CONSERVATION OF THE CYANOBACTERIAL CIRCADIAN CLOCK:

COMPARATIVE STUDIES IN *NOSTOC* SP. STRAIN PCC 9709, A CYANOBACTERIUM ISOLATED FROM THE LICHEN *PELTIGERA MEMBRANACEA*

by

AMY LEE

B.Sc., The University of British Columbia, 1998

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Department of Microbiology and Immunology)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 2003

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Microbiology and Immunology

The University of British Columbia
Vancouver, Canada

Date August 25, 2003

ABSTRACT

The Cyanobacteria are the only group of prokaryotic organisms known to possess a circadian program. Efforts to elucidate the molecular mechanism underlying the cyanobacterial circadian clock have focused almost exclusively on the *kaiABC* gene cluster of the unicellular model organism *Synechococcus elongatus* PCC 7942. The considerable morphological and habitat diversity of cyanobacteria raises obvious questions: Do all members of the Cyanobacteria utilize circadian programs? How similar might other cyanobacterial clocks be to the *S. elongatus* model? To investigate whether clock molecular biology is conserved, I undertook a comparative study with *Nostoc* strain PCC 9709 as the subject. *Nostoc* PCC 9709 is a heterocyst-forming filamentous cyanobacterium originally isolated as a symbiont within the lichen *Peltigera membranacea*. **Methods:** The *Nostoc* PCC 9709 genome was screened for clock gene homologues using PCR-based methods and Southern blot hybridization. To support genetic identification with functional data, the temporal gene expression profiles of putative *kai* homologues were compared against the previously determined rhythmic abundances of *kaiA* and *kaiBC* transcripts of *S. elongatus* with the use of Northern blot hybridization and relative RT-PCR assays. **Conclusions:** The prokaryotic circadian clock gene cluster *kaiABC* is conserved in *Nostoc* PCC 9709. Deduced amino acid sequence homology (46 % identity for KaiA, 81 % for KaiB, 80 % for KaiC) is consistent with conservation of function in accordance with available structure-function data for the *S. elongatus* model. A relative RT-PCR assay method was validated on the *S. elongatus* model system. Time course studies on *kai* gene transcription demonstrated peak-and-trough patterns in cultured *Nostoc* and field specimens of *P. membranacea* that are consistent with circadian oscillation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	adenine
bp	basepair
C	cytosine
cDNA	complementary deoxyribonucleic acid
CT	circadian time
DEPC	diethyl pyrocarbonate
DNA	deoxyribonucleic acid
dCTP	2'-deoxycytidine 5'-triphosphate
dNTP	2'-deoxynucleoside 5'-triphosphate
EDTA	ethylene-diamine-tetraacidic acid
G	guanine
I	inosine
kb	kilobase
kD	kilodalton
M	amino (adenine or cytosine)
mRNA	messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
PCR	polymerase chain reaction
PFU	plaque-forming unit
R	purine (adenine or guanine)
RNA	ribonucleic acid
rDNA	ribosomal DNA
rRNA	ribosomal RNA
RT	reverse transcription
RT-PCR	reverse transcription-polymerase chain reaction
SDS	sodium dodecyl sulfate
SSC	standard saline citrate; 1× solution = 15 mM sodium citrate, 0.15 M NaCl, pH 7.0
T	thymine
TBE	Tris-borate-EDTA; composition of 1× solution = 89 mM Tris, 89 mM Boric acid, 2 mM EDTA, pH 8.3
TE	Tris-EDTA; composition of 1× solution = 10 mM Tris, 1 mM EDTA, pH 8.0
Y	pyrimidine (cytosine or thymine)

ACKNOWLEDGEMENTS

I would like to thank Julian and Tom for providing this open opportunity to pursue a research interest. I must also thank them for their patience. I am indebted to Dr. Susan Golden for graciously welcoming me to her circadian research laboratory at Texas A&M University. Special thanks go to Vivian Miao as well as Marie-Françoise Coeffet-LeGal for guidance, care and advice – day-in and day-out – in the lab.

