys Ile Asn Ser Glu Pro Asp    1616
4801  TTG GCT AAG TTT ATA CTG ACT CAA ATA TGT ATA AAT TCT GAG CCA GAT    4848

1617  Thr Ser Ala Val Leu Thr Pro Asp Gly Asn Gly Glu Glu Lys Asp Ser    1632
4849  ACC TCA GCA GTT CTC ACA CCA GAT GGA AAT GGG GAG GAA AAA GAC AGT    4896

1633  Asn Lys Asp Arg Ser Thr Ser Thr Ser Val Asp Ser Asn Ile Thr Ser    1648
4897  AAT AAG GAC AGA AGT ACC AGC ACT TCC GTT GAC TCA AAT ATT ACT TCC    4944

1649  Met Ala Glu Asp Leu Phe Ala Leu Glu Ser Leu Leu Asn Lys Ile Lys    1664
4945  ATG GCA GAA GAT TTA TTC GCA CTC GAA TCC TTA CTA AAT AAA ATA AAA    4992

1665  Arg Asp Gln ***                                                   1668
4993  AGA GAT CAA TAA                                                   5004
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : ORF2.SEQ       SEQUENCE : NORMAL    318 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -    318

    TRANSLATION REGION :      1 -    318


*** DNA TRANSLATION ***

   1  Met Leu Asn Leu Asp Arg Ile Leu Asn Gln Glu Arg Leu Leu Arg Ala      16
   1  ATG TTG AAC TTA GAC CGC ATC CTG AAT CAA GAG CGA CTG CTA CGA GCA      48

  17  Met Thr Gly Leu Asn Arg Gln Ala Phe Asn Glu Leu Leu Ser Gln Phe      32
  49  ATG ACT GGA CTT AAC CGC CAA GCA TTC AAC GAG CTG TTA TCT CAG TTT      96

  33  Ala Asp Thr Tyr Glu Arg Thr Val Phe Asn Ser Leu Ala Asn Arg Lys      48
  97  GCT GAT ACC TAT GAA CGC ACC GTG TTC AAC TCC TTA GCA AAC CGC AAA     144

  49  Arg Ala Pro Gly Gly Gly Arg Lys Pro Thr Leu Arg Ser Ile Glu Glu      64
 145  CGT GCG CCC GGG GGC GGA CGC AAG CCT ACA CTC AGA AGT ATA GAG GAA     192

  65  Lys Leu Phe Tyr Ile Leu Leu Tyr Cys Lys Phe Tyr Pro Thr Tyr Arg      80
 193  AAA CTA TTT TAT ATC CTG CTG TAC TGC AAA TTT TAT CCC ACA TAT AGG     240

  81  Asn Arg Val Thr Asp Phe Asp Asp Gln Leu Met Leu Val Ser Ala Gly      96
 241  AAT CGT GTG ACC GAT TTT GAT GAT CAA TTG ATG TTG GTC AGT GCA GGG     288

  97  Leu Trp Asn Phe Tyr Leu Asp Ala Ala ***                          106
 289  CTA TGG AAC TTC TAT TTA GAT GCA GCA TAA                          318
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : CYNG.SEQ       SEQUENCE : NORMAL    600 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -    600

    TRANSLATION REGION :      1 -    600


*** DNA TRANSLATION ***

   1  Met Lys Ser Ser Asn Ile Gln His Thr Lys His Leu Val Ser Pro Phe      16
   1  ATG AAA TCA AGC AAT ATT CAA CAT ACA AAA CAT CTT GTG TCC CCA TTT      48

  17  Asp Arg Cys Leu Lys Asn Gly His Arg Asn Gly Ile Leu Trp Phe Thr      32
  49  GAT CGT TGT TTA AAG AAT GGT CAC CGA AAT GGA ATT CTA TGG TTC ACT      96

  33  Gly Leu Ser Gly Ser Gly Lys Thr Thr Leu Ala Leu Glu Leu Glu Arg      48
  97  GGA TTA TCA GGA TCA GGT AAA ACA ACA TTG GCT CTA GAA TTA GAG CGG     144

  49  Ile Leu Phe Gln Lys Gly Trp Tyr Thr Phe Val Leu Asp Gly Asp Asn      64
 145  ATT TTG TTC CAG AAA GGT TGG TAT ACT TTT GTC CTT GAT GGT GAC AAT     192

  65  Ile Arg His Gly Leu Cys Ala Asp Leu Gly Phe Ser Ala Ala Asp Arg      80
 193  ATC CGT CAT GGA TTA TGT GCT GAC TTG GGA TTT TCT GCT GCT GAT CGT     240

  81  Ala Glu Asn Ile Arg Arg Ile Gly Glu Val Ala Lys Leu Phe Ala Asp      96
 241  GCG GAA AAT ATT CGT CGT ATT GGC GAG GTG GCC AAA CTT TTT GCG GAT     288

  97  Ser Gly Cys Leu Val Ile Thr Ala Phe Ile Ser Pro Tyr Arg Lys Asp     112
 289  TCG GGT TGT TTA GTG ATT ACT GCT TTC ATT TCA CCT TAC AGG AAA GAC     336

 113  Arg Glu Gln Ile Arg Ile Leu Ser Gly Asp Leu Phe His Glu Val Tyr     128
 337  CGG GAG CAA ATC CGT ATT CTG TCC GGA GAT CTG TTT CAT GAA GTA TAT     384

 129  Ile Ala Thr Pro Met Glu Ile Cys Glu Gln Arg Asp Pro Lys Gly Leu     144
 385  ATT GCA ACT CCG ATG GAA ATT TGT GAA CAA CGT GAC CCC AAG GGT CTC     432

 145  Tyr Ala Lys Ala Arg Arg Gly Glu Ile Glu Ser Phe Thr Gly Ile Ser     160
 433  TAT GCA AAA GCA CGC CGT GGA GAA ATT GAG TCT TTT ACG GGA ATT AGC     480

 161  Ala Pro Tyr Glu Pro Pro Arg Thr Pro Asp Leu Leu Val Glu Thr Ser     176
 481  GCT CCC TAC GAA CCA CCG AGA ACA CCA GAT TTA CTG GTG GAA ACT TCG     528

 177  Glu Leu Ser Ile Glu Ala Ser Leu Gln Gln Leu Leu Glu Tyr Val Gly     192
 529  GAA CTT TCC ATC GAG GCC AGC TTA CAA CAG CTA CTC GAA TAC GTG GGA     576

 193  Lys Lys Phe Thr Phe Thr Gln ***                                     200
 577  AAG AAG TTC ACT TTC ACT CAG TAG                                     600
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : ORF4.SEQ       SEQUENCE : NORMAL   1371 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -   1371

    TRANSLATION REGION :      1 -   1371


*** DNA TRANSLATION ***

   1  Leu Ile Val Glu Lys Tyr Thr Ala Val Ile Leu Ser Val Glu Val Gly      16
   1  CTG ATA GTG GAA AAA TAT ACC GCA GTT ATT CTT TCC GTT GAA GTG GGA      48

  17  Ser Val Phe Ser His Asn Pro Gln Gln Ala Glu Gln Val Ile Leu Glu      32
  49  TCA GTA TTC TCC CAT AAT CCC CAG CAA GCG GAG CAG GTT ATT TTA GAA      96

  33  Glu Trp Lys Gln Ala Ile Lys Phe Tyr Leu Pro Lys Glu Leu Gln Pro      48
  97  GAA TGG AAA CAG GCA ATC AAA TTT TAT TTA CCC AAA GAA CTA CAA CCA     144

  49  Ser Tyr Trp Pro Glu Arg Glu Thr Asp Ser Gly Ile Gly Lys Thr Leu      64
 145  TCC TAT TGG CCA GAG CGT GAA ACA GAC TCA GGA ATA GGC AAA ACT TTA     192

  65  Ser Glu Trp Ser Ala Gln Ser Pro Arg Pro Leu Val Ile Phe Leu Asp      80
 193  AGT GAG TGG TCC GCA CAA TCT CCA AGA CCT CTT GTA ATC TTT TTA GAT     240

  81  Glu Ile Asp Ser Leu Thr Asp Glu Ala Leu Ile Leu Ile Leu Arg Gln      96
 241  GAA ATC GAT TCC CTA ACA GAT GAA GCT TTA ATC CTA ATT TTA AGA CAA     288

  97  Leu Arg Ser Gly Phe Pro Arg Arg Pro Arg Gly Phe Pro His Ser Val     112
 289  TTA CGC TCA GGT TTT CCC CGT CGT CCT CGG GGA TTT CCC CAT TCG GTG     336

 113  Gly Leu Ile Gly Met Arg Asp Val Arg Asp Tyr Lys Val Lys Ser Gly     128
 337  GGG TTA ATT GGT ATG CGG GAT GTG CGG GAC TAT AAG GTT AAA TCT GGT     384

 129  Gly Ser Glu Arg Leu Asn Thr Ser Ser Pro Phe Asn Ile Lys Ala Glu     144
 385  GGA AGT GAA CGA CTG AAT ACG TCA AGT CCT TTC AAT ATC AAA GCG GAA     432

 145  Ser Leu Thr Leu Ser Asn Phe Thr Leu Ser Glu Val Glu Glu Leu Tyr     160
 433  TCC TTG ACT TTA AGT AAT TTC ACT CTG TCA GAG GTG GAA GAA CTT TAC     480

 161  Leu Gln His Thr Gln Ala Thr Gly Gln Ile Phe Thr Pro Glu Ala Ile     176
 481  TTA CAA CAT ACG CAA GCT ACA GGA CAA ATT TTT ACC CCG GAA GCA ATT     528

 177  Lys Gln Ala Phe Tyr Leu Thr Asp Gly Gln Pro Trp Leu Val Asn Ala     192
 529  AAA CAA GCA TTT TAT TTA ACC GAT GGG CAA CCA TGG TTA GTA AAC GCC     576

 193  Leu Ala Arg Gln Ala Thr Gln Val Leu Val Lys Asp Ile Thr Gln Pro     208
 577  CTA GCT CGT CAA GCC ACT CAG GTG TTA GTG AAA GAT ATT ACT CAA CCC     624

 209  Ile Thr Ala Glu Val Ile Asn Gln Ala Lys Glu Val Leu Ile Gln Arg     224
 625  ATT ACC GCT GAA GTA ATT AAC CAA GCC AAA GAA GTT CTG ATT CAG CGC     672

 225  Gln Asp Thr His Leu Asp Ser Leu Ala Glu Arg Leu Arg Glu Asp Arg     240
 673  CAG GAT ACC CAT TTG GAT AGT TTG GCA GAG CGC TTA CGG GAA GAT CGG     720

 241  Val Lys Ala Ile Ile Gln Pro Met Leu Ala Gly Ser Asp Leu Pro Asp     256
 721  GTC AAA GCC ATT ATT CAA CCT ATG TTA GCT GGA TCG GAC TTA CCA GAT     768

 257  Thr Pro Glu Asp Asp Arg Arg Phe Leu Leu Asp Leu Gly Leu Val Lys     272
 769  ACC CCA GAG GAT GAT CGC CGT TTC TTG CTA GAT TTA GGC TTG GTA AAG     816
```

```
 273  Arg Ser Pro Leu Gly Gly Leu Thr Ile Ala Asn Pro Ile Tyr Gln Glu     288
 817  CGC AGT CCC TTG GGA GGA CTA ACC ATT GCC AAT CCC ATT TAC CAG GAG     864

 289  Val Ile Pro Arg Val Leu Ser Gln Gly Ser Gln Asp Ser Leu Pro Gln     304
 865  GTG ATT CCT CGT GTT TTG TCC CAG GGT AGT CAG GAT AGT CTA CCC CAG     912

 305  Ile Gln Pro Thr Trp Leu Asn Thr Asp Asn Thr Leu Asn Pro Asp Lys     320
 913  ATT CAA CCT ACT TGG TTA AAT ACT GAT AAT ACT TTA AAT CCT GAC AAA     960

 321  Leu Leu Asn Ala Phe Leu Glu Phe Trp Arg Gln His Gly Glu Pro Leu     336
 961  CTC TTA AAT GCT TTC CTA GAG TTT TGG CGA CAA CAT GGG GAA CCA TTA    1008

 337  Leu Lys Ser Ala Pro Tyr His Glu Ile Ala Pro His Leu Val Leu Met     352
1009  CTC AAA AGT GCG CCT TAT CAT GAA ATT GCT CCC CAT TTA GTT TTG ATG    1056

 353  Ala Phe Leu His Arg Val Val Asn Gly Gly Gly Thr Leu Glu Arg Glu     368
1057  GCG TTT TTA CAT CGG GTA GTG AAT GGT GGT GGC ACT TTA GAA CGG GAA    1104

 369  Tyr Ala Val Gly Ser Gly Arg Met Asp Ile Cys Leu Arg Tyr Gly Lys     384
1105  TAT GCC GTT GGT TCT GGA AGA ATG GAT ATT TGT TTA CGC TAT GGC AAG    1152

 385  Val Val Met Gly Ile Glu Leu Lys Val Trp Gly Gly Lys Ser Asp Pro     400
1153  GTA GTG ATG GGC ATA GAG TTA AAG GTT TGG GGG GGA AAA TCG GAT CCG    1200

 401  Leu Thr Lys Gly Leu Thr Gln Leu Asp Lys Tyr Leu Gly Gly Leu Gly     416
1201  TTA ACG AAG GGT TTG ACC CAA TTG GAT AAA TAT CTG GGT GGG TTA GGA    1248

 417  Leu Asp Arg Gly Trp Leu Val Ile Phe Asp His Arg Pro Gly Leu Pro     432
1249  TTA GAT AGA GGT TGG TTA GTA ATT TTT GAT CAC CGT CCG GGA TTA CCA    1296

 433  Pro Met Gly Glu Arg Ile Ser Met Glu Gln Ala Ile Ser Pro Glu Gly     448
1297  CCC ATG GGT GAG AGG ATT AGT ATG GAA CAG GCC ATT AGT CCA GAG GGA    1344

 449  Arg Thr Ile Thr Val Ile Arg Ser ***                              457
1345  AGA ACC ATT ACA GTG ATT CGT AGC TAG                              1371
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : HYPC.SEQ       SEQUENCE : NORMAL    279 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -    279

    TRANSLATION REGION :      1 -    279


*** DNA TRANSLATION ***

   1  Met Cys Leu Gly Ile Pro Gly Gln Ile Ile Glu Ile Ser Asn Pro Glu      16
   1  ATG TGT TTA GGA ATA CCT GGA CAA ATA ATA GAA ATC AGC AAT CCA GAA      48

  17  His Lys Leu Ala Ile Val Lys Val Gly Gly Val Lys Arg Gln Ile Asn      32
  49  CAT AAA CTA GCA ATA GTT AAG GTT GGA GGA GTT AAG CGT CAA ATC AAT      96

  33  Ile Ala Cys Ile Val Asp Glu Glu His Pro Pro Glu Lys Cys Leu Gly      48
  97  ATA GCA TGT ATT GTT GAT GAG GAA CAC CCA CCC GAA AAA TGT CTG GGA     144

  49  Asp Trp Val Leu Val His Val Gly Phe Ala Met Asn Arg Ile Asn Glu      64
 145  GAT TGG GTG TTA GTT CAT GTC GGT TTT GCT ATG AAC AGA ATT AAT GAA     192

  65  Gln Glu Ala Ala Glu Thr Leu Ala Leu Leu Gln Glu Ile Ala Asn Ser      80
 193  CAG GAA GCA GCA GAA ACC TTA GCC CTT TTA CAG GAA ATT GCT AAC AGT     240

  81  Tyr Asp Asp Thr Lys Asp Leu Gln Arg Ser Asn Leu ***                  93
 241  TAT GAT GAT ACC AAA GAC CTG CAA AGG AGC AAT TTA TAA                 279
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : HYPD.SEQ       SEQUENCE : NORMAL   1128 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -   1128

    TRANSLATION REGION :      1 -   1128


*** DNA TRANSLATION ***

   1  Met Lys Tyr Val Asp Glu Phe Arg Asn Pro Glu Lys Ala Gln Gly Leu      16
   1  ATG AAA TAT GTA GAT GAA TTC CGT AAC CCC GAA AAA GCT CAA GGT TTA      48

  17  Gln Lys Glu Ile Ala Gln Leu Ser Leu Gln Ile Ser Arg Asn Ser His      32
  49  CAA AAA GAA ATT GCA CAA CTG AGT CTT CAA ATC TCA AGA AAT TCC CAC      96

  33  Lys Asn Lys His Leu Lys Ile Met Glu Val Cys Gly Gly His Thr His      48
  97  AAA AAC AAA CAC CTC AAA ATA ATG GAA GTA TGT GGT GGA CAC ACC CAT     144

  49  Ala Ile Phe Lys Tyr Gly Ile Glu Glu Ile Leu Pro Asp Asn Ile Glu      64
 145  GCG ATT TTT AAA TAT GGC ATT GAG GAA ATA TTA CCA GAT AAC ATT GAA     192

  65  Leu Ile His Gly Pro Gly Cys Pro Val Cys Val Met Pro Lys Gly Arg      80
 193  CTG ATT CAT GGA CCT GGA TGT CCG GTA TGT GTC ATG CCA AAA GGT AGA     240

  81  Val Asp Asp Ala Ile Ala Leu Cys Gln Asn Gln Gln Ile Ile Phe Thr      96
 241  GTA GAT GAT GCG ATC GCT CTT TGC CAA AAC CAG CAG ATT ATT TTC ACC     288

  97  Thr Phe Gly Asp Ala Met Arg Val Pro Gly Ser Lys Thr Ser Leu Leu     112
 289  ACC TTT GGG GAT GCT ATG AGG GTT CCA GGG TCT AAA ACC AGC TTG TTA     336

 113  Gln Ala Lys Ala Glu Gly Ala Asp Ile Arg Met Val Tyr Ser Pro Leu     128
 337  CAA GCA AAA GCT GAG GGA GCA GAT ATT AGA ATG GTC TAT TCT CCT CTG     384

 129  Asp Ser Leu Lys Ile Ala Lys Glu Asn Pro Glu Lys Glu Ile Val Phe     144
 385  GAT AGT TTA AAG ATT GCC AAG GAA AAC CCT GAA AAA GAG ATA GTA TTT     432

 145  Phe Gly Leu Gly Phe Glu Thr Thr Ala Pro Ser Thr Ala Phe Thr Ile     160
 433  TTT GGT TTA GGA TTT GAA ACC ACT GCT CCT AGC ACT GCA TTT ACC ATA     480

 161  Leu Gln Ala Ala Ala Glu Asn Ile Arg Asn Phe Ser Leu Phe Ser Asn     176
 481  TTA CAA GCA GCA GCA GAA AAC ATT AGA AAC TTT AGT TTA TTT TCC AAT     528

 177  His Val Leu Val Ile Pro Ala Leu Glu Ala Leu Leu Ala Asn Pro Asp     192
 529  CAT GTA TTA GTT ATA CCA GCA TTG GAA GCA TTA TTA GCA AAT CCA GAT     576

 193  Leu Gln Leu Asp Gly Phe Val Gly Pro Gly His Val Ser Met Val Ile     208
 577  TTA CAA CTA GAT GGT TTT GTT GGT CCA GGT CAT GTA AGC ATG GTT ATT     624

 209  Gly Thr Asp Pro Tyr Glu Phe Ile Pro Glu Arg Tyr His Lys Pro Ile     224
 625  GGA ACA GAT CCA TAT GAG TTT ATT CCC GAA AGA TAT CAT AAA CCG ATT     672

 225  Val Ile Ser Gly Phe Glu Pro Leu Asp Ile Phe Gln Ser Ile Trp Met     240
 673  GTC ATT TCT GGA TTT GAA CCA TTG GAT ATT TTC CAA TCT ATA TGG ATG     720

 241  Leu Leu Lys Gln Ile Glu Glu Asn Arg Cys Glu Val Glu Asn Gln Tyr     256
 721  TTA TTA AAA CAA ATA GAA GAA AAT CGG TGT GAA GTA GAA AAC CAA TAT     768

 257  Asn Arg Leu Val Glu Lys Gly Gly Asn Gln Asn Ala Ile Gln Ala Met     272
 769  AAT CGG TTA GTT GAG AAA GGG GGA AAT CAA AAT GCC ATT CAA GCC ATG     816
```

```
 273  Asn Gln Val Phe Thr Val Arg Glu Lys Phe Glu Trp Arg Gly Leu Gly     288
 817  AAT CAG GTT TTT ACA GTT CGC GAG AAA TTT GAA TGG CGG GGG TTG GGA     864

 289  Glu Ile Pro Asn Ser Gly Leu Lys Ile Arg Pro Glu Tyr Ala Gln Phe     304
 865  GAA ATT CCT AAT TCT GGT TTG AAA ATA CGT CCC GAA TAT GCT CAA TTT     912

 305  Asp Ala Glu Ala Lys Phe Thr Ile Pro Asn Leu Lys Val Pro Asp His     320
 913  GAT GCA GAA GCC AAA TTT ACT ATT CCC AAT TTG AAA GTG CCC GAT CAT     960

 321  Lys Ala Cys Gln Cys Gly Glu Ile Leu Lys Gly Val Leu Lys Pro Trp     336
 961  AAA GCC TGT CAA TGT GGA GAA ATT CTT AAA GGC GTT TTA AAA CCT TGG    1008

 337  Glu Cys Lys Val Phe Gly Thr Ala Cys Thr Pro Glu Thr Pro Ile Gly     352
1009  GAA TGT AAA GTA TTT GGC ACA GCT TGC ACA CCA GAA ACA CCC ATA GGA    1056

 353  Thr Cys Met Val Ser Tyr Glu Gly Ala Cys Ala Ala Tyr Tyr Lys Tyr     368
1057  ACT TGT ATG GTT TCC TAT GAA GGT GCT TGT GCA GCA TAT TAT AAA TAT    1104

 369  Gly Arg Leu Ser Ala Leu Ile ***                                     376
1105  GGC AGA TTA TCA GCT CTT ATA TAG                                    1128
```

```
*** INPUT INFORMATION ***

    FILE  NAME  : HYPE.SEQ       SEQUENCE : NORMAL    747 BP

    CODON TABLE : UNIV.TCN

    SEQUENCE    REGION :      1 -    747

    TRANSLATION REGION :      1 -    747


*** DNA TRANSLATION ***

   1  Met Lys Val Leu Val Gln His Ile Ile Asn Met Ala Asp Tyr Gln Leu      16
   1  ATG AAG GTG CTT GTG CAG CAT ATT ATA AAT ATG GCA GAT TAT CAG CTC      48

  17  Leu Tyr Ser Asn Cys Met Ala Ile Asp Lys Asn Gln Lys Asn Gln Asn      32
  49  TTA TAT AGC AAT TGT ATG GCA ATA GAT AAA AAT CAG AAA AAT CAA AAT      96

  33  Pro Leu Phe Gln Lys Ile Ala Gln Val Gly Arg Arg Ser His Gln Val      48
  97  CCC CTA TTT CAA AAA ATA GCA CAG GTA GGT CGT CGC TCC CAT CAA GTT     144

  49  Lys Asp Ser His Ile Asn Leu Ser His Gly Ser Gly Gly Lys Ala Met      64
 145  AAA GAT AGC CAT ATT AAT CTA TCC CAT GGT AGT GGT GGT AAA GCT ATG     192

  65  Ser Asp Leu Ile His Asp Val Phe Val Lys Asn Phe His Asn Glu Ile      80
 193  AGC GAT TTA ATT CAT GAT GTA TTT GTG AAA AAT TTC CAT AAT GAG ATT     240

  81  Leu Ser Gln Leu Glu Asp Gln Ala Ser Phe Asp Leu Ser Thr Leu Gly      96
 241  TTA TCA CAA TTA GAA GAT CAG GCA AGT TTT GAT TTA TCA ACC CTG GGA     288

  97  Lys Leu Gly Asp Arg Leu Ala Phe Thr Thr Asp Ser Tyr Val Val Asp     112
 289  AAA CTG GGA GAT AGA TTA GCA TTT ACC ACC GAT TCT TAC GTG GTA GAT     336

 113  Pro Leu Phe Phe Pro Gly Gly Asn Ile Gly Asp Leu Ala Val Asn Gly     128
 337  CCC TTG TTT TTT CCC GGT GGC AAT ATT GGT GAT TTA GCT GTT AAT GGC     384

 129  Thr Ile Asn Asp Leu Ala Val Gly Gly Ala Lys Pro Leu Tyr Leu Ser     144
 385  ACA ATT AAT GAT CTG GCA GTT GGG GGT GCT AAA CCC CTA TAT TTA AGT     432

 145  Cys Ser Met Ile Leu Glu Glu Gly Leu Pro Leu Glu Thr Leu Arg Lys     160
 433  TGT AGC ATG ATC TTA GAA GAG GGA TTG CCA TTA GAA ACA TTA AGA AAA     480

 161  Val Val Gly Ser Met Gln His Ala Ala Gln Lys Ala Gly Val Gln Ile     176
 481  GTA GTG GGG AGT ATG CAA CAT GCT GCT CAA AAA GCG GGA GTG CAA ATA     528

 177  Val Thr Gly Asp Thr Lys Val Val His Arg Gly Cys Ala Asp Lys Leu     192
 529  GTT ACA GGA GAC ACC AAG GTA GTG CAT CGT GGT TGT GCG GAT AAA TTA     576

 193  Phe Ile Asn Thr Ala Gly Ile Gly Ile Ile Pro Gln Gly Ile Lys Ile     208
 577  TTT ATT AAC ACT GCT GGG ATA GGG ATA ATA CCC CAG GGA ATT AAA ATT     624

 209  Ser Pro Arg His Ile Gln Pro Gly Asp Met Val Ile Ile Asn Gly Glu     224
 625  TCA CCT CGC CAT ATT CAA CCG GGG GAT ATG GTA ATT ATT AAT GGA GAA     672

 225  Ile Gly Asn His Gly Ala Ala Ile Leu Ile Ala Arg Gly Glu Leu Glu     240
 673  ATA GGG AAT CAT GGT GCG GCA ATT TTA ATA GCT AGA GGA GAA TTA GAA     720

 241  Leu Glu Thr Asp Ile Glu Ser Asp Ser                              249
 721  TTA GAA ACA GAT ATA GAA AGC GAC AGT                              747
```

Appendix 2 Detailed results of gene-directed PCR toward the CYN biosynthetic cluster

GENUS	SPECIES	ID CODE	*cynA* (AT)	*cynB** (KS)	*cynB* (AT)	*cynD* (AT)	*cynC*	*cynF*	*cynL*	*orf4*	*orf3*	*cynE*	*cynG* (ASK)	*cynH* (P-OH)	*cynJ* (AT)	*cynK*(AT)	CYN Toxicity	Comments
Anabaena	*aphanizomeniodes*	ANA259C	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	HPLC/MS/PI
	aphanizomeniodes	ANA262A	-		-	-	-	-	-	-	-	-	-	-	-		-	HPLC/MS
	bergii	ANA283A(old)	+	+	+	+	+	+			-	+	-	-	+		+	Fergusson (2003)
	bergii	ANA283A(new)	-		-	-	-	-	-	-	-	-	-	-	-	-	-	HPLC/MS
	bergii	ANA283B	-		-	-	-	-	-	-	-	-	-	-	-		-	HPLC/MS
	bergii	ANA360H	-	-	-	-	-	-	-	-	-	-	-		-		-	Fergusson (2003)
	bergii	ANA366A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Fergusson (2003)
	bergii	ANA366B	+	+	+	+	+	+	-	-	-	+	-	-	+	+	+	Fergusson (2003)
	bergii	ANA372E	-		-	-	-	-	-	-	-	-	-	-	-		-	HPLC/MS
	bergii	ANA372J	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	HPLC/MS
	*lapponica***	966	-	-	-***	-	-***	-	-	-	-	+	-	+	-		+	3 x HPLC analysis
Aphanizomenon	*issatschenkoi*	APH016B	-	-	-	-	-	-	-	-	-	-	-	-	-		-	HPLC/MS
	ovalisporum	APH031E	+	+	+	+	+	+	-	-	-	+	-	-	+		Not tested	
	ovalisporum	APH031G	+	+	+	+	+	+	-	-	-	+	-	-	+	+	+	Fergusson (2003)
	ovalisporum	APHO33B	+		+	+	+	+	-	-	-	+	-	-	+	+	+	PI
	ovalisporum	APH033C	+	+	+	+	+	+	-	-	-	+	-	-	+		+	Fergusson (2003)
	ovalisporum	APH033E	+		+	+	+	+	-	-	-	+	-	-	+		Not tested	
	ovalisporum	APH035C	+	+	+	+	+	+	-	-	-	+	-	-	+	+	+	PI
	ovalisporum	APH035D	+		+	+	+	+	-	-	-	+	-	-	+		+	PI
	ovalisporum	APH035E	+		+	+	+	+	-	-	-	+	-	-	+		Not tested	
	ovalisporum	APH035F	+	+	+	+	+	+	-	-	-	+	-	-	+		Not tested	
	issatschenkoi	CAWBG02	+	+	+	+	+	+	+	-	+		+		+		-	Wood et al 2007b
	sp †	10E6	+	§	+	+	+	+				+			+	+	+	Pressul et al 2006
	ovalisporum	ACO284	+	+	+	+	+	+	-	-	-	+	-	-	+	+	+	HPLC/MS
Cylindrospermopsis	*raciborskii*	CYP005F	+		+	+	+	+	+	+	+	+	+	+	+		Not tested	
	raciborskii	CYP014A	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-	Fergusson (2003)
	raciborskii	CYP018A	-	-	-	+	-	-	-	+	+	-	-	-	-	-	-	HPLC/MS
	raciborskii	CYLI-19	-	-	-	-	-	-	-	-	+	-	+	-	-		-	Fergusson (2003)
	raciborskii	CYP020A	+	+	+	+	+	+	+	+	+	+	+	+	+		+	Fergusson (2003)
	raciborskii	CYP020B	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Fergusson (2003)
	raciborskii	CYP023A	+	+	+	+	+	+	+	+	+	+	+	+	+		+	Fergusson (2003)
	raciborskii	CYP023E	+	+	+	+	+	+	+	+	+	+	+	+	+		+	Fergusson (2003)
	raciborskii	CYP024C	+	+	+	+	+	+	+	+	+	+	+	+	+		+	Fergusson (2003)
	raciborskii	CYP025B	+	+	+	+	+	+	+	+	+	+	+	+	+		+	Fergusson (2003)
	raciborskii	CYP026J	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Fergusson (2003)
	raciborskii	CYP028A	+		+	+	+	+	+	+	+	+	+	+	+	+	+	HPLC/MS
	raciborskii	CYP029	-		-	-	+	-	-	+	+	-	+	-	-		-	Fergusson (2003)
	raciborskii	CYP030A	+		+	+	+	+	+	+	+	+	+	+	+	+	+	HPLC/MS
	raciborskii	CYLI-31	-	-	-	-	-	-	-	+	+	-	+	-	-		-	Fergusson (2003)
	raciborskii	CYP032	+	+	+	+	+	+	+	+	+	+	+	+	+		+	HPLC/MS
	raciborskii	CYP033	+		+	+	+	+	+	+	+	+	+	+	+		Not tested	
	raciborskii	CYP034	+		+	+	+	+	+	+	+	+	+	+	+		Not tested	
	raciborskii	CYP040	+		-	-	+	-	-	-	-	-	+	-	-		-	Fergusson (2003)
	raciborskii	CYP041	-		-	+	+	+	-	-	+	-	-	-	-		-	HPLC/MS/saxitoxin +ve
	raciborskii	CYP042	-		-	-	-	+	-	+	+	-	-	-	-		-	HPLC/MS/saxitoxin+ve
	raciborskii	CYN43	+		+	+	+	+	+	+	+	+	+	+	+	+	+	Fergusson (2003)
	raciborskii	CYL1-53	-	-	-	-	+	-	-	+	+	-	+	-	-		-	Fergusson (2003)
	raciborskii	CYLI-75	-	-	+	-	-	+	-	+	+	+	+	-	-		-	Fergusson (2003)
Lyngbya	*wollei*			-	+**							+					+	per Barbara Sendall

* Using primers K18/M4 (Fergusson et al 2003)

** + amplification products for cynB using degenerate (inhouse) primers designed and performed by Barbara Sendall QHSS

*** + amplification products using primers AL1/AL2 for cynB (this work) and for cynC (Kellman et al 2006)

† Sequences confirmed by Anke Stüken/ cynE primer seq demonstarted 100% homology to *Aph. sp.* 10E6 sequence.

§ Preussel et al. 2006

PCR not perfomed

Appendix 3 Distribution of Chi square Probability

	Probability			
Degrees of freedom	0.10	0.05	0.01	0.001
1	2.7	3.8	6.6	10.8
2	4.6	6.0	9.2	13.8
3	6.3	7.8	11.3	16.3
4	7.8	9.5	13.3	18.5
5	9.2	11.1	15.1	20.5
6	10.6	12.6	16.8	22.5
7	12.0	14.1	18.5	24.3
8	13.4	15.5	20.1	26.1
9	14.7	16.9	21.7	27.9
10	16.0	18.3	23.2	29.6

Appendix 4 Multiple alignment of 16S rRNA global