Finally, I particularly wish to acknowledge Denis and my friends and co-workers at Protiva Biotherapeutics – this dissertation would never have been completed without your understanding and morale-boosting support.

This research was financially supported in part by a Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship.

CHAPTER ONE
General Introduction

Circadian programs are regulatory systems which allow organisms to anticipate regular changes in their local environment and alter their physiological state as appropriate in order to gain an adaptive advantage. A circadian system can be described as consisting of a core oscillator (the rhythm-generator), an input component which entrains the internal oscillator to extracellular environmental cues, and an output component which converts the oscillation into synchronized metabolic rhythms. Circadian programs display three diagnostic characteristics: 1) rhythms have a roughly 24 hour period and persist in the absence of external timing signals, 2) the phase of the rhythm can be reset by light or temperature cues to allow adaptation to the local environment, and 3) the period is temperature-compensated such that rhythms are stable at different constant ambient temperatures.

Circadian rhythms are found in most eukaryotes and the genetic and biochemical bases of these systems have been extensively studied in a number of model organisms including *Drosophila*, *Neurospora*, *Arabidopsis*, and *Mus musculus* (reviewed in 16, 28). Currently, the only known examples of prokaryotic circadian programs are found in the Cyanobacteria. Many different types of cyanobacteria have been observed to carry out temporal regulation of critical metabolic functions such as nitrogen fixation and photosynthesis (reviewed in 24). Genes encoding the cyanobacterial circadian oscillator were first identified in *Synechococcus elongatus* strain PCC 7942 (32). This free-living unicellular cyanobacterium has a three-component circadian oscillator encoded in a 2.9 kb gene cluster (named *kaiABC*) on the chromosome. There appears to be no genetic similarity between the cyanobacterial clock and any of the known eukaryotic clock genes.

Because *S. elongatus* has as a small genome, grows rapidly in laboratory culture, and is amenable to a large assortment of genetic manipulations, it is an ideal model organism for investigation of the molecular basis of circadian rhythms. Consequently, a great deal of biochemical information on the S. elongatus *kaiABC*-encoded oscillator has been gathered since the gene cluster was first reported in 1998. We know that abundances of KaiB and KaiC proteins vary rhythmically in a circadian manner (77). Kai proteins function by physically interacting with each other as well as other molecules suspected to be involved in input and output pathways for circadian programs (34, 37). KaiC seems to be the pivotal component of the

oscillator; this protein has rhythmic patterns of auto-phosphorylation (33, 55) and forms ATP-dependent hexameric rings which can bind DNA (29, 53). KaiA protein enhances and KaiB protein attenuates phosphorylation of KaiC (42, 76, 78). Despite all this research information, it is yet unknown how circadian signals are transmitted and distributed through the cell infrastructure to orchestrate expression of the cyanobacterial genome (45).

Thesis objectives: The goal of this study was to assess whether a detailed understanding of the molecular mechanism of circadian oscillation in *S. elongatus* PCC 7942 is relevant to a broader analysis of circadian programs associated with the diverse group of organisms known as the Cyanobacteria. To address this question, I chose to conduct a comparative study in *Nostoc* PCC 9709. This cyanobacterium is ecologically and morphologically dissimilar to *S. elongatus*, being a symbiotically-associated heterocyst-forming filamentous organism originally isolated from the lichen *Peltigera membranacea* (51) whereas *S. elongatus* is a unicellular free-living cyanobacterium that is incapable of nitrogen-fixation. These studies were organized into two parts: 1) identification of potential *kaiABC* homologues on the genetic level, using PCR-based and Southern blot techniques, and 2) expression characterization of the putative *kai* homologues at the mRNA level, using Northerm blot and relative RT-PCR methods.

?