```
C._raciborskii_LMECYA132          AATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAATACCGCGTGAGGGAGGAAGGCTCTTGGGTCGTAAACCTCTTTTCT [ 80]
Aph._ovalisporum_ILC-146(KA1.1)   ............................................................T............. [ 80]
Ana._flos-aquae_LMECYA165         ............................................................................ [ 80]
Ana._lapponica_966                ............................................................T............. [ 80]
U._natans_TAC101                  ............................................................T............. [ 80]
A._issatschenkoi_LMECYA_31        ............................................................T............. [ 80]
M._aeruginosa_PCC7806             ................................CG................TCT....A.T............. [ 80]
Nostoc_sp._IO-102                 ............................................................................ [ 80]
Ana._flos-aquae_NIES-73           ............................................................................ [ 80]
Ana._kisseleviana_NIES-74         ............................................................................ [ 80]
Ana._bergii_283B                  ............................................................................ [ 80]
C._raciborskii_T3                 ............................................................................ [ 80]
Ana._bergii_360B                  ............................................................T............. [ 80]
Ana._bergii_366A                  ............................................................T............. [ 80]
Aph._aphanizomenoides_217A        ............................................................................ [ 80]
Nostochopsis_sp.89-45             ............................................................T............. [ 80]
Nostoc._sp._SAG                   ............................................................................ [ 80]
Aph._aphanizomenoides_235C        ............................................................................ [ 80]
G._violaceus_PCC7421              ......................R.........CG.......G.T..T......C.G...........G.C.....G. [ 80]
```

```
P._agardhii_213                 .....................................G........G..........T......A.T..C....C....... [ 80]
P._agardhii_NIVA-CYA_127        .....................................G........G..........T......A.T..C....C....... [ 80]
Fischerella_sp._MV9             ...............................................................T.............. [ 80]
F._muscicola_PCC7414            ...............................................................T.............. [ 80]
Ana._bergii_366B_(283A)         ...............................................................T.............. [ 80]
Aph._issatschenkoi_CAWBG02      ...............................................................T.............. [ 80]
Aph._ovalisporum_031E           ...............................................................T.............. [ 80]
Aph._ovalisporum_033B_(033E)    ...............................................................T.............. [ 80]
Aph._ovalisporum_033C_(031G)    ...............................................................T.............. [ 80]
Aph._sp._10E9                   ...............................................................T.............. [ 80]
Aph._sp._10E6                   ...............................................................T.............. [ 80]
Aph._sp._22D11                  ...............................................................T.............. [ 80]
Aph._ovalisporum_035C_(035D)    ...........C...................................................T.............. [ 80]
R._curvata_HB1                  ................................................................................ [ 80]
Aph._ovalisporum_035F_(035E)    ...........C...................................................T.............. [ 80]
C._raciborskii_AWT205(23E_24C_26 ................................................................................ [ 80]
C._raciborskii_030A_(023D_032A_0 ................................................................................ [ 80]
N._spumigena_PCC73104           ...............................................................T.............. [ 80]
Ana._flos-aquae_UTEX_LB2557     ...............................................................T.............. [ 80]
N._spumigena_BY1                ...............................................................T.............. [ 80]
Aph._aphanizomenoides_259C      ...............................................................T.............. [ 80]
```

```
Tri.variabilis_KCTC_AG10269      ..............................................................T.............. [ 80]
Ana._bergii_ZIE26AB              ..............................................................T.............. [ 80]
Tri.variabilis_HINDAK_2001/4     ..............................................................T.............. [ 80]
Aph.ovalisporum_FAS-AP1          ..............................................................T.............. [ 80]

C._raciborskii_LMECYA132         CAAGGAAGAAGAAAGTGACGGTACTTGAGGAATAAGCATCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGATG [160]
Aph._ovalisporum_ILC-146(KA1.1)  ..G.......A.TT--........C..................................................... [160]
Ana._flos-aquae_LMECYA165        ..................................................................................... [160]
Ana._lapponica_966               ..G.......A...A.........C..................................................... [160]
U._natans_TAC101                 ..G.......A.TT--........C..................................................... [160]
A._issatschenkoi_LMECYA_31       ..G.........C.A.........C..................................................... [160]
M._aeruginosa_PCC7806            ...........TTC-...................C...C.....................................G...G. [160]
Nostoc_sp._IO-102                ..G.......C.C.A.............................................................. [160]
Ana._flos-aquae_NIES-73          ..........TGC.A.............................................................. [160]
Ana._kisseleviana_NIES-74        ..........TGC.A.............................................................. [160]
Ana._bergii_283B                 ..................................................................................... [160]
C._raciborskii_T3                ..................................................................................... [160]
Ana._bergii_360B                 ..G.......A...--........C..................................................... [160]
Ana._bergii_366A                 ..G.......A...--........C..................................................... [160]
Aph._aphanizomenoides_217A       ..................................................................................... [160]
Nostochopsis_sp.89-45            ..G....T...C......A.....C........C............................................ [160]
```

```
Nostoc._sp._SAG                 ..G.......C.C.A.........C..................................................... [160]
Aph._aphanizomenoides_235C      ..............................................................................  [160]
G._violaceus_PCC7421            ..G...C...ACCT--........C...C................................................. [160]
P._agardhii_213                 ..G.......C.C.A.........C.......A.....................................G..... [160]
P._agardhii_NIVA-CYA_127        ..G.......C.C.A.........C.......A.....................................G..... [160]
Fischerella_sp._MV9             ............TG.-...................C..................................... [160]
F._muscicola_PCC7414            ...........-TCT-...................C..................................... [160]
Ana._bergii_366B_(283A)         ..G.......A.TT--........C..................................................... [160]
Aph._issatschenkoi_CAWBG02      ..G.......A...A.........C..................................................... [160]
Aph._ovalisporum_031E           ..G.......A.TT--........C..................................................... [160]
Aph._ovalisporum_033B_(033E)    ..G.......A.TT--........C..................................................... [160]
Aph._ovalisporum_033C_(031G)    ..G.......A.TT--........C..................................................... [160]
Aph._sp._10E9                   ..G.......C...A.........C..................................................... [160]
Aph._sp._10E6                   ..G.......C.G.A.........C..................................................... [160]
Aph._sp._22D11                  ..G.......C...A.........C..................................................... [160]
Aph._ovalisporum_035C_(035D)    ..G.......A.TT--........C..................................................... [160]
R._curvata_HB1                  ..............................................................................  [160]
Aph._ovalisporum_035F_(035E)    ..G.......A.TT--........C..................................................... [160]
C._raciborskii_AWT205(23E_24C_26 ..............................................................................  [160]
C._raciborskii_030A_(023D_032A_0 ..............................................................................  [160]
N._spumigena_PCC73104           ..G.......A...A.........C..................................................... [160]
```

```
Ana._flos-aquae_UTEX_LB2557      ..G.......C.C.A.........C.........C.......................................... [160]
N._spumigena_BY1                 ..G.......A...A.........C.................................................... [160]
Aph._aphanizomenoides_259C       ..G....T..........A.....C.................................................... [160]
Tri.variabilis_KCTC_AG10269      ..G.......A...A.........C.................................................... [160]
Ana._bergii_ZIE26AB              ..G.......A...--........C.................................................... [160]
Tri.variabilis_HINDAK_2001/4     ..G.......A...A.........C.................................................... [160]
Aph.ovalisporum_FAS-AP1          ..G.......A.TT--........C.................................................... [160]

C._raciborskii_LMECYA132         CAAGCGTTATCCGGAATGATTGGGCGTAAAGGGTCTGCAGGTGGAACTGAAAGTCTGCTGTTAAAGAGTTTGGCTTAACC [240]
Aph._ovalisporum_ILC-146(KA1.1)  ..........................................CCG..T.................A.C.......... [240]
Ana._flos-aquae_LMECYA165        ....................................T......T......................CAA....C.... [240]
Ana._lapponica_966               ..............................C...C........TTA..T..............T...CAT...C...A [240]
U._natans_TAC101                 ..................................C........CTA..T.................A.GA......... [240]
A._issatschenkoi_LMECYA_31       ..................................C........T.G..T.................A.CAC...C...G [240]
M._aeruginosa_PCC7806            ................T.............C...C........TCAGCC........C..C...TCAGG.T.......G [240]
Nostoc_sp._IO-102                ..............................C...C........C.A..T..................C.A...C...T [240]
Ana._flos-aquae_NIES-73          ....................................T......T......................CAA....C.... [240]
Ana._kisseleviana_NIES-74        ....................................T......T......................CAA....C.... [240]
Ana._bergii_283B                 ....................................T......T......................CAA....C.... [240]
C._raciborskii_T3                ................................................................................ [240]
Ana._bergii_360B                 ..........................................CTA..T.................A.C.A.......T [240]
```

```
Ana._bergii_366A                ..........................................CTA..T.................A.C.A.......T [240]
Aph._aphanizomenoides_217A      ....................................T......T......................CAA....C.... [240]
Nostochopsis_sp.89-45           ..............................C...C.T....A.C.G..TGT....AT.................C.... [240]
Nostoc._sp._SAG                 ..............................C...C........CTA..T...................C.A...C...T [240]
Aph._aphanizomenoides_235C      ....................................T......T......................CAA....C.... [240]
G._violaceus_PCC7421            ................T.............C..AC.T...C..CCAGCC.......GG..C....CCCGCA..C....T [240]
P._agardhii_213                 ..............................A...C.T....A.TCA.CC..................CGA......... [240]
P._agardhii_NIVA-CYA_127        ..............................A...C.T....A.TCA.CC..................CGA......... [240]
Fischerella_sp._MV9             ..............................C...C........TTG..TGT....AT...C....CA.CAT.......A [240]
F._muscicola_PCC7414            ..............................C...C........TTG..TGT....AT...C....G..CAT.......A [240]
Ana._bergii_366B_(283A)         ..........................................CCG..T.................A.C.......... [240]
Aph._issatschenkoi_CAWBG02      ....................................C........T.G..T...................CAC...C...G [240]
Aph._ovalisporum_031E           ..........................................CCG..T.................A.C.......... [240]
Aph._ovalisporum_033B_(033E)    ..........................................CCG..T.................A.C.......... [240]
Aph._ovalisporum_033C_(031G)    ..........................................CCG..T.................A.C.......... [240]
Aph._sp._10E9                   ....................................C........C.T..T.......................C.... [240]
Aph._sp._10E6                   ....................................C........C.T..T.......................C.... [240]
Aph._sp._22D11                  ....................................C........C.T..T.......................C.... [240]
Aph._ovalisporum_035C_(035D)    ..........................................CCG..T.................A.C.......... [240]
R._curvata_HB1                  ...................................................................C.... [240]
Aph._ovalisporum_035F_(035E)    ..........................................CCG..T.................A.C.......... [240]
```

```
C._raciborskii_AWT205(23E_24C_26 ..................................................................... [240]
C._raciborskii_030A_(023D_032A_0 ..................................................................... [240]
N._spumigena_PCC73104            ..................................C........CTA..T.................ACC.A.......T [240]
Ana._flos-aquae_UTEX_LB2557      ...............................C...C........CTG..T...........C......CAAA...C...T [240]
N._spumigena_BY1                 ..................................C........CTA..T.................ACC.A.......T [240]
Aph._aphanizomenoides_259C       ...............................C...C........C....T.................CAA....C.... [240]
Tri.variabilis_KCTC_AG10269      ..................................C........CTG..T..............T...CAT...C...A [240]
Ana._bergii_ZIE26AB              ........................................CTA..T.................A.C.A.......T [240]
Tri.variabilis_HINDAK_2001/4     ..................................C........CTA..T..............T...CAT...C...A [240]
Aph.ovalisporum_FAS-AP1          ........................................CCG..T.................A.C.......... [240]

C._raciborskii_LMECYA132         AAATAAAAGCGGTGGAAACTACAGAACTAGAGTGCGGTAGGGGCAAAAGGAATTCCTGGTGTAGCGGTGAAATGCGTAGA [320]
Aph._ovalisporum_ILC-146(KA1.1)  .G........A.............GG..........T.C....T.G.G.............................. [320]
Ana._flos-aquae_LMECYA165        TTG...G...A..................................................................... [320]
Ana._lapponica_966               TG....GG..A.............AG...........T.C......G.G.............................. [320]
U._natans_TAC101                 TC..C.G...A.............C.G..........T.C....T.G.G.............................. [320]
A._issatschenkoi_LMECYA_31       TG..C.....A.............C.........A...........G................................ [320]
M._aeruginosa_PCC7806            .CC....G.............GGCAG.......A..A.......T.GC.........CA..................... [320]
Nostoc_sp._IO-102                .G....G...A.............T.G.......A..T.C......G.G.............................. [320]
Ana._flos-aquae_NIES-73          TTG...G...A.................G.................................................... [320]
Ana._kisseleviana_NIES-74        TTG...G...A.................G.................................................... [320]
```

```
Ana._bergii_283B              TTG...G...A.......................................................................... [320]
C._raciborskii_T3             ..................................................................................... [320]
Ana._bergii_360B              .G........A.............GG..........T.C....T.G.G..................................... [320]
Ana._bergii_366A              .G........A.............G..........T.C....T.G.G..................................... [320]
Aph._aphanizomenoides_217A    TTG...G...A.......................................................................... [320]
Nostochopsis_sp.89-45         ......GG.....A.........C.G..........T.C......G.G..................................... [320]
Nostoc._sp._SAG               .G....G...A............TGG..........T.C......G.G..................................... [320]
Aph._aphanizomenoides_235C    TTG...G...A.......................................................................... [320]
G._violaceus_PCC7421          GCGG...G..CC........GGTTGG.....................GG........C.................T.G... [320]
P._agardhii_213               TCG....G..A........CGG.AG.........TA.........G.G..................................... [320]
P._agardhii_NIVA-CYA_127      TCG....G..A.........GG.AG.........TA.........G.G..................................... [320]
Fischerella_sp._MV9           TG.....G..A..A.........CG...........T.C......G.G..................................... [320]
F._muscicola_PCC7414          TG.....G..A............CG...........T.C......G.G..................................... [320]
Ana._bergii_366B_(283A)       .G........A.............GG..........T.C....T.G.G..................................... [320]
Aph._issatschenkoi_CAWBG02    TG..C.....A............C.........A...........G..................................... [320]
Aph._ovalisporum_031E         .G........A.............GG..........T.C....T.G.G..................................... [320]
Aph._ovalisporum_033B_(033E)  .G........A.............GG..........T.C....T.G.G..................................... [320]
Aph._ovalisporum_033C_(031G)  .G........A.............GG..........T.C....T.G.G..................................... [320]
Aph._sp._10E9                 ..........A............A.G........T...C......G.G..................................... [320]
Aph._sp._10E6                 ..........A............A.G........T...C......G.G..................................... [320]
Aph._sp._22D11                ..........A............A.G........T...C......G.G..................................... [320]
```