CHAPTER TWO
Materials and Methods

2.1. Cyanobacterial strains and culture conditions

2.1.1. *S. elongatus* PCC 7942

A modified strain of *S. elongatus* PCC 7942 (designated AMC541), containing a P_{kaiB}:*luc* reporter construct inserted into the NSII site of the chromosome, was obtained from the laboratory of Susan Golden (Texas A&M University, Texas, USA). This strain was grown on modified BG-11 growth medium (8) supplemented with 1 mM sodium thiosulfate, *L*-proline (2 mg/mL), chloramphenicol (7.5 μg/mL) and solidified with 1.5 % agar.

2.1.2. *Nostoc* sp. PCC 9709

Nostoc sp. PCC 9709 was obtained from the Pasteur Culture Collection (Paris, France) and was maintained on modified BG-11 growth medium (2) supplemented with a vitamins solution (57) and solidified with 1 % agar. Liquid cultures of *Nostoc* PCC 9709 were maintained in 250 mL Erlenmeyer flasks with sponge stoppers to allow gas exchange. Periodically, flask cultures were assessed for contamination by microscopic examination and by testing for growth in Luria Bertani broth at 30 °C.

2.1.3. Preparation of culture media

Water used to prepare BG-11 media was purified to a resistance of 18 MΩ with a Milli-Q Biocel water system (Millipore). To make solid BG-11, equal volumes of 2× agar solution (containing Difco Select Agar, BD Diagnostic Systems) and 2× nutrient solution were prepared, autoclaved separately, and then combined according to the method of Allen (2).

2.1.4. Environmental settings

Strains were maintained at 25 °C under a daily lighting regime of 16 hours illumination followed by 8 hours darkness. Cool white light from two 8 W fluorescent lamps at the top of the incubator provided illumination at a photon flux density of approximately 25 μmol•m^{-2}•s^{-1} photosynthetically active radiation (400–700 nm wavelength range). To decrease the rate of water evaporation from slow-growing cultures, plates were wrapped in Parafilm and open containers of sterile water were placed in the incubator to raise the ambient humidity.

2.2. DNA sources and preparation

DNA samples were prepared in TE buffer. DNA concentration was determined by fluorometric analysis, utilizing a TKO 100 Mini-Fluorometer (Hoefer Scientific Instruments) to

measure the 460 nm wavelength emission of Hoechst 33258 dye (bis benzimidazole). Samples were quantified in assay solution (0.1 μg/mL Hoeschst 33258 dye, 10 mM Tris-Cl, 1 mM EDTA, and 0.2 M NaCl, pH 7.4) against DNA standard pGEM-3Zf(+), which was obtained as a 0.2 μg/μL solution from the ABI Prism Dye Terminator Cycle Sequencing (Applied Biosystems).

2.2.1. Genomic DNA

S. elongatus PCC 7942 genomic DNA was obtained from Dr. Didier Mazel (Pasteur Institute, Paris, France). *Nostoc* sp. PCC 9709 genomic DNA was recovered by back-extraction of the organic phases generated during phenol-chloroform RNA extraction (11, 41) and concentrated using a standard isopropanol precipitation procedure (4). Genomic DNA from *P. membranacea* specimen TDI#AR95 (51) was obtained from TerraGen Discovery (Vancouver, BC). Genomic DNA was extracted from *P. membranacea* specimens AL#21 and AL#24 using the FastPrep MH DNA extraction kit (Qbiogene).

2.2.2. Plasmids and host strains

Plasmids and *E. coli* host strains used in this study are listed in Table 2.1. Plasmid pJS1 (47) was obtained from TerraGen Discovery (Vancouver, BC). *E. coli* strains were obtained from Invitrogen. Plasmids were propagated by culturing host bacteria at 37 °C in Luria Bertani broth supplemented with the appropriate selective antibiotic. Plasmids were purified from cell cultures using a standard alkaline lysis procedure (6).