```
Aph._ovalisporum_035C_(035D)     .G........A.............GG..........T.C....T.G.G.............................. [320]
R._curvata_HB1                   ..................................T.......................................... [320]
Aph._ovalisporum_035F_(035E)     .G........A.............GG..........T.C....T.G.G.............................. [320]
C._raciborskii_AWT205(23E_24C_26 ..................................................................................... [320]
C._raciborskii_030A_(023D_032A_0 ..................................................................................... [320]
N._spumigena_PCC73104            .GG.......A.............AGG..........T.C....T.G.G.............................. [320]
Ana._flos-aquae_UTEX_LB2557      TTG....G..A.............CGG..........T.C......G.G.............................. [320]
N._spumigena_BY1                 .GG.......A.............AGG..........T.C....T.G.G.............................. [320]
Aph._aphanizomenoides_259C       TTG....G..A..............G.......A..T.C......G.G.............................. [320]
Tri.variabilis_KCTC_AG10269      TG.....G..A.............T.G..........T.C......G.G.............................. [320]
Ana._bergii_ZIE26AB              GG........A.............GG..........T.C....T.G.G.............................. [320]
Tri.variabilis_HINDAK_2001/4     TG..C..G..A.............AGG..........T.C......G.G.............................. [320]
Aph.ovalisporum_FAS-AP1          .G........A.............GG..........T.C....T.G.G.............................. [320]

C._raciborskii_LMECYA132         TATCAGGAAGAACACCGGTGGCGAAAGCGTTTTGCTAGACCGTAACTGACACTGAGGGACGAAAGCTAGGGGAGCGAATG [400]
Aph._ovalisporum_ILC-146(KA1.1)  ................A.............C.C.A....G...C..................................... [400]
Ana._flos-aquae_LMECYA165        ................................C.A..G.G.T...T........C.......................A. [400]
Ana._lapponica_966               ..............................C.C......G...C..................................... [400]
U._natans_TAC101                 ................A.............C.C.A....G...C..................................... [400]
A._issatschenkoi_LMECYA_31       ................................C.........TGT..................................... [400]
M._aeruginosa_PCC7806            G..TG.........T...............GC.A..G.G.T...T........C.......................A. [400]
```

```
Nostoc_sp._IO-102              G.....................G...C.C......G..................................... [400]
Ana._flos-aquae_NIES-73        .........................................C..................................... [400]
Ana._kisseleviana_NIES-74      .........................................C..................................... [400]
Ana._bergii_283B               .........................................C..................................... [400]
C._raciborskii_T3              ........................................................................... [400]
Ana._bergii_360B               ................A.A.........C.C.A....G...C..................................... [400]
Ana._bergii_366A               ................A.A.........C.C.A....G...C..................................... [400]
Aph._aphanizomenoides_217A     .........................................C..................................... [400]
Nostochopsis_sp.89-45          G.........................C.C......G...C..................................... [400]
Nostoc._sp._SAG                G.....................G...C.C......G...C..................................... [400]
Aph._aphanizomenoides_235C     .........................................C..................................... [400]
G._violaceus_PCC7421           ....G...........A.C.........CC.....G.....C.......G......T........C........A.... [400]
P._agardhii_213                G.....................G...C.C....G.G.TA..................................... [400]
P._agardhii_NIVA-CYA_127       G.....................G...C.C....G.G.TA..................................... [400]
Fischerella_sp._MV9            G.........................C.C......G...C..................................... [400]
F._muscicola_PCC7414           G.........................C.C......G..TGC..................................... [400]
Ana._bergii_366B_(283A)        ................A...........C.C.A....G...C..................................... [400]
Aph._issatschenkoi_CAWBG02     ............................C.........TGT..................................... [400]
Aph._ovalisporum_031E          ................A...........C.C.A....G...C..................................... [400]
Aph._ovalisporum_033B_(033E)   ................A...........C.C.A....G...C..................................... [400]
Aph._ovalisporum_033C_(031G)   ................A...........C.C.A....G...C..................................... [400]
```

```
Aph._sp._10E9                  ..............T.........G...C.C......G..AAG..................................... [400]
Aph._sp._10E6                  ..............T.........G...C.C......G..AAG..................................... [400]
Aph._sp._22D11                 ........................G...C.C......G..AAG..................................... [400]
Aph._ovalisporum_035C_(035D)   ................A...........C.C.A....G...C..................................... [400]
R._curvata_HB1                 .........................................A..................................... [400]
Aph._ovalisporum_035F_(035E)   ................A...........C.C.A....G...C..................................... [400]
C._raciborskii_AWT205(23E_24C_26 ..................................................................................... [400]
C._raciborskii_030A_(023D_032A_0 ..................................................................................... [400]
N._spumigena_PCC73104          ................A.......G...C.C.A....G...C..................................... [400]
Ana._flos-aquae_UTEX_LB2557    G......G................G...C.C......G...C..................................... [400]
N._spumigena_BY1               ................A.......G...C.C.A....G...C..................................... [400]
Aph._aphanizomenoides_259C     G..........................C.C......G.......................................... [400]
Tri.variabilis_KCTC_AG10269    ..........................C.C......G...C..................................... [400]
Ana._bergii_ZIE26AB            ................A.A.........C.C.A....G...C..................................... [400]
Tri.variabilis_HINDAK_2001/4   ..........................C.C......G...C..................................... [400]
Aph.ovalisporum_FAS-AP1        ................A...........C.C.A....G...C..................................... [400]

C._raciborskii_LMECYA132       GGATTAGATACCCCAGTAGTCCTAGCCGTAAACGATGGATACTAGGCGTGGCTTGTA [457]
Aph._ovalisporum_ILC-146(KA1.1) ..................................................... [457]
Ana._flos-aquae_LMECYA165      ..................................................... [457]
Ana._lapponica_966             ..............................................T.......... [457]
```

```
U._natans_TAC101                ..................................................... [457]
A._issatschenkoi_LMECYA_31      .............................................T.......... [457]
M._aeruginosa_PCC7806           ..............T....................................... [457]
Nostoc_sp._IO-102               ..................................................... [457]
Ana._flos-aquae_NIES-73         ..................................................... [457]
Ana._kisseleviana_NIES-74       ..................................................... [457]
Ana._bergii_283B                ..................................................... [457]
C._raciborskii_T3               ..................................................... [457]
Ana._bergii_360B                ..................................................... [457]
Ana._bergii_366A                ..................................................... [457]
Aph._aphanizomenoides_217A      ..................................................... [457]
Nostochopsis_sp.89-45           ...............................................T.AGA... [457]
Nostoc._sp._SAG                 ..................................................... [457]
Aph._aphanizomenoides_235C      ..................................................... [457]
G._violaceus_PCC7421            ......................G..............C......T..T..GC... [457]
P._agardhii_213                 ........................G............A......T.....C.... [457]
P._agardhii_NIVA-CYA_127        ........................G............A......T.....C.... [457]
Fischerella_sp._MV9             ...............................................T.AGA... [457]
F._muscicola_PCC7414            ...............................................T.AGA... [457]
Ana._bergii_366B_(283A)         ..................................................... [457]
Aph._issatschenkoi_CAWBG02      .............................................T.......... [457]
```

```
Aph._ovalisporum_031E            ..................................................... [457]
Aph._ovalisporum_033B_(033E)     ..................................................... [457]
Aph._ovalisporum_033C_(031G)     ..................................................... [457]
Aph._sp._10E9                    .................................................A...C... [457]
Aph._sp._10E6                    .................................................A...C... [457]
Aph._sp._22D11                   .................................................A...C... [457]
Aph._ovalisporum_035C_(035D)     ..................................................... [457]
R._curvata_HB1                   ..................................................... [457]
Aph._ovalisporum_035F_(035E)     ..................................................... [457]
C._raciborskii_AWT205(23E_24C_26 ..................................................... [457]
C._raciborskii_030A_(023D_032A_0 ..................................................... [457]
N._spumigena_PCC73104            ..................................................... [457]
Ana._flos-aquae_UTEX_LB2557      ........................T............................ [457]
N._spumigena_BY1                 ..................................................... [457]
Aph._aphanizomenoides_259C       ..................................................... [457]
Tri.variabilis_KCTC_AG10269      ..................................................... [457]
Ana._bergii_ZIE26AB              ..................................................... [457]
Tri.variabilis_HINDAK_2001/4     ..................................................... [457]
Aph.ovalisporum_FAS-AP1          ..................................................... [457]
;
end;
```

Appendix 5 Multiple alignment of 16S rRNA CYN-producing cyanobacteria only

```
Aph._aphanizomenoides_259C       AATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAATACCGCGTGAGGGAGGAAGGCTCTTGGGTTGTAAACCTCTTTTCT [ 80]
Ana._lapponica_966               ..............................................................................  [ 80]
Ana._bergii_366B_(283A)          ..............................................................................  [ 80]
Aph._issatschenkoi_CAWBG02       ..............................................................................  [ 80]
Aph._ovalisporum_031E            ..............................................................................  [ 80]
A._ovalisporum_033B_(033E)       ..............................................................................  [ 80]
A._ovalisporum_033C_(031G)       ..............................................................................  [ 80]
Aph_sp._10E9                     ..............................................................................  [ 80]
Aph_sp._10E6                     ..............................................................................  [ 80]
Aph_sp._22D11                    ..............................................................................  [ 80]
Aph._ovalisporum_035C_(035D)     ...........C..................................................................  [ 80]
Aph._ovalisporum_035F_(035E)     ...........C..................................................................  [ 80]
C._raci._AWT205_(023E_024C_026J_ ..............................................................C...............  [ 80]
C._raciborskii_030A_(023D_032A_0 ..............................................................C...............  [ 80]
Aph.ovalisporum_ILC-146_(KA1.1)  ..............................................................................  [ 80]

Aph._aphanizomenoides_259C       CAGGGAATAAGAAAGTGAAGGTACCTGAGGAATAAGCATCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGATG [160]
Ana._lapponica_966               .......G..A...A...C..........................................................  [160]
Ana._bergii_366B_(283A)          .......G..--..T...C..........................................................  [160]
```

```
Aph._issatschenkoi_CAWBG02       .......G..A...A...C.......................................................... [160]
Aph._ovalisporum_031E            .......G..--..T...C.......................................................... [160]
A._ovalisporum_033B_(033E)       .......G..--..T...C.......................................................... [160]
A._ovalisporum_033C_(031G)       .......G..--..T...C.......................................................... [160]
Aph_sp._10E9                     .......G..C...A...C.......................................................... [160]
Aph_sp._10E6                     .......G..C.G.A...C.......................................................... [160]
Aph_sp._22D11                    .......G..C...A...C.......................................................... [160]
Aph._ovalisporum_035C_(035D)     .......G..--..T...C.......................................................... [160]
Aph._ovalisporum_035F_(035E)     .......G..--..T...C.......................................................... [160]
C._raci._AWT205_(023E_024C_026J_ ..A....G.........C.....T..................................................... [160]
C._raciborskii_030A_(023D_032A_0 ..A....G.........C.....T..................................................... [160]
Aph.ovalisporum_ILC-146_(KA1.1)  .......G..--..T...C.......................................................... [160]

Aph._aphanizomenoides_259C       CAAGCGTTATCCGGAATGATTGGGCGTAAAGCGTCCGCAGGTGGCACTGTAAGTCTGCTGTTAAAGAGCAAGGCTCAACC [240]
Ana._lapponica_966               ..........................................TTA..................T..TC.T.......A [240]
Ana._bergii_366B_(283A)          ............................G...T.........CG...................ATCT....T.... [240]
Aph._issatschenkoi_CAWBG02       ............................G............T.G.....................TC.C.......G [240]
Aph._ovalisporum_031E            ............................G...T.........CG...................ATCT....T.... [240]
A._ovalisporum_033B_(033E)       ............................G...T.........CG...................ATCT....T.... [240]
A._ovalisporum_033C_(031G)       ............................G...T.........CG...................ATCT....T.... [240]
Aph_sp._10E9                     ............................G..............T...................TTT......... [240]
```

```
Aph_sp._10E6                    ..............................G.............T.....................TTT......... [240]
Aph_sp._22D11                   ..............................G.............T.....................TTT......... [240]
Aph._ovalisporum_035C_(035D)    ..............................G...T.........CG...................ATCT....T.... [240]
Aph._ovalisporum_035F_(035E)    ..............................G...T.........CG...................ATCT....T.... [240]
C._raci._AWT205_(023E_024C_026J_ ..............................G...T........A....A.................TTT....T.... [240]
C._raciborskii_030A_(023D_032A_0 ..............................G...T........A....A.................TTT....T.... [240]
Aph.ovalisporum_ILC-146_(KA1.1) ..............................G...T.........CG...................ATCT....T.... [240]

Aph._aphanizomenoides_259C      TTGTAAAGGCAGTGGAAACTACAGAGCTAGAGTACGTTCGGGGCAGAGGGAATTCCTGGTGTAGCGGTGAAATGCGTAGA [320]
Ana._lapponica_966              .GA...G.................AGA.......G.......................................... [320]
Ana._bergii_366B_(283A)         AGA....A................G........G.........T................................... [320]
Aph._issatschenkoi_CAWBG02      .GA.C..A................C.A..........G.A........A............................... [320]
Aph._ovalisporum_031E           AGA....A................G........G.........T................................... [320]
A._ovalisporum_033B_(033E)      AGA....A................G........G.........T................................... [320]
A._ovalisporum_033C_(031G)      AGA....A................G........G.........T................................... [320]
Aph_sp._10E9                    AAA....A................A.........GT.G....................................... [320]
Aph_sp._10E6                    AAA....A................A.........GT.G....................................... [320]
Aph_sp._22D11                   AAA....A................A.........GT.G....................................... [320]
Aph._ovalisporum_035C_(035D)    AGA....A................G........G.........T................................... [320]
Aph._ovalisporum_035F_(035E)    AGA....A................G........G.........T................................... [320]
C._raci._AWT205_(023E_024C_026J_ AAA....A..G..............A.......G..G.A......A.A............................... [320]
```

```
C._raciborskii_030A_(023D_032A_0 AAA....A..G.............A.......G..G.A......A.A.............................. [320]
Aph.ovalisporum_ILC-146_(KA1.1)  AGA....A.............G........G.........T................................... [320]

Aph._aphanizomenoides_259C       GATCAGGAAGAACACCGGTGGCGAAAGCGCTCTGCTAGGCCGTAACTGACACTGAGGGACGAAAGCTAGGGGAGCGAATG [400]
Ana._lapponica_966               T.....................................C..................................... [400]
Ana._bergii_366B_(283A)          T..............A..............A........C..................................... [400]
Aph._issatschenkoi_CAWBG02       T............................T........A..TGT................................... [400]
Aph._ovalisporum_031E            T..............A..............A........C..................................... [400]
A._ovalisporum_033B_(033E)       T..............A..............A........C..................................... [400]
A._ovalisporum_033C_(031G)       T..............A..............A........C..................................... [400]
Aph_sp._10E9                     T.............T.........G..............AAG.................................... [400]
Aph_sp._10E6                     T.............T.........G..............AAG.................................... [400]
Aph_sp._22D11                    T.......................G..............AAG.................................... [400]
Aph._ovalisporum_035C_(035D)     T..............A..............A........C..................................... [400]
Aph._ovalisporum_035F_(035E)     T..............A..............A........C..................................... [400]
C._raci._AWT205_(023E_024C_026J_ T............................T.T......A..................................... [400]
C._raciborskii_030A_(023D_032A_0 T............................T.T......A..................................... [400]
Aph.ovalisporum_ILC-146_(KA1.1)  T..............A..............A........C..................................... [400]

Aph._aphanizomenoides_259C       GGATTAGATACCCCAGTAGTCCTAGCCGTAAACGATGGATACTAGGCGTGGCTTGTA [457]
Ana._lapponica_966               ..........................................T.......... [457]
```

```
Ana._bergii_366B_(283A)          ..................................................... [457]
Aph._issatschenkoi_CAWBG02       ..............................................T..... [457]
Aph._ovalisporum_031E            ..................................................... [457]
A._ovalisporum_033B_(033E)       ..................................................... [457]
A._ovalisporum_033C_(031G)       ..................................................... [457]
Aph_sp._10E9                     ..............................................A...C... [457]
Aph_sp._10E6                     ..............................................A...C... [457]
Aph_sp._22D11                    ..............................................A...C... [457]
Aph._ovalisporum_035C_(035D)     ..................................................... [457]
Aph._ovalisporum_035F_(035E)     ..................................................... [457]
C._raci._AWT205_(023E_024C_026J_ ..................................................... [457]
C._raciborskii_030A_(023D_032A_0 ..................................................... [457]
Aph.ovalisporum_ILC-146_(KA1.1)  ..................................................... [457]
```

Appendix 6 Multiple alignment of *rpo*C1