2.3. Oligonucleotide primers

Oligonucleotide primers used in this study are described in Table 2.2 and Table 2.3. Primers were designed using the Primer3 software program (61) or by manual sequence inspection. Custom-designed and standard oligonucleotides were synthesized on an Oligo 1000M DNA Synthesizer (Beckman Coulter). Following application of the Ultra*FAST* Cleavage and Deprotection kit (Beckman Coulter), synthesized oligonucleotides were not further purified before research use. Primer stocks and working dilutions were prepared in water and stored at - 20 °C. When required, a single aliquot of working solution was thawed and kept at 4 °C for the duration of use. Oligonucleotide stock concentration was determined by spectrophotometric measurement of absorbance at 260 nm (using water as diluent and blank).

Table 2.1. Plasmids and bacterial host strains used in this study

Plasmid	Description	Host Strain	Source / Reference
pJS1	PCC 7942 *kaiABC*	*not applicable*	Lorne et al., 2000
pKN10A1	PCC 9709 *kaiC* amplicon	*E. coli* DH10B	This study
pGEM-3Zf(+)	DNA standard	*not applicable*	Applied Biosciences
pkai2A-1	PCC 9709 genomic library clone	*E. coli* XL0LR	This study
pkai2B-3	PCC 9709 genomic library clone	*E. coli* XL0LR	This study
pkai2C-1	same as pkai2B-3	*E. coli* XL0LR	This study
pkai2D-5	PCC 9709 genomic library clone	*E. coli* XL0LR	This study
pkai2E-1	PCC 9709 genomic library clone	*E. coli* XL0LR	This study
pkai2F-1	PCC 9709 genomic library clone	*E. coli* XL0LR	This study
pkai2G-1	PCC 9709 genomic library clone	*E. coli* XL0LR	This study
pkai2H-1	PCC 9709 genomic library clone	*E. coli* XL0LR	This study
pkai2I-1	same as pkai2M-1	*E. coli* XL0LR	This study
pkai2J-1	PCC 9709 genomic library clone	*E. coli* XL0LR	This study
pkai2K-1	PCC 9709 genomic library clone	*E. coli* XL0LR	This study
pkai2L-2	PCC 9709 genomic library clone	*E. coli* XL0LR	This study
pkai2M-1	PCC 9709 genomic library clone	*E. coli* XL0LR	This study
pkai2N-1	same as pkai2G-1	*E. coli* XL0LR	This study
pNrnp-1	PCC 9709 *rnpB* amplicon	*E. coli* DH5α	This study

Table 2.2. Oligonucleotide primers specific for *Nostoc* PCC 9709 *kaiABC*

Name	Sequence (5'–3')	Reference
O-495	CGACGGCGACGTTCTCCTTC	This study
O-496	CCTTGATCAGTGGTACCTCC	This study
O-497	GGAGGTACCACTGATCAAGG	This study
O-498	GAAGGAGAACGTCGCCGTCG	This study
O-510	TCTTCAATTACCTCCATGTG	This study
O-511	GCGCATTCGATTACTGATC	This study
O-532	GTGGAACCCTCCGGCATTTC	This study
O-533	GCCAAACTTATCAGCCCACC	This study
O-642	CTTGATTCTGCCACCTGAT	This study
O-643	CCTACGTTCTCAAGCTTTAC	This study
O-644	CACTCAGTTCTTCATAGAGC	This study
O-651	GACTGAGCTCTGCAGATCTAGAGTTGCGT ATTACTGGTGTC	This study
O-662	AACGTTAGACGGTAATCCAG	This study
O-676	ACCGATCTTGATTCTGCCAC	This study
O-677	GGGGCACAGGAATATTAGCA	This study
O-678	CGATTGAAACTCGTTTAGCTTT	This study
O-685	GGGYGTGGAAATTKATTGAC	This study
O-686	GATATTAGTCATGGTGGTTTACCAA	This study