```
Aph.aphanizomenoides_259C        CCAGAAACGATTAATTACCGGACTCTCAAGCCAGAAATGGATGGCTTGTTTTGTGAGCGCATCTTTGGCCCCGCGAAA [ 78]
Ana._bergii_283A                 ........A..........T..G........C........C..........C.................C... [ 78]
Ana.berg_366B                    ........A..........T..G........C........C..........C.................C... [ 78]
Aph.issatschenkoi_CAWBG02        ..T.....T.....C.....C..C.C...A.....................C........T.....A..G..T... [ 78]
Aph._ovalisporum_031E            ........A..........T..G........C........C..........C.................C... [ 78]
Aph._ovalisporum_031G(033C)      ........A..........T..G........C........C..........C.................C... [ 78]
Aph._ovalisporum_033B(035E)      ........A..........T..G........C........C..........C.................C... [ 78]
Aph._ovalisporum_033E            ........A..........T..G........C........C..........C.................C... [ 78]
Aph._ovalisporum_035C            ........A..........T..G........C........C..........C.................C... [ 78]
Aph._ovalisporum_035D            ........A..........T..G........C........C..........C.................C... [ 78]
Aph._ovalisporum_035F            ........A..........T..G........C........C..........C.................C... [ 78]
C.raciborskii_AWT205(023E_024C_0 ..T.....C.....C.....C..C.....................T..A..........G..........T..T..G [ 78]
C.raciborskii_030A(023D_032A_034 ..T.....C.....C.....C..C.....................T..A..........G..........T..T..G [ 78]
Aph_sp._22D11                    ........T..........C..........G.....................A.G..T..C..A..A..T..G [ 78]
Aph_sp._10E6                     ........T..........C..........G.....................A.G..T..C..A..A..T..G [ 78]
Aph_sp._10E9                     ........T..........C..........G.....................A.G..T..C..A..A..T..G [ 78]
Ana.lapponica_966                ........T..C.............G.....G...........T.....C..C..A..G........A..G..T..G [ 78]
Aph._ovalisporumILC-146          ........A..........T..G........C........C..........C.................C... [ 78]
```

```
Aph.aphanizomenoides_259C        GATTGGGAATGCCATTGCGGTAAGTATAAGAGAGTCCGCCATAGAGGCATTGTCTGTGAGCGTTGTGGCGTGGAAGTT [156]
Ana._bergii_283A                 ..C........T..C..T.....A.....A........T..CC.G..T..A.....C..A..C..C.....A...... [156]
Ana.berg_366B                    ..C........T..C..T.....A.....A........T..CC.G..T..A.....C..A..C..C.....A...... [156]
Aph.issatschenkoi_CAWBG02        ..C........T.....T..C..A..C..AC.G..G...........T..A..T.....A..C..C..T......... [156]
Aph._ovalisporum_031E            ..C........T..C..T.....A.....A........T..CC.G..T..A.....C..A..C..C.....A...... [156]
Aph._ovalisporum_031G(033C)      ..C........T..C..T.....A.....A........T..CC.G..T..A.....C..A..C..C.....A...... [156]
Aph._ovalisporum_033B(035E)      ..C........T..C..T.....A.....A........T..CC.G..T..A.....C..A..C..C.....A...... [156]
Aph._ovalisporum_033E            ..C........T..C..T.....A.....A........T..CC.G..T..A.....C..A..C..C.....A...... [156]
Aph._ovalisporum_035C            ..C........T..C..T.....A.....A........T..CC.G..T..A.....C..A..C..C.....A...... [156]
Aph._ovalisporum_035D            ..C........T..C..T.....A.....A........T..CC.G..T..A.....C..A..C..C.....A...... [156]
Aph._ovalisporum_035F            ..C........T..C..T.....A.....A........T..CC.G..T..A.....C..A..C..C.....A...... [156]
C.raciborskii_AWT205(023E_024C_0 ...........T..C..T.....A.....AC.G..T..T........T.............C.....A.....G..A [156]
C.raciborskii_030A(023D_032A_034 ...........T..C..T.....A.....AC.G..T..T........T.............C.....A.....G..A [156]
Aph_sp._22D11                    ........G........T............C.G..G...........G.....G..C.....C.....T..A.....G [156]
Aph_sp._10E6                     ........G........T............C.G..G...........G.....G..C.....C.....T..A.....G [156]
Aph_sp._10E9                     ........G........T............C.G..G...........G.....G..C.....C.....T..A.....G [156]
Ana.lapponica_966                ...........T..................C.T..G..T..CC.T..T...........A..C..C..T..A.....G [156]
Aph._ovalisporumILC-146          ..C........T..C..T.....A.....A........T..CC.G..T..A.....C..A..C..C.....A...... [156]


Aph.aphanizomenoides_259C        ACCGAGTCACGGGTACGCCGTCACCGTATGGGGTACATCAAACTTGCCGCCCCTGTTGCCCACGTTTGGTATCTCAAA [234]
Ana._bergii_283A                 ..T.....C.....T..T........C.....T.....T.....G..G..T..A..G.....T.............. [234]
```

```
Ana.berg_366B                    ..T.....C.....T..T........C.....T.....T.....G..G..T..A..G.....T.............. [234]
Aph.issatschenkoi_CAWBG02        ..T.....T..C..C..T..C...A.A.....C.TT..T.....G..A..A..C..A..T..T.............G [234]
Aph._ovalisporum_031E            ..T.....C.....T..T........C.....T.....T.....G..G..T..A..G.....T.............. [234]
Aph._ovalisporum_031G(033C)      ..T.....C.....T..T........C.....T.....T.....G..G..T..A..G.....T.............. [234]
Aph._ovalisporum_033B(035E)      ..T.....C.....T..T........C.....T.....T.....G..G..T..A..G.....T.............. [234]
Aph._ovalisporum_033E            ..T.....C.....T..T........C.....T.....T.....G..G..T..A..G.....T.............. [234]
Aph._ovalisporum_035C            ..T.....C.....T..T........C.....T.....T.....G..G..T..A..G.....T.............. [234]
Aph._ovalisporum_035D            ..T.....C.....T..T........C.....T.....T.....G..G..T..A..G.....T.............. [234]
Aph._ovalisporum_035F            ..T.....C.....T..T........C.....T.....T.....G..G..T..A..G.....T.............. [234]
C.raciborskii_AWT205(023E_024C_0 ..A..A.....A..T.....C..TA.A.....C.T...T..GT.A..A..A.....A.....T.............. [234]
C.raciborskii_030A(023D_032A_034 ..A..A.....A..T.....C..TA.A.....C.T...T..GT.A..A..A.....A.....T.............. [234]
Aph_sp._22D11                    ..A..A..C..C..C..T......A.G.....T.TT..T..GT.G..T.....G..G..T..T.........T.A..G [234]
Aph_sp._10E6                     ..A..A..C..C..C..T......A.G.....T.TT..T..GT.G..T.....G..G..T..T.........T.A..G [234]
Aph_sp._10E9                     ..A..A..C..C..C..T......A.G.....T.TT..T..GT.G..T.....G..G..T..T.........T.A..G [234]
Ana.lapponica_966                ..T..A.....C..T..T........C.....T.TT..T..GT.G..A..G..A..A..T..T..A.....C..G..G [234]
Aph._ovalisporumILC-146          ..T.....C.....T..T........C.....T.....T.....G..G..T..A..G.....T.............. [234]

Aph.aphanizomenoides_259C        GGGATTCCTAGCTATATTTCTATTCTGCTAGATATGCCTTTGCGGGATGTGGAGCAGATTGTCTATTTCAACTCTTAT [312]
Ana._bergii_283A                 ..C..C..C.....C..C..C.....C..G..............A..C..A.....A.....T...........C..C [312]
Ana.berg_366B                    ..C..C..C.....C..C..C.....C..G..............A..C..A.....A.....T...........C..C [312]
Aph.issatschenkoi_CAWBG02        ...........T......G....CT....G........CC.....................T...........C... [312]
```

```
Aph._ovalisporum_031E           ..C..C..C.....C..C..C.....C..G.............A..C..A.....A.....T..........C..C [312]
Aph._ovalisporum_031G(033C)     ..C..C..C.....C..C..C.....C..G.............A..C..A.....A.....T..........C..C [312]
Aph._ovalisporum_033B(035E)     ..C..C..C.....C..C..C.....C..G.............A..C..A.....A.....T..........C..C [312]
Aph._ovalisporum_033E           ..C..C..C.....C..C..C.....C..G.............A..C..A.....A.....T..........C..C [312]
Aph._ovalisporum_035C           ..C..C..C.....C..C..C.....C..G.............A..C..A.....A.....T..........C..C [312]
Aph._ovalisporum_035D           ..C..C..C.....C..C..C.....C..G.............A..C..A.....A.....T..........C..C [312]
Aph._ovalisporum_035F           ..C..C..C.....C..C..C.....C..G.............A..C..A.....A.....T..........C..C [312]
C.raciborskii_AWT205(023E_024C_0 ..C........T......G.C..A..A.....C.....CC.A..A.....T.....A..C.....C..T.....C... [312]
C.raciborskii_030A(023D_032A_034 ..C........T......G.C..A..A.....C.....CC.A..A.....T.....A..C.....C..T.....C... [312]
Aph_sp._22D11                   ..................G.G.....C..G........GC.....................T.............C [312]
Aph_sp._10E6                    ..................G.G.....C..G........GC.....................T.............C [312]
Aph_sp._10E9                    ..................G.G.....C..G........GC.....................T.............C [312]
Ana.lapponica_966               ...........T..C...G....C..CT..........C..........A.....A.....T..C..T..T.....C [312]
Aph._ovalisporumILC-146         ..C..C..C.....C..C..C.....C..G.............A..C..A.....A.....T..........C..C [312]

Aph.aphanizomenoides_259C       GTTGTTCTCAGTCCCGGTAATGCTGAAACCCTAACCTACAAGCAGCTACTGAGTGAAGACCAGTGGTTAGAAATTGAA [390]
Ana._bergii_283A                ..C..C.............C..C.....AT..........A.....T..C..C........A...........A... [390]
Ana.berg_366B                   ..C..C.............C..C.....AT..........A.....T..C..C........A...........A... [390]
Aph.issatschenkoi_CAWBG02       ..A..CT.AGC...T...........T.....T.....T..A..A..C..C.......................... [390]
Aph._ovalisporum_031E           ..C..C.............C..C.....AT..........A.....T..C..C........A...........A... [390]
Aph._ovalisporum_031G(033C)     ..C..C.............C..C.....AT..........A.....T..C..C........A...........A... [390]
```

```
Aph._ovalisporum_033B(035E)      ..C..C.............C..C.....AT.........A.....T..C..C........A..........A... [390]
Aph._ovalisporum_033E            ..C..C.............C..C.....AT.........A.....T..C..C........A..........A... [390]
Aph._ovalisporum_035C            ..C..C.............C..C.....AT.........A.....T..C..C........A..........A... [390]
Aph._ovalisporum_035D            ..C..C.............C..C.....AT.........A.....T..C..C........A..........A... [390]
Aph._ovalisporum_035F            ..C..C.............C..C.....AT.........A.....T..C..C........A..........A... [390]
C.raciborskii_AWT205(023E_024C_0 ..A...T.AGA...T..........C..A..CGTT..T..A..A..GT.A.CA........A.............G [390]
C.raciborskii_030A(023D_032A_034 ..A...T.AGA...T..........C..A..CGTT..T..A..A..GT.A.CA........A.............G [390]
Aph_sp._22D11                    TG....T.AGCG..T.....C..G..T..T..C.GT.....A..A..GT.................C..........G [390]
Aph_sp._10E6                     TG....T.AGCG..T........G..T..T..C.GT.....A..A..GT.................C..........G [390]
Aph_sp._10E9                     TG....T.AGCG..T........G..T..T..C.GT.....A..A..GT.................C..........G [390]
Ana.lapponica_966                TGC...T.GGC......A.....C..T.....C.G...T..A..AT.G..............A.............. [390]
Aph._ovalisporumILC-146          ..C..C.............C..C.....AT.........A.....T..C..C........A..........A... [390]

Aph.aphanizomenoides_259C        GACCAAATCTATAGTGAAGATTCTCAGTTGCAAGGTGTAGAAGTAGGTATCGGTGCTGAAGCACTGTTGCGCTTGCTG [468]
Ana._bergii_283A                 ........T..........C....TAC.C..G..C........G..C..T.....A......T.AC.......A..C [468]
Ana.berg_366B                    ........T..........C....TAC.C..G..C........G..C..T.....A......T.AC.......A..C [468]
Aph.issatschenkoi_CAWBG02        ..TGCT.....C..C.....C..G.....AG.......G.....G..C..T.....G..G...T..C....G..AT.. [468]
Aph._ovalisporum_031E            ........T..........C....TAC.C..G..C........G..C..T.....A......T.AC.......A..C [468]
Aph._ovalisporum_031G(033C)      ........T..........C....TAC.C..G..C........G..C..T.....A......T.AC.......A..C [468]
Aph._ovalisporum_033B(035E)      ........T..........C....TAC.C..G..C........G..C..T.....A......T.AC.......A..C [468]
Aph._ovalisporum_033E            ........T..........C....TAC.C..G..C........G..C..T.....A......T.AC.......A..C [468]
```

```
Aph._ovalisporum_035C            ........T..........C....TAC.C..G..C........G..C..T.....A......T.AC.......A..C [468]
Aph._ovalisporum_035D            ........T..........C....TAC.C..G..C........G..C..T.....A......T.AC.......A..C [468]
Aph._ovalisporum_035F            ........T..........C....TAC.C..G..C........G..C..T.....A......T.AC.......A..C [468]
C.raciborskii_AWT205(023E_024C_0 ..TAGG..T..............A.A..AGT...G...........A..T.....A.....TT.AC.A..T..AT.A [468]
C.raciborskii_030A(023D_032A_034 ..TAGG..T..............A..AGT...G...........A..T.....A.....TT.AC.A..T..AT.A [468]
Aph_sp._22D11                    ..TGCT.....C...........C.....AG.G.....G.....G..C..T..G.....G..T........G...T.. [468]
Aph_sp._10E6                     ..TGCT.....C...........C.....AG.G..A..G.....G..C..T..G.....G..T........G...T.. [468]
Aph_sp._10E9                     ..TGCT.....C...........C.....AG.G..A..G.....G..C..T..G.....G..T........G...T.. [468]
Ana.lapponica_966                ..TG.C..T.....C........C..AC.CAC...G..G.....G..A..T.....A............A.A...T.. [468]
Aph._ovalisporumILC-146          ........T..........C....TAC.C..G..C........G..C..T.....A......T.AC.......A..C [468]

Aph.aphanizomenoides_259C        GCTGATATTAATTTAGAGCAAGAAGCAGAAAGCCTCCGGGAAGAAATTGGTAGTGCTAAAGGACAAAAACGAGCAAAA [546]
Ana._bergii_283A                 ..C.......G..............C..G.AT.............CC.C..A...C..G.............C... [546]
Ana.berg_366B                    ..C.......G..............C..G.AT.............CC.C..A...C..G.............C... [546]
Aph.issatschenkoi_CAWBG02        ..G..............A........T...TCT...A.A..........CC.AC.....G..G........C..C... [546]
Aph._ovalisporum_031E            ..C.......G..............C..G.AT.............CC.C..A...C..G.............C... [546]
Aph._ovalisporum_031G(033C)      ..C.......G..............C..G.AT.............CC.C..A...C..G.............C... [546]
Aph._ovalisporum_033B(035E)      ..C.......G..............C..G.AT.............CC.C..A...C..G.............C... [546]
Aph._ovalisporum_033E            ..C.......G..............C..G.AT.............CC.C..A...C..G.............C... [546]
Aph._ovalisporum_035C            ..C.......G..............C..G.AT.............CC.C..A...C..G.............C... [546]
Aph._ovalisporum_035D            ..C.......G..............C..G.AT.............CC.C..A...C..G.............C... [546]
```

```
Aph._ovalisporum_035F           ..C.......G..............C..G.AT.............CC.C..A...C..G.............C... [546]
C.raciborskii_AWT205(023E_024C_0 T.A.G............AG............AA..G....GG.......AAGCA..A..G...........G..T... [546]
C.raciborskii_030A(023D_032A_034 T.A.G............AG............AA..G....GG.......AAGCA..A..G...........G..T... [546]
Aph_sp._22D11                   ..G...........G.....G..G..G........G.............AA.AA..C..G.....G.....G..G..G [546]
Aph_sp._10E6                    ..G...........G.....G..G..G........G.............AG.AA..C..G.....G.....G..G..G [546]
Aph_sp._10E9                    ..G...........G.....G..G..G........G.............AG.AA..C..G.....G.....G..G..G [546]
Ana.lapponica_966               ......C.G........A........C....AT..G..C.........AA..A......G.................. [546]
Aph._ovalisporumILC-146         ..C.......G..............C..G.AT.............CC.C..A...C..G.............C... [546]

Aph.aphanizomenoides_259C       CTCATTAAACGCCTGCGGGTA [567]
Ana._bergii_283A                T.A........A..CA.A... [567]
Ana.berg_366B                   T.A........A...A.A... [567]
Aph.issatschenkoi_CAWBG02       T.A........G......... [567]
Aph._ovalisporum_031E           T.A........A..TA.A... [567]
Aph._ovalisporum_031G(033C)     T.A........A..CA.A... [567]
Aph._ovalisporum_033B(035E)     T.A........A..CA.A... [567]
Aph._ovalisporum_033E           T.A........A..CA.A... [567]
Aph._ovalisporum_035C           T.A........A..CA.A... [567]
Aph._ovalisporum_035D           T.A........A..CA.A... [567]
Aph._ovalisporum_035F           T.A........A..CA.A... [567]
C.raciborskii_AWT205(023E_024C_0 T.G........T......... [567]
```

```
C.raciborskii_030A(023D_032A_034 T.G........T......... [567]
Aph_sp._22D11                    T.A........G........G [567]
Aph_sp._10E6                     T.A........G........G [567]
Aph_sp._10E9                     T.A........G........G [567]
Ana.lapponica_966                T.A.........T........ [567]
Aph._ovalisporumILC-146          T.A........A..TA.A... [567]
;
```

Appendix 7 Multiple Alignment of *cynC*

```
C._raci_AWT205_(023E_024C_026J_0 AACCGCCAAACTCAAAGACATTGCGAAAGGGAATTGGATAGACCTCAAAGCCGTGTTTATCTAGCAACTCATGAAGCG [ 78]
Ana_aphanizomenoides_259C        ............................................................................. [ 78]
Ana_bergii_283A                  ....A.....................G...........G..................................... [ 78]
Ana_bergii_366B                  ....A.....................G...........G..................................... [ 78]
Aph._ovalisporum_31E             ....A.....................G...........G..................................... [ 78]
Aph._ovalisporum_31G             ....A.....................G...........G..................................... [ 78]
Aph._ovalisporum_33B             ....A.....................G...........G..................................... [ 78]
Aph._issatschenkoi_CAWG02        ............................................................................. [ 78]
Aph._ovalisporum_ILC-146         ....A.....................G...........G..................................... [ 78]
Ana._lapponica_966               .T..A.....T........G.....G........C..G.........................A........C.... [ 78]
Aph._sp._22D11                   ....................................A........................................ [ 78]
Aph._sp._10E6                    ....................................A........................................ [ 78]
Aph._sp._10E9                    ....................................A........................................ [ 78]
C._raci._023D_(030A_032A_034)    ............................................................................. [ 78]
Aph._ovalisporum_35C             ....A.....................G...........G..................................... [ 78]
Aph._ovalisporum_33E             ....A.....................G...........G..................................... [ 78]
Aph._ovalisporum_35D             ....A.....................G...........G..................................... [ 78]
Aph._ovalisporum_35E             ....A.....................G...........G..................................... [ 78]
Aph._ovalisporum_35F             ....A.....................G...........G..................................... [ 78]
```

```
C._raci_AWT205_(023E_024C_026J_0 GATGCTCTTGCTCTTCACAGATGACCTTCTTGGGGGAAATGCTTAACACATTCATCGCCAACCACTTACTGGACTGGC [156]
Ana_aphanizomenoides_259C        ............................................................................ [156]
Ana_bergii_283A                  ...........................................................A................ [156]
Ana_bergii_366B                  ...........................................................A................ [156]
Aph._ovalisporum_31E             ...........................................................A................ [156]
Aph._ovalisporum_31G             ...........................................................A................ [156]
Aph._ovalisporum_33B             ...........................................................A................ [156]
Aph._issatschenkoi_CAWG02        ............................................................................ [156]
Aph._ovalisporum_ILC-146         ...........................................................A................ [156]
Ana._lapponica_966               C.........T.................A........A.....................T............. [156]
Aph._sp._22D11                   ............................................................................ [156]
Aph._sp._10E6                    ............................................................................ [156]
Aph._sp._10E9                    ............................................................................ [156]
C._raci._023D_(030A_032A_034)    ............................................................................ [156]
Aph._ovalisporum_35C             ...........................................................A................ [156]
Aph._ovalisporum_33E             ...........................................................A................ [156]
Aph._ovalisporum_35D             ...........................................................A................ [156]
Aph._ovalisporum_35E             ...........................................................A................ [156]
Aph._ovalisporum_35F             ...........................................................A................ [156]
```

```
C._raci_AWT205_(023E_024C_026J_0 AGAACATAGGCATCTCATCGTCGGTGGAAGTGGGGAGGGGTGCTTCGATGAATTGCCAACCGTTATCCATGAAGAGTT [234]
Ana_aphanizomenoides_259C       ............................................................................ [234]
Ana_bergii_283A                 .......G..........A.............A..A.......T.............T.A.............. [234]
Ana_bergii_366B                 .......G..........A.............A..A.......T.............T.A....T.......... [234]
Aph._ovalisporum_31E            .......G..........A.............A..A.......T.............T.A....T.......... [234]
Aph._ovalisporum_31G            .......G..........A.............A..A.......T.............T.A.............. [234]
Aph._ovalisporum_33B            .......G..........A.............A..A.......T.............T.A.............. [234]
Aph._issatschenkoi_CAWG02       .......G.................................................................... [234]
Aph._ovalisporum_ILC-146        .......G..........A.............A..A.......T.............T.A....T.......... [234]
Ana._lapponica_966              ......GG........................T.....................T....G....G........ [234]
Aph._sp._22D11                  ............................................................................ [234]
Aph._sp._10E6                   ............................................................................ [234]
Aph._sp._10E9                   ............................................................................ [234]
C._raci._023D_(030A_032A_034)   ............................................................................ [234]
Aph._ovalisporum_35C            .......G..........A.............A..A.......T.............T.A.............. [234]
Aph._ovalisporum_33E            .......G..........A.............A..A.......T.............T.A.............. [234]
Aph._ovalisporum_35D            .......G..........A.............A..A.......T.............T.A.............. [234]
Aph._ovalisporum_35E            .......G..........A.............A..A.......T.............T.A.............. [234]
Aph._ovalisporum_35F            .......G..........A.............A..A.......T.............T.A.............. [234]

C._raci_AWT205_(023E_024C_026J_0 TCTCTTCACCCTCTTTGATGGGGCGATCTGGATTCACTAACACAACCCCAGGTGCTAAGGGGACAAAAGTACAATCAA [312]
```

```
Ana_aphanizomenoides_259C       ..................................................................... [312]
Ana_bergii_283A                 ..............G...A.....C...........C.G..............C..................... [312]
Ana_bergii_366B                 ..............G...A.....C...........C.G..............C..................... [312]
Aph._ovalisporum_31E            ..............G...A.....C...........C.G..............C..................... [312]
Aph._ovalisporum_31G            ..............G...A.....C...........C.G..............C..................... [312]
Aph._ovalisporum_33B            ..............G...A.....C...........C.G..............C..................... [312]
Aph._issatschenkoi_CAWG02       ..............G...................................................... [312]
Aph._ovalisporum_ILC-146        ..............G...A.....C...........C.G..............C..................... [312]
Ana._lapponica_966              ........................C.............G......G...........C........G..G....... [312]
Aph._sp._22D11                  ..................................................................... [312]
Aph._sp._10E6                   ..................................................................... [312]
Aph._sp._10E9                   ..................................................................... [312]
C._raci._023D_(030A_032A_034)   ..................................................................... [312]
Aph._ovalisporum_35C            ..............G...A.....C...........C.G..............C..................... [312]
Aph._ovalisporum_33E            ..............G...A.....C...........C.G..............C..................... [312]
Aph._ovalisporum_35D            ..............G...A.....C...........C.G..............C..................... [312]
Aph._ovalisporum_35E            ..............G...A.....C...........C.G..............C..................... [312]
Aph._ovalisporum_35F            ..............G...A.....C...........C.G..............C..................... [312]


C._raci_AWT205_(023E_024C_026J_0 TGTGGGATGGGAAAATATCTAGTGGGAAGTGAATATCATGCACGCGGAAGCGACGCGGCTCTAAATGCCGTTTGAGCC [390]
Ana_aphanizomenoides_259C       ..................................................................... [390]
```

```
Ana_bergii_283A                 ..........A.....................C.................C...T.....C.............. [390]
Ana_bergii_366B                 ..........A.....................C.................C...T.....C.............. [390]
Aph._ovalisporum_31E            ..........A.....................C.................C...T.....C.............. [390]
Aph._ovalisporum_31G            ..........A.....................C.................C...T.....C.............. [390]
Aph._ovalisporum_33B            ..........A.....................C.................C...T.....C.............. [390]
Aph._issatschenkoi_CAWG02       ........................................................................... [390]
Aph._ovalisporum_ILC-146        ..........A.....................C.................C...T.....C.............. [390]
Ana._lapponica_966              ..........A....G.........A.......C......T..........C...T.....C.............. [390]
Aph._sp._22D11                  ...................................................C........................ [390]
Aph._sp._10E6                   ...................................................C........................ [390]
Aph._sp._10E9                   ...................................................C........................ [390]
C._raci._023D_(030A_032A_034)   ........................................................................... [390]
Aph._ovalisporum_35C            ..........A.....................C.................C...T.....C.............. [390]
Aph._ovalisporum_33E            ..........A.....................C.................C...T.....C.............. [390]
Aph._ovalisporum_35D            ..........A.....................C.................C...T.....C.............. [390]
Aph._ovalisporum_35E            ..........A.....................C.................C...T.....C.............. [390]
Aph._ovalisporum_35F            ..........A.....................C.................C...T.....C.............. [390]

C._raci_AWT205_(023E_024C_026J_0 AGCGAATCCCTGCACGATTAGTCGTCATTGACTCCTGCACAAAAATATCACGGCCAAAGCGGCTACAGTCTGCTGCGT [468]
Ana_aphanizomenoides_259C       ........................................................................... [468]
Ana_bergii_283A                 ............A...........................................................C..A. [468]
```

```
Ana_bergii_366B                 ............A.....................................................C..A. [468]
Aph._ovalisporum_31E            ............A.....................................................C..A. [468]
Aph._ovalisporum_31G            ............A.....................................................C..A. [468]
Aph._ovalisporum_33B            ............A.....................................................C..A. [468]
Aph._issatschenkoi_CAWG02       ..................................................................... [468]
Aph._ovalisporum_ILC-146        ............A.....................................................C..A. [468]
Ana._lapponica_966              ..T....T...........G........G..............G.CG.....C.................C..A. [468]
Aph._sp._22D11                  .....................................................................A. [468]
Aph._sp._10E6                   .....................................................................A. [468]
Aph._sp._10E9                   .....................................................................A. [468]
C._raci._023D_(030A_032A_034)   ..................................................................... [468]
Aph._ovalisporum_35C            ............A.....................................................C..A. [468]
Aph._ovalisporum_33E            ............A.....................................................C..A. [468]
Aph._ovalisporum_35D            ............A.....................................................C..A. [468]
Aph._ovalisporum_35E            ............A.....................................................C..A. [468]
Aph._ovalisporum_35F            ............A.....................................................C..A. [468]

C._raci_AWT205_(023E_024C_026J_0 CAAAAATCACCTCATCCTGGGTGACGCAGAACTCAAAATCATGCATACTCTCAAATCGATCTTCCATCGGACACTCCC [546]
Ana_aphanizomenoides_259C       ..................................................................... [546]
Ana_bergii_283A                 ....................A.........................G.....................C....... [546]
Ana_bergii_366B                 ....................A.........................G.....................C....... [546]
```

```
Aph._ovalisporum_31E            .....................A.....................G.................C..... [546]
Aph._ovalisporum_31G            .....................A.....................G.................C..... [546]
Aph._ovalisporum_33B            .....................A.....................G.................C..... [546]
Aph._issatschenkoi_CAWG02       ........................................................................ [546]
Aph._ovalisporum_ILC-146        .....................A.....................G.................C..... [546]
Ana._lapponica_966              .......T.............T.T.....T.......T....G.CTG..G...G..C.T...T.AT..C..... [546]
Aph._sp._22D11                  ..............................................G.................A.C..... [546]
Aph._sp._10E6                   ..............................................G.................A.C..... [546]
Aph._sp._10E9                   ..............................................G.................A.C..... [546]
C._raci._023D_(030A_032A_034)   ........................................................................ [546]
Aph._ovalisporum_35C            .....................A.....................G.................C..... [546]
Aph._ovalisporum_33E            .....................A.....................G.................C..... [546]
Aph._ovalisporum_35D            .....................A.....................G.................C..... [546]
Aph._ovalisporum_35E            .....................A.....................G.................C..... [546]
Aph._ovalisporum_35F            .....................A.....................G.................C..... [546]

C._raci_AWT205_(023E_024C_026J_0 AGAAATCTTCGCGGTACATGGCATTTTGCATAGTCGGTTTAGGCGCAGCATTCCAGATCATATCTGGATCTTTATGCC [624]
Ana_aphanizomenoides_259C       ........................................................................ [624]
Ana_bergii_283A                 ..........A.............................................G.........A...G.... [624]
Ana_bergii_366B                 ..........A.....................................................CA...G.... [624]
Aph._ovalisporum_31E            ..........A.....................................................CA...G.... [624]
```

```
Aph._ovalisporum_31G            ..........A.....................................................G.........A...G.... [624]
Aph._ovalisporum_33B            ..........A.....................................................G.........A...G.... [624]
Aph._issatschenkoi_CAWG02       ..................................................................................... [624]
Aph._ovalisporum_ILC-146        ..........A.....................................................CA...G.... [624]
Ana._lapponica_966              ....C....G.............C......G........G......CT.C............G......G..G.... [624]
Aph._sp._22D11                  ..............................................T.......................... [624]
Aph._sp._10E6                   ..............................................T.......................... [624]
Aph._sp._10E9                   ..............................................T.......................... [624]
C._raci._023D_(030A_032A_034)   ..................................................................................... [624]
Aph._ovalisporum_35C            ..........A.....................................................G.........A...G.... [624]
Aph._ovalisporum_33E            ..........A.....................................................G.........A...G.... [624]
Aph._ovalisporum_35D            ..........A.....................................................G.........A...G.... [624]
Aph._ovalisporum_35E            ..........A.....................................................G.........A...G.... [624]
Aph._ovalisporum_35F            ..........A.....................................................G.........A...G.... [624]


C._raci_AWT205_(023E_024C_026J_0 AATATTCATAGACTAGTTTGCGATAGGGTAAATACTCAAAGAAGC [669]
Ana_aphanizomenoides_259C       ..................................... [669]
Ana_bergii_283A                 ..............C..C..................... [669]
Ana_bergii_366B                 ..............C..C..................... [669]
Aph._ovalisporum_31E            ..............C..C..................... [669]
Aph._ovalisporum_31G            ..............C..C..................... [669]
```

```
Aph._ovalisporum_33B            .................C..C......................... [669]

Aph._issatschenkoi_CAWG02       .................C............................ [669]

Aph._ovalisporum_ILC-146        .................C..C......................... [669]

Ana._lapponica_966              .......G.....C....G......A..C.G.............. [669]

Aph._sp._22D11                  ............................................... [669]

Aph._sp._10E6                   ............................................... [669]

Aph._sp._10E9                   ............................................... [669]

C._raci._023D_(030A_032A_034)   ............................................... [669]

Aph._ovalisporum_35C            .................C..C......................... [669]