Table 2.3. Other oligonucleotide primers used in this study

Name	Sequence (5'–3')	Reference
Specific for cyanobacterial *kaiC*		
O-488	ACIATGATIGARGGITTYGAYGAIAT	Lorne et al., 2000
O-489	GTRAAIGGRTAYTCIGGYTTCATRTG	Lorne et al., 2000
Specific for *S. elongatus* PCC 7942 *kaiABC*		
O-672	GACCAAATCGACTGCCTGAT	This study
O-673	GAGCTTCTTGGCAAACTCGT	This study
O-674	CCCTCAAAACGCTCAAGAAC	This study
O-675	CGAGCTACCAACCGAAAGAG	This study
O-679	TCTTGAAGTTCGCCGTAGAT	This study
O-682	ATTCTGAGCACCAAGCCATC	This study
Specific for bacterial 16S rRNA genes[a]		
16S.0007.F21	GAGAGTTTGATCCTGGCTCAG	Miao et al., 1997
16S.0515.F16	GTGCCAGCAGCCGCGG	Miao et al., 1997
16S.1100.F16	CAACGAGCGCAACCCT	This study
16S.1511.R21	CGGCTACCTTGTTACGACTTC	Miao et al., 1997
Specific for cyanobacterial *rnpB*		
O-663	GGAGGAAAGTCCGGGCTC	This study
O-664	AAGCCGGGTTCTGTTMTC	This study
Specific for common plasmid vectors		
S1224[b]	CGCCAGGGTTTTCCCAGTCACGAC	New England Biolabs
S1228[c]	ATTAACCCTCACTAAAGGGA	New England Biolabs
S1233[d]	AGCGGATAACAATTTCACACAGGA	New England Biolabs
S1248[e]	TAATACGACTCACTATAGGG	New England Biolabs

[a] These primers were designed to bind to consensus sequences thought to be present in all bacterial 16S rRNA genes and incorporate in their names information about the corresponding position on *the E. coli* 16S rDNA sequence to which their 5' terminal nucleotide align, their direction (forward 'F' or reverse 'R'), and their size (e.g. primer 16S.0007.F21 is a 21-base oligonucleotide starting at position 7 and allows sequencing forward).
[b] M13/pUC forward sequencing primer (-47)
[c] T3 promoter primer
[d] M13/pUC reverse sequencing primer (-48)
[e] T7 universal primer

2.4. Bacteriophage library screen

A *Nostoc* PCC 9709 genomic library (designated λZAP:773), composed of 4-10 kb fragments prepared by partial *Sau*3A I digestion and inserted in the *Bam*H I site of the Lambda ZAP Express vector (Stratagene), was constructed by Marie-Françoise Coeffet-LeGal and kindly made available by TerraGen Discovery (Vancouver, BC). Aliquots of the λZAP:773 library were plated on *E. coli* strain XL1-Blue MRF' and plaques were transferred onto Nytran+ membranes (Schleicher & Schuell) in duplicate according to instructions provided by the manufacturer of the vector. Probe was prepared by incorporation of radioisotope [α-^{32}P]dCTP (Amersham Biosciences) into gel-purified PCC 9709 *kaiC* PCR product (using a random primer method) followed by purification on a ProbeQuant G-50 Micro Column (Amersham Biosciences). Hybridization of radiolabeled probe DNA to plaque lifts was performed for 12 to 16 hours (overnight) at 60 °C in hybridization buffer consisting of 0.25 M Na$_2$HPO$_4$, 1 mM EDTA, 1 % bovine serum albumin and 7 % SDS. Afterwards, membranes were washed at 60 °C twice in 2× SSC/0.1 % SDS for 5 minutes and twice in 0.2x SSC/0.1 % SDS for 15 minutes before being exposed to BioMax Light X-ray film (Kodak) for one to six hours with intensifying screens. Where strong signals were detected on duplicate membranes, bacteriophage plaque cores were extracted and subjected to a second purifying round of screening by hybridization. A pBK-CMV phagemid containing the genomic insert was prepared by *in vivo* excision from each purified bacteriophage clone using ExAssist helper phage and transferred into *E. coli* XLOLR (for standard plasmid propagation) according to the instructions provided by the manufacturer of the Lambda ZAP Express vector.