Aph._ovalisporum_33E            .................C..C......................... [669]

Aph._ovalisporum_35D            .................C..C......................... [669]

Aph._ovalisporum_35E            .................C..C......................... [669]

Aph._ovalisporum_35F            .................C..C......................... [669]

;

end;
```

Appendix 8 Multiple Alignment of *cynD*

```
Aph._aphanizomeniodes_259C       CAAGATGGCACAAATAGCATCAATCGCTTGCTGTTGCCCAGAAATTACAACACTTTGTGGGCCATTAATCGCGGCGAT [ 78]
Ana._bergii_283A                 ..............................................................................  [ 78]
Ana._bergii_366B                 ........T.G...................................................................  [ 78]
Aph._issatschenkoi_CAWBGO2       ..............................................................................  [ 78]
Aph._ovalisporum_ILC-146         ........T.G...................................................................  [ 78]
Aph._ovalisporum_031E            ........T.G...................................................................  [ 78]
Aph._ovalisporum_031G_(033C)     ..............................................................................  [ 78]
Aph._ovalisporum_033B_(035E)     ..............................................................................  [ 78]
Aph._ovalisporum_033E            ..............................................................................  [ 78]
Aph._ovalisporum_035C            ..............................................................................  [ 78]
Aph_ovalisporum_035D             ..............................................................................  [ 78]
Aph._ovalisporum_035F            ..............................................................................  [ 78]
C.raciborskii_AWT205(23E_24C_26J ..............................................................................  [ 78]
C.raciborskii_023D_(030A_032A_03 ..............................................................................  [ 78]
Aph._sp_10E6                     ...............................................................T..............  [ 78]
Aph._sp_10E9                     ...............................................................T..............  [ 78]
Aph._sp_22D11                    ...............................................................T..............  [ 78]


Aph._aphanizomeniodes_259C       CCCTACCTGTCCCCCATAAGGTAAAATGATATC------TGCAATACGCTTCTCGCTTGCCATCACGGAAACCATGCT [156]
```

```
Ana._bergii_283A                .G.............................------..................................... [156]
Ana._bergii_366B                .G.............................------..................................... [156]
Aph._issatschenkoi_CAWBGO2      .G.............................------..................................... [156]
Aph._ovalisporum_ILC-146        .G.............................GAATTC..................................... [156]
Aph._ovalisporum_031E           .G.............................------..................................... [156]
Aph._ovalisporum_031G_(033C)    .G.............................------..................................... [156]
Aph._ovalisporum_033B_(035E)    .G.............................------..................................... [156]
Aph._ovalisporum_033E           .G.............................------..................................... [156]
Aph._ovalisporum_035C           .G.............................------..................................... [156]
Aph_ovalisporum_035D            .G.............................------..................................... [156]
Aph._ovalisporum_035F           .G.............................------..................................... [156]
C.raciborskii_AWT205(23E_24C_26J ..............................------..................................... [156]
C.raciborskii_023D_(030A_032A_03 ..............................------..................................... [156]
Aph._sp_10E6                    ..............................------..................................... [156]
Aph._sp_10E9                    ..............................------..................................... [156]
Aph._sp_22D11                   ..............................------..................................... [156]

Aph._aphanizomeniodes_259C      CCCATCACGAGGAAGTGCCTGTATCAGTCGTCCTCGTTCTGCAATCAGTTTCAACCCATCCTCTAAGCTAAAGACTCC [234]
Ana._bergii_283A                ...................................................G...................... [234]
Ana._bergii_366B                ...................................................G.............T.......... [234]
Aph._issatschenkoi_CAWBGO2      ...................................................G...................... [234]
```

```
Aph._ovalisporum_ILC-146          ..................................................G..................... [234]
Aph._ovalisporum_031E             ..................................................G..................... [234]
Aph._ovalisporum_031G_(033C)      ..................................................G..................... [234]
Aph._ovalisporum_033B_(035E)      ..................................................G..................... [234]
Aph._ovalisporum_033E             ..................................................G..................... [234]
Aph._ovalisporum_035C             ..................................................G..................... [234]
Aph_ovalisporum_035D              ..................................................G..................... [234]
Aph._ovalisporum_035F             ..................................................G..................... [234]
C.raciborskii_AWT205(23E_24C_26J  ..................................................................... [234]
C.raciborskii_023D_(030A_032A_03  ..................................................................... [234]
Aph._sp_10E6                      ..................................................................... [234]
Aph._sp_10E9                      ..................................................................... [234]
Aph._sp_22D11                     ..................................................................... [234]

Aph._aphanizomeniodes_259C        CGCCACACAAGCGGACACATATTCACCTACACTATGACCTATCACCACTGAGGGCTGGATGCCCCACGACTTCCATAA [312]
Ana._bergii_283A                  ..............A...........G.....................................G..T.......G [312]
Ana._bergii_366B                  ..............A...........G.................C.................G..T.......G [312]
Aph._issatschenkoi_CAWBGO2        ..............A...........G.....................................G..T........ [312]
Aph._ovalisporum_ILC-146          ..............A...........G.................C.................G..T.......G [312]
Aph._ovalisporum_031E             ..............A...........G.................C.................G..T.......G [312]
Aph._ovalisporum_031G_(033C)      ..............A...........G.....................................G..T........ [312]
```

```
Aph._ovalisporum_033B_(035E)     ..............A..........G.....................................G..T.......G [312]
Aph._ovalisporum_033E            ..............A..........G.................C.................G..T.......G [312]
Aph._ovalisporum_035C            ..............A..........G.................C.................G..T.......G [312]
Aph_ovalisporum_035D             ..............A..........G.....................................G..T.......G [312]
Aph._ovalisporum_035F            ..............A..........G.................C.................G..T.......G [312]
C.raciborskii_AWT205(23E_24C_26J ..................................................................... [312]
C.raciborskii_023D_(030A_032A_03 ..................................................................... [312]
Aph._sp_10E6                     ..................................................................... [312]
Aph._sp_10E9                     ..................................................................... [312]
Aph._sp_22D11                    ..................................................................... [312]

Aph._aphanizomeniodes_259C       TTGATAGAGCGCGTATTCGATGGCAAAGAGTGCGGGCTGGGTATAGGCTGTGTCATCTAATGAGAGTCCATGCTCAGA [390]
Ana._bergii_283A                 ...................A.........................................A..... [390]
Ana._bergii_366B                 ...................A.........................................A..... [390]
Aph._issatschenkoi_CAWBGO2       ...................A.........................................A..... [390]
Aph._ovalisporum_ILC-146         ...................A.........................................A..... [390]
Aph._ovalisporum_031E            ...................A.........................................A..... [390]
Aph._ovalisporum_031G_(033C)     ...................A.........................................A..... [390]
Aph._ovalisporum_033B_(035E)     ...................A.........................................A..... [390]
Aph._ovalisporum_033E            ...................A.........................................A..... [390]
Aph._ovalisporum_035C            ...................A.........................................A..... [390]
```

```
Aph_ovalisporum_035D            ....................A.........................................A.............  [390]
Aph._ovalisporum_035F           ....................A.........................................A.............  [390]
C.raciborskii_AWT205(23E_24C_26J ..................................................................................  [390]
C.raciborskii_023D_(030A_032A_03 ..................................................................................  [390]
Aph._sp_10E6                    ..................................................................................  [390]
Aph._sp_10E9                    ..................................................................................  [390]
Aph._sp_22D11                   ..................................................................................  [390]

Aph._aphanizomeniodes_259C      TGAACCAAAAATAACGGAGAGAATGGACTCGATCCCTAAATTTTGGAGAATGGTTTCACATCGGTCAAGGGCTGCACG  [468]
Ana._bergii_283A                ............G.....................................................................  [468]
Ana._bergii_366B                ............G.....................................................................  [468]
Aph._issatschenkoi_CAWBGO2      ............G.......C.............................................................  [468]
Aph._ovalisporum_ILC-146        ............G.....................................................................  [468]
Aph._ovalisporum_031E           ............G.....................................................................  [468]
Aph._ovalisporum_031G_(033C)    ............G.......C.............................................................  [468]
Aph._ovalisporum_033B_(035E)    ............G.....................................................................  [468]
Aph._ovalisporum_033E           ............G.....................................................................  [468]
Aph._ovalisporum_035C           ............G.....................................................................  [468]
Aph_ovalisporum_035D            ............G.....................................................................  [468]
Aph._ovalisporum_035F           ............G.....................................................................  [468]
C.raciborskii_AWT205(23E_24C_26J ..................................................................................  [468]
```

```
C.raciborskii_023D_(030A_032A_03 ..................................................................... [468]
Aph._sp_10E6                     ..................................................................... [468]
Aph._sp_10E9                     ..................................................................... [468]
Aph._sp_22D11                    ..................................................................... [468]

Aph._aphanizomeniodes_259C       GAATGTTGATTCAGTATCGTAGAGGGTGCGCCCCATGTTAATGTATTGTGAGCCTTGACCTGTAAACAGGAATGCGAT [546]
Ana._bergii_283A                 ...................................................................G..... [546]
Ana._bergii_366B                 ...................................................................G..... [546]
Aph._issatschenkoi_CAWBGO2       ...................................................................G..A... [546]
Aph._ovalisporum_ILC-146         ...................................................................G..... [546]
Aph._ovalisporum_031E            ...................................................................G..... [546]
Aph._ovalisporum_031G_(033C)     ...................................................................G..A... [546]
Aph._ovalisporum_033B_(035E)     ...................................................................G..... [546]
Aph._ovalisporum_033E            ...................................................................G..... [546]
Aph._ovalisporum_035C            ...................................................................G..... [546]
Aph_ovalisporum_035D             ...................................................................G..... [546]
Aph._ovalisporum_035F            ...................................................................G..... [546]
C.raciborskii_AWT205(23E_24C_26J ..................................................................... [546]
C.raciborskii_023D_(030A_032A_03 ..................................................................... [546]
Aph._sp_10E6                     ..................................................................... [546]
Aph._sp_10E9                     ..................................................................... [546]
```

```
Aph._sp_22D11                  ..................................................................................... [546]


Aph._aphanizomeniodes_259C     CGTCGGCGGGGCATTTTTTTTCTGTGCTGAAA [578]
Ana._bergii_283A               ............................... [578]
Ana._bergii_366B               ............................... [578]
Aph._issatschenkoi_CAWBGO2     ............................... [578]
Aph._ovalisporum_ILC-146       ............................... [578]
Aph._ovalisporum_031E          ............................... [578]
Aph._ovalisporum_031G_(033C)   ............................... [578]
Aph._ovalisporum_033B_(035E)   ............................... [578]
Aph._ovalisporum_033E          ............................... [578]
Aph._ovalisporum_035C          ............................... [578]
Aph_ovalisporum_035D           ............................... [578]
Aph._ovalisporum_035F          ............................... [578]
C.raciborskii_AWT205(23E_24C_26J ............................. [578]
C.raciborskii_023D_(030A_032A_03 ............................. [578]
Aph._sp_10E6                   ............................... [578]
Aph._sp_10E9                   ............................... [578]
Aph._sp_22D11                  ............................... [578]
```

Appendix 9 Multiple Alignment of *cynA*

```
Aph.aphanizomenoides_259C       GCGCGATTTATTGGCCATAACTGAAACCATCGCCCCATCGGGAGGCAATTCTTGCATCAGACGGCCTCTTTCAGCAAT [ 78]
Ana._bergii_283A                ............A.............................................................. [ 78]
Ana._bergii_366B                ............A.............................................................. [ 78]
Aph._issatschenkoi_CAWBG02      .............................................................................. [ 78]
Aph._ovalisporum_031E           ............A..............................................C............... [ 78]
Aph._ovalisporum_031G_(033C)    ............A.............................................................. [ 78]
Aph._ovalisporum_033B_(035E)    ............A.............................................................. [ 78]
Aph._ovalisporum_033E           ............A.............................................................. [ 78]
Aph._ovalisporum_035C           gcg.........A.............................................................. [ 78]
Aph._ovalisporum_035D           ............A.............................................................. [ 78]
Aph._ovalisporum_035F           ............A.............................................................. [ 78]
C._raciborskii_AWT205_(23E_24C_2 .............................................................................. [ 78]
C._raciborskii_023D_(030A_032A_0 gcg........................................................................... [ 78]
Aph._sp.10E6                    .............................................................................. [ 78]

Aph._aphanizomenoides_259C      TAGTTTGAGTCCATCCTCTAAAGAAAAGACACCCGCCACACAAGCGGCCACATATTCTCCCACACTATGCCCCATCAC [156]
Ana._bergii_283A                .........A.................................................................. [156]
Ana._bergii_366B                .........A.................................................................. [156]
Aph._issatschenkoi_CAWBG02      .............................................................................. [156]
```

```
Aph._ovalisporum_031E            .........A.......................................................................... [156]
Aph._ovalisporum_031G_(033C)     .........A.......................................................................... [156]
Aph._ovalisporum_033B_(035E)     .........A.......................................................................... [156]
Aph._ovalisporum_033E            .........A.......................................................................... [156]
Aph._ovalisporum_035C            .........A.......................................................................... [156]
Aph._ovalisporum_035D            .........A.......................................................................... [156]
Aph._ovalisporum_035F            .........A.......................................................................... [156]
C._raciborskii_AWT205_(23E_24C_2 .......................................................................... [156]
C._raciborskii_023D_(030A_032A_0 .......................................................................... [156]
Aph._sp.10E6                     .........G.......................................................................... [156]

Aph._aphanizomenoides_259C       CACATCAGGCTCAATACCCCAGGAGCGCCACAACTGCGCCAGTGCATATTCAAGAGAAAAAAGAGTGGGTTGAGTATA [234]
Ana._bergii_283A                 ............................T.....................C.............CC........... [234]
Ana._bergii_366B                 ....................................................C.............C........... [234]
Aph._issatschenkoi_CAWBG02       .......................................................................... [234]
Aph._ovalisporum_031E            ....................................................C.............C........... [234]
Aph._ovalisporum_031G_(033C)     ............................T.....................C.............CC........... [234]
Aph._ovalisporum_033B_(035E)     ............................T.....................C.............CC........... [234]
Aph._ovalisporum_033E            ....................................................C.............C........... [234]
Aph._ovalisporum_035C            ....................................................C.............C........... [234]
Aph._ovalisporum_035D            ............................T.....................C.............CC........... [234]
```

```
Aph._ovalisporum_035F            ..............................T.....................C.............CC.......... [234]
C._raciborskii_AWT205_(23E_24C_2 ..................................................................................... [234]
C._raciborskii_023D_(030A_032A_0 ..................................................................................... [234]
Aph._sp.10E6                     ..........................................................G.......C.............. [234]

Aph._aphanizomenoides_259C       GGCGGTTTCATCAATTTGGGATGGCGTTTCCAATCCCATTTGTTGGCTGGGATAGAGTATTGAGAGGATTGATTTGCC [312]
Ana._bergii_283A                 ..................................................................................... [312]
Ana._bergii_366B                 ..................................................................................... [312]
Aph._issatschenkoi_CAWBG02       ..................................................................................... [312]
Aph._ovalisporum_031E            ..................................................................................... [312]
Aph._ovalisporum_031G_(033C)     ..................................................................................... [312]
Aph._ovalisporum_033B_(035E)     ..................................................................................... [312]
Aph._ovalisporum_033E            ..................................................................................... [312]
Aph._ovalisporum_035C            ..................................................................................... [312]
Aph._ovalisporum_035D            ..................................................................................... [312]
Aph._ovalisporum_035F            ..................................................................................... [312]
C._raciborskii_AWT205_(23E_24C_2 ..................................................................................... [312]
C._raciborskii_023D_(030A_032A_0 ..................................................................................... [312]
Aph._sp.10E6                     A.................................................................................... [312]

Aph._aphanizomenoides_259C       CAACAGTGAACGAAGAATCTCATCACATCGGTCAATGGTTTGCCGGAAGGTGGGTTGGCTCTCATAAAGTTCTTGCCC [390]
```