2.5. DNA restriction map-analysis

Restriction endonucleases *Kpn* I, *Not* I, *Pst* I, *Sac* I and associated reagents were obtained from New England Biolabs. For map-analysis of library clones (Section 3.3.2), plasmid DNA (200 ng per 20 μL reaction) was digested by *Kpn* I and *Sac* I in 1× NEBuffer 1 and also by *Not* I and *Pst* I in 1× NEBuffer 3. Digest reactions were supplemented with BSA (100 μg/mL) and were allowed to proceed for at least one hour at 37 °C. Fragment size was determined by comparison against 1 Kb Plus DNA Ladder (Invitrogen) after electrophoretic separation of digested DNA through a 0.8 % agarose gel in 0.5× TBE buffer.

2.6. PCR amplification

PCR reactions contained 225 μM dNTP mix, 1 unit recombinant *Taq* DNA polymerase, 1× PCR buffer, and varying concentrations of primers, $MgCl_2$ and template DNA in 20 μL total volume. DNA polymerase, PCR buffer, $MgCl_2$ and dNTP mix were obtained from Invitrogen. Amplification reactions were carried out in PTC-100 thermal cyclers (MJ Research). PCR products were separated by agarose gel electrophoresis in 0.5× TBE buffer for analysis. Product size was determined by comparison against 1 Kb Plus DNA Ladder (Invitrogen).

2.6.1. *kaiC* gene

PCR reactions contained 1 μM each of primers O-488 and O-489, 2.0 mM $MgCl_2$ and up to 5 ng plasmid or genomic DNA in addition to the reagents described in Section 2.6 above. For use as PCR template, a small amount of the λZAP:773 genomic library bacteriophage stock (5 x 10^8 PFU/mL) was diluted 1:10 in water and 1 μL of the dilution was added directly to the PCR reaction mixture. The amplification program was comprised of 40 cycles of template denaturation for 45 seconds at 92 °C, primer annealing for 45 seconds at 51 °C, and polymerization for 1 minute at 72 °C. After analysis of completed reactions on a 1 % agarose gel, PCR products of interest were extracted from the gel matrix using a Geneclean Spin Kit (Qbiogene) and eluted into TE buffer. The gel-isolated PCR products were cloned into a plasmid vector as described in Section 2.7.1 and also directly subjected to DNA sequence analysis as described in Section 2.8.

2.6.2. Map-analysis of library clones

The inserts of genomic library clones were subjected to PCR analysis with multiple combinations of vector-specific (T3, T7) and probe-specific (O-488, O-489, O-495, O-496, O-497, O-498) primers. PCR reactions contained 0.2 μM primers, 1.5 mM $MgCl_2$ and 50 pg phagemid DNA in addition to the reagents described in Section 2.6 above. The amplification program comprised 32 cycles of template denaturation for 30 seconds at 92 °C, primer annealing for 30 seconds at 56 °C, and polymerization for 3 minutes and 45 seconds at 72 °C. Completed reactions were analyzed on 0.8 % agarose gels.

2.6.3. 16S rRNA gene

Amplification of 16S rRNA genes was performed as described by Miao et al. (51). PCR reactions contained 0.375 μM each of primers 16S.0007.F21 and 16S.1511.R21, 2.0 mM $MgCl_2$ and 2 ng genomic DNA in addition to the reagents described in Section 2.6 above. The

amplification program was comprised of 30 cycles of template denaturation for 40 seconds at 92 °C, primer annealing for 40 seconds at 55 °C, and polymerization for 2 minutes at 72 °C. After analysis of completed reactions on a 1 % gel, PCR products were removed from the reaction mixture using a standard ethanol/ammonium acetate precipitation procedure (4), redissolved in water and directly subjected to DNA sequence analysis as described in Section 2.8.

2.6.4. *P. membranacea kaiABC* gene cluster

PCR reactions contained 0.2 μM each of primers O-495 and O-676, 1.5 mM MgCl$_2$ and 1 ng genomic DNA in addition to the reagents described in Section 2.6 above. The amplification program was comprised of 40 cycles of template denaturation for 45 seconds at 92 °C, primer annealing for 45 seconds at 51 °C, and polymerization for 2 minutes at 72 °C. After analysis of completed reactions on a 1 % gel, PCR products were removed from the reaction mixture using a standard ethanol/ammonium acetate precipitation procedure (4), redissolved in water and directly subjected to DNA sequence analysis as described in Section 2.8.

2.6.5. RNase P (*rnpB*) gene

PCR reactions contained 0.2 μM each of primers O-663 and O-664, 1.5 mM MgCl$_2$ and 1 ng genomic DNA in addition to the reagents described in Section 2.6 above. The amplification program was comprised of 35 cycles of template denaturation for 45 seconds at 92 °C, primer annealing for 45 seconds at 52 °C, and polymerization for 45 seconds at 72 °C. After analysis of completed reactions on a 1 % gel, PCR products were directly cloned into a plasmid vector (Section 2.7.2) without the use of any post-amplification purification procedures.

2.6.6. Colony PCR

Cell lysates were prepared by transferring (with the use of a sterile toothpick) a small amount of a bacterial colony into a 0.5 mL tube containing 50 μL water and incubating the mixture at 100 °C for 5 minutes in a PTC-100 thermal cycler (MJ Research). Before use, cell lysates were briefly centrifuged to clear cell debris from the sample. PCR reactions (40 μL volume) contained 0.2 μM primers, 1.5 mM MgCl$_2$ and 5 μL cell lysate in addition to the reagents described in Section 2.6 above. Primer combinations are described in Section 2.7.1 and Section 2.7.2. The amplification program was comprised of 30 cycles of template denaturation for 45 seconds at 92 °C, primer annealing for 45 seconds at 54 °C, and polymerization for 45 seconds at 72 °C. Completed reactions were analyzed on 1 % agarose gels.

2.7. Cloning PCR products

2.7.1. *kaiC* amplicon

Gel-isolated *kaiC* PCR products amplified from Nostoc PCC 9709 (Section 2.6.1) were ligated into the pGEM-T vector (pGEM-T vector system, Promega) according to the manufacturer's instructions. *E. coli* DH10B (Invitrogen) cells, which had been made electrocompetent according to a procedure described in (64), were transformed with ligation products and cultured at 37 °C on solid Luria Bertani medium containing ampicillin (100 μg/mL) for selection. Transformants were screened by colony PCR (Section 2.6.5) using vector-specific primers S1224 and S1233 to identify clones containing inserts of the correct size. The insert sequence of the chosen clone (named pKN10A1) was determined by subjecting colony PCR products (purified and resuspended in water using the Qbiogene Geneclean Spin Kit) to DNA sequence analysis as described in Section 2.8.

2.7.2. RNase P (*rnpB*) amplicon

After amplification (Section 2.6.5), *rnpB* PCR products were directly ligated into the pCR2.1-TOPO vector (TOPO TA Cloning Kit, Invitrogen) according to the manufacturer's instructions. *E. coli* DH5α (Invitrogen) cells, which had been made electrocompetent according to a procedure described in (64), were transformed with ligation products and cultured at 37 °C on solid Luria Bertani medium containing ampicillin (100 μg/mL) for selection. Transformants were screened by colony PCR (Section 2.6.6) using vector-specific primers S1233 and S1248 to identify clones containing inserts of the correct size. The insert sequence of the chosen clone was determined by subjecting purified plasmid to DNA sequence analysis as described in Section 2.8.