```
Ana._bergii_283A                ...........G.............................A.............................. [390]
Ana._bergii_366B                ...........G.............................A.............................. [390]
Aph._issatschenkoi_CAWBG02      ..................................................................... [390]
Aph._ovalisporum_031E           ...........G.............................A.............................. [390]
Aph._ovalisporum_031G_(033C)    ...........G.............................A.............................. [390]
Aph._ovalisporum_033B_(035E)    ...........G.............................A.............................. [390]
Aph._ovalisporum_033E           ...........G.............................A.............................. [390]
Aph._ovalisporum_035C           ...........G.............................A.............................. [390]
Aph._ovalisporum_035D           ...........G.............................A.............................. [390]
Aph._ovalisporum_035F           ...........G.............................A.............................. [390]
C._raciborskii_AWT205_(23E_24C_2 ..................................................................... [390]
C._raciborskii_023D_(030A_032A_0 ..................................................................... [390]
Aph._sp.10E6                    ..................................................................... [390]

Aph._aphanizomenoides_259C      CATCCCCACATACTGTGAGCCTTGACCTGTAAAGAGAAAAGCTATTTTTGAGAGAGATGTCCTCTTCCCCTGCCCCTC [468]
Ana._bergii_283A                ............................................................G............ [468]
Ana._bergii_366B                ............................................................G............ [468]
Aph._issatschenkoi_CAWBG02      ..................................................................... [468]
Aph._ovalisporum_031E           ............................................................G............ [468]
Aph._ovalisporum_031G_(033C)    ............................................................G............ [468]
Aph._ovalisporum_033B_(035E)    ............................................................G............ [468]
```

```
Aph._ovalisporum_033E            ..............................................................G.............. [468]
Aph._ovalisporum_035C            ..............................................................G.............. [468]
Aph._ovalisporum_035D            ..............................................................G.............. [468]
Aph._ovalisporum_035F            ..............................................................G.............. [468]
C._raciborskii_AWT205_(23E_24C_2 ................................................................................ [468]
C._raciborskii_023D_(030A_032A_0 ................................................................................ [468]
Aph._sp.10E6                     ................................................................................ [468]

Aph._aphanizomenoides_259C       TCCC-----------------------------GATTGGGCAAACGTTTCCAATTGTTGGCGCAACTGGGTTTTAGA [546]
Ana._bergii_283A                 ....CATTGGGCAAAGGTTTCCAATTGTTGTCCC...........G.............................. [546]
Ana._bergii_366B                 ....CATTGGGCAAACGTTTCCAATTGTTGTCCC.............................................. [546]
Aph._issatschenkoi_CAWBG02       ....-----------------------------.............................................. [546]
Aph._ovalisporum_031E            ....CATTGGGCAAACGTTTCCAATTGTTGTCCC.............................................. [546]
Aph._ovalisporum_031G_(033C)     ....CATTGGGCAAAGGTTTCCAATTGTTGTCCC...........G.............................. [546]
Aph._ovalisporum_033B_(035E)     ....CATTGGGCAAAGGTTTCCAATTGTTGTCCC...........G.............................. [546]
Aph._ovalisporum_033E            ....CATTGGGCAAAGGTTTCCAATTGTTGTCCC...........G.............................. [546]
Aph._ovalisporum_035C            ....CATTGGGCAAAGGTTTCCAATTGTTGTCCC...........G.............................. [546]
Aph._ovalisporum_035D            ....CATTGGGCAAAGGTTTCCAATTGTTGTCCC...........G.............................. [546]
Aph._ovalisporum_035F            ....CATTGGGCAAAGGTTTCCAATTGTTGTCCC...........G.............................. [546]
C._raciborskii_AWT205_(23E_24C_2 ....-----------------------------.............................................. [546]
C._raciborskii_023D_(030A_032A_0 ....-----------------------------.............................................. [546]
```

```
Aph._sp.10E6                    ....------------------------------..................................... [546]


Aph._aphanizomenoides_259C      CTCTGCTACTACTGCAAAGCGATGTCTAAAGTGTTTGCGCCCTGTGTTGGCTGTAAAGCAAATATCCGCCAGACTCTC [624]
Ana._bergii_283A                ........................T.............................................. [624]
Ana._bergii_366B                ........................T.............................................. [624]
Aph._issatschenkoi_CAWBG02      ..................................................................... [624]
Aph._ovalisporum_031E           ........................T.............................................. [624]
Aph._ovalisporum_031G_(033C)    ........................T.............................................. [624]
Aph._ovalisporum_033B_(035E)    ........................T.............................................. [624]
Aph._ovalisporum_033E           ........................T.............................................. [624]
Aph._ovalisporum_035C           ........................T.............................................. [624]
Aph._ovalisporum_035D           ........................T.............................................. [624]
Aph._ovalisporum_035F           ........................T.............................................. [624]
C._raciborskii_AWT205_(23E_24C_2 ..................................................................... [624]
C._raciborskii_023D_(030A_032A_0 ..................................................................... [624]
Aph._sp.10E6                    ..................................................................... [624]


Aph._aphanizomenoides_259C      TTGTAAGTGTTCAGTGAGATAAGTTACATAA [655]
Ana._bergii_283A                ............................... [655]
Ana._bergii_366B                ............................... [655]
Aph._issatschenkoi_CAWBG02      ............................... [655]
```

*Aph._ovalisporum*_031E ..... [655]

*Aph._ovalisporum*_031G_(033C) ..... [655]

*Aph._ovalisporum*_033B_(035E) ..... [655]

*Aph._ovalisporum*_033E ..... [655]

*Aph._ovalisporum*_035C ..... [655]

*Aph._ovalisporum*_035D .....taa [655]

*Aph._ovalisporum*_035F ..... [655]

*C._raciborskii*_AWT205_(23E_24C_2 ..... [655]

*C._raciborskii*_023D_(030A_032A_0 .....taa [655]

*Aph.*_sp.10E6 ..... [655]

Appendix 10 Multiple Alignment of *cynB*

```
Aph.aphanizomenoides_259          TAAATGATTTACCCAGTAACCAGGGGAAGCGATCACTCCTTCCTCCAGGGAAGACTCAGGTGAAATCAAGTCGCCGGT [ 78]
Ana.bergii_283A                   ..C.............................................................................. [ 78]
Ana.bergii_366B                   ..C.............................................................................. [ 78]
Aph.issatschenkoi_CAWBGO2         ..C.............................................................................. [ 78]
Aph.ovalisporum_ILC-146           ..C.............................................................................. [ 78]
A.lapponica_966                   C.C.............G.............G.........A..AC........TCG.C.........C..T.. [ 78]
Aph.ovalisporum_31E               ..C.............................................................................. [ 78]
Aph.ovalisporum_31G               ..C.............................................................................. [ 78]
Aph.ovalisporum_33B               ..C.............................................................................. [ 78]
Aph.ovalisporum_33E               ..C.............................................................................. [ 78]
Aph.ovalisporum_35C               ..C.............................................................................. [ 78]
Aph.ovalisporum_35D               ..C.............................................................................. [ 78]
Aph.ovalisporum_35F               ..C.............................................................................. [ 78]
C.raciborskii_AWT205(23E_24C_26J  ..C.............................................................................. [ 78]
C.raciborskii_23D(030_032A_034)   ..C.............................................................................. [ 78]
Aph.sp._10E6                      ..C.............................................................................. [ 78]
Aph.sp._10E9                      ..C.............................................................................. [ 78]

Aph.aphanizomenoides_259          TAGATTACTGACAAGAGCTATTTTTGGAGGACTATAGTTGATTTCCCTGGCCACCTGTGCGAACTCCTGCTCCATTGG [156]
```

```
Ana.bergii_283A                 ..............T.............................................................. [156]
Ana.bergii_366B                 ..............T.............................................................. [156]
Aph.issatschenkoi_CAWBGO2       ..............T.............................................................. [156]
Aph.ovalisporum_ILC-146         ..............T.............................................................. [156]
A.lapponica_966                 ...G..............A..........G..................C...........A..........G...... [156]
Aph.ovalisporum_31E             ..............T.............................................................. [156]
Aph.ovalisporum_31G             ..............T.............................................................. [156]
Aph.ovalisporum_33B             ..............T.............................................................. [156]
Aph.ovalisporum_33E             ..............T.............................................................. [156]
Aph.ovalisporum_35C             ..............T.............................................................. [156]
Aph.ovalisporum_35D             ..............T.............................................................. [156]
Aph.ovalisporum_35F             ..............T.............................................................. [156]
C.raciborskii_AWT205(23E_24C_26J ..............T.............................................................. [156]
C.raciborskii_23D(030_032A_034) ..............T.............................................................. [156]
Aph.sp._10E6                    ..............T.............................................................. [156]
Aph.sp._10E9                    ..............T.............................................................. [156]

Aph.aphanizomenoides_259        TTTCATCAATGGAGAGTGGAAAGCGTGTGAAACAGACAGTCTCTTGCACCGGATTCCTTGATCTTCTAGGGTGTTGGT [234]
Ana.bergii_283A                 ....................................A.............................T..... [234]
Ana.bergii_366B                 ....................................A.............................T..... [234]
Aph.issatschenkoi_CAWBGO2       ..............T.............................................................. [234]
```

```
Aph.ovalisporum_ILC-146          ....................................A.................................T..... [234]
A.lapponica_966                  .GC.........C.....................G.G..A..........T...........CT......CC....... [234]
Aph.ovalisporum_31E              ....................................A.................................T..... [234]
Aph.ovalisporum_31G              ....................................A.................................T..... [234]
Aph.ovalisporum_33B              ....................................A.................................T..... [234]
Aph.ovalisporum_33E              ....................................A.................................T..... [234]
Aph.ovalisporum_35C              ....................................A.................................T..... [234]
Aph.ovalisporum_35D              ....................................A.................................T..... [234]
Aph.ovalisporum_35F              ....................................A.................................T..... [234]
C.raciborskii_AWT205(23E_24C_26J ..................................................................................... [234]
C.raciborskii_23D(030_032A_034)  ..................................................................................... [234]
Aph.sp._10E6                     ..................................................................................... [234]
Aph.sp._10E9                     ..................................................................................... [234]

Aph.aphanizomenoides_259         GAAGACTTGCACAGCTTGGCGATCGCCCGAAATGACGACACTTCGTGGTCCATTTACCGCCGCTATCACAACTCGCTG [312]
Ana.bergii_283A                  ..............C....C.................................................................. [312]
Ana.bergii_366B                  ..............C....C.................................................................. [312]
Aph.issatschenkoi_CAWBGO2        ..................................................................................... [312]
Aph.ovalisporum_ILC-146          ..............C....C.................................................................. [312]
A.lapponica_966                  T.........................G...........G.............C.....T..G...G.G...T.G.. [312]
Aph.ovalisporum_31E              ..............C....C.................................................................. [312]
```

```
Aph.ovalisporum_31G             ..............C....C.......................................................... [312]
Aph.ovalisporum_33B             ..............C....C.......................................................... [312]
Aph.ovalisporum_33E             ..............C....C.......................................................... [312]
Aph.ovalisporum_35C             ..............C....C.......................................................... [312]
Aph.ovalisporum_35D             ..............C....C.......................................................... [312]
Aph.ovalisporum_35F             ..............C....C.......................................................... [312]
C.raciborskii_AWT205(23E_24C_26J ................................................................................ [312]
C.raciborskii_23D(030_032A_034) ................................................................................ [312]
Aph.sp._10E6                    ................................................................................ [312]
Aph.sp._10E9                    ................................................................................ [312]

Aph.aphanizomenoides_259        GTCAGATTGAATAGCCGGCTTAACTTCCTCAAGGGAGGCCGATATAGCAACCATCTCGCCATTTTGTGGTAGCGCCTG [390]
Ana.bergii_283A                 ..............A..............T.....................C...C.......... [390]
Ana.bergii_366B                 ..............A..............T.....................C.............. [390]
Aph.issatschenkoi_CAWBGO2       ................................................................................ [390]
Aph.ovalisporum_ILC-146         ..............A..............T.....................C.............. [390]
A.lapponica_966                 .....G..........ATC.......G.T.....G...A.....................C...C.....T..... [390]
Aph.ovalisporum_31E             ..............A..............T.....................C.............. [390]
Aph.ovalisporum_31G             ..............A..............T.....................C...C.......... [390]
Aph.ovalisporum_33B             ..............A..............T.....................C...C.......... [390]
Aph.ovalisporum_33E             ..............A..............T.....................C...C.......... [390]
```

```
Aph.ovalisporum_35C              .................A.................T.....................C...C.......... [390]
Aph.ovalisporum_35D              .................A.................T.....................C...C.......... [390]
Aph.ovalisporum_35F              .................A.................T.....................C...C.......... [390]
C.raciborskii_AWT205(23E_24C_26J ..................................................................... [390]
C.raciborskii_23D(030_032A_034)  ..................................................................... [390]
Aph.sp._10E6                     ........................T.................................C.............. [390]
Aph.sp._10E9                     ........................T.................................C.............. [390]

Aph.aphanizomenoides_259         CATCAAACGGCCTCTTTCGGCAGCAAGTTTGAGACCATCCTCCAGAGAAAAGATCCCCGCCACACAAGCTGCCACATA [468]
Ana.bergii_283A                  .........C........................................................... [468]
Ana.bergii_366B                  .........C........................................................... [468]
Aph.issatschenkoi_CAWBGO2        ..................................................................... [468]
Aph.ovalisporum_ILC-146          .........C........................................................... [468]
A.lapponica_966                  .........A.............C....C.............................G...........T..... [468]
Aph.ovalisporum_31E              .........C........................................................... [468]
Aph.ovalisporum_31G              ..................................................................... [468]
Aph.ovalisporum_33B              ..................................................................... [468]
Aph.ovalisporum_33E              ..................................................................... [468]
Aph.ovalisporum_35C              ..................................................................... [468]
Aph.ovalisporum_35D              ..................................................................... [468]
Aph.ovalisporum_35F              ..................................................................... [468]
```

```
C.raciborskii_AWT205(23E_24C_26J ..................................................................... [468]
C.raciborskii_23D(030_032A_034)  ..................................................................... [468]
Aph.sp._10E6                    ..................................................................... [468]
Aph.sp._10E9                    .....................................------------------------------ [468]

Aph.aphanizomenoides_259        TTCACCTATGCTATGACCCATTACGATATCTGGCTCGATTCCCCACGACCGCCACATCTGTGCTAGGGCATATTCTAC [546]
Ana.bergii_283A                 .....................................T..................G........ [546]
Ana.bergii_366B                 .....................................T..................G........ [546]
Aph.issatschenkoi_CAWBGO2       ..................................................................... [546]
Aph.ovalisporum_ILC-146         ......................------.....................T..................G........ [546]
A.lapponica_966                 ..................G..G..A.......................A...........CA.A.............. [546]
Aph.ovalisporum_31E             .....................................T..................G........ [546]
Aph.ovalisporum_31G             .....................................T..................G........ [546]
Aph.ovalisporum_33B             .....................................T..................G........ [546]
Aph.ovalisporum_33E             .....................................T..................G........ [546]
Aph.ovalisporum_35C             .....................................T..................G........ [546]
Aph.ovalisporum_35D             .....................................T..................G........ [546]
Aph.ovalisporum_35F             .....................................T..................G........ [546]
C.raciborskii_AWT205(23E_24C_26J ..................................................................... [546]
C.raciborskii_23D(030_032A_034)  ..................................................................... [546]
Aph.sp._10E6                    ..................................................................... [546]
```

```
Aph.sp._10E9                    ------------------------------------------------------------------------------- [546]


Aph.aphanizomenoides_259        AGCAAAAAGGGCAGG [561]
Ana.bergii_283A                 ............... [561]
Ana.bergii_366B                 ............... [561]
Aph.issatschenkoi_CAWBGO2       ............... [561]
Aph.ovalisporum_ILC-146         ............... [561]
A.lapponica_966                 .CT............ [561]
Aph.ovalisporum_31E             ..A............ [561]
Aph.ovalisporum_31G             ............... [561]
Aph.ovalisporum_33B             ............... [561]
Aph.ovalisporum_33E             ............... [561]
Aph.ovalisporum_35C             ............... [561]
Aph.ovalisporum_35D             ............... [561]
Aph.ovalisporum_35F             ............... [561]
C.raciborskii_AWT205(23E_24C_26J ............... [561]
C.raciborskii_23D(030_032A_034) ............... [561]
Aph.sp._10E6                    ............... [561]
Aph.sp._10E9                    --------------- [561]
;
end;
```