2.8. DNA sequence analysis

DNA sequences were determined using the Sanger method (63) with fluorescence labeled dideoxynucleotide chain terminators. Each 20 μL sequencing reaction contained 6 μL Ready Reaction Mix from the ABI Prism Dye Terminator Cycle Sequencing kit (Applied Biosystems) and was carried out in a PTC-100 thermal cycler (MJ Research) in a fashion equivalent to the manufacturer's recommendations. DNA templates and sequencing primers are described in Sections 2.8.1 to 2.8.4. After reaction products were purified through CentriSep columns (Princeton Separations), data was collected on a 373 Stretch automated DNA sequencer (Applied Biosystems) operated by TerraGen Discovery (Vancouver, BC) or on

similar instruments operated by the University of British Columbia Nucleic Acids Protein Service Unit (Vancouver, BC).

Sequence data were corrected manually upon inspection of electropherograms and assembled using GeneWorks 2.4 (Accelrys) and AssemblyLign 1.0.9b (Accelrys) computer software programs. MacVector 7.0 (Accelrys) software was used to perform amino acid translations and Clustal W alignments (69). Nucleic acid and deduced amino acid sequences were compared to sequences in public databases using the BLAST program (3, 23). Protein motif searches and molecular mass calculations were performed using computational tools available on the PROSITE database (http://www.expasy.org/prosite/) (20, 65). Prokaryotic promoter searches were performed using the Neural Net Promoter Prediction software program available on the Lawrence Berkeley National Laboratory's Berkeley Drosophila Genome Project website (http://www.fruitfly.org/).

2.8.1. *kaiC*

The 0.65 kb lengths of amplified *kaiC* PCR products (Section 2.6.1) and the cloned PCC 9709 *kaiC* amplicon (Section 2.7.1) were sequenced on both strands with the use of primers O-488 and O-489.

2.8.2. Nostoc PCC 9709 library clones

A 2.7 kb region of the 6.5 kb insert on plasmid pkai2A-1 was sequenced using primers O-495, O-497, O-510, O-532, O-533 and S1228. A 1.9 kb region of the 4.5 kb insert on plasmid pkai2D-5 was sequenced using primers O-498, O-511, O-651 and S1228. The entire 3.2 kb insert on plasmid pkai2E-1 was sequenced using primers O-495, O-497, O-510, O-533, S1228 and S1248. A 3.1 kb region (minus a 0.2 kb gap) of the 7 kb insert on plasmid pkai2L-2 was sequenced using primers O-496, O-498, O-511, O-643, O-651 and S1228.

2.8.3. 16S rRNA

The 1.4 kb lengths of amplified 16S rDNA PCR products (Section 2.6.3) were sequenced (70% multiple-pass coverage) with the use of primers 16S.0007.F21, 16S.0515.F16 and 16S.1511.R21.

2.8.4. *P. membranacea kaiABC*

The 1.9 kb lengths of amplified kaiABC PCR products (Section 2.6.4) were sequenced (minimum double-pass coverage) with the use of primers O-495, O-643, O-644, O-676 and O-678.

2.8.5. RNase P (*rnpB*)

The 0.39 kb length of the cloned *Nostoc* PCC 9709 *rnpB* amplicon (Section 2.7.2) was sequenced on one strand with use of vector primer S1248.

2.9. Preparation of biological material for time course studies

Cultures of *S. elongatus* PCC 9742, *Nostoc* PCC 9709 and *P. membranacea* specimens were pre-conditioned with at least 12 hours darkness to reset the circadian clock before being subjected to constant light for the duration of the experiment.

2.9.1. *S. elongatus*

Plate cultures of *S. elongatus* PCC 7942 were monitored by visual inspection and time course experiments were scheduled so as to utilize actively growing cultures and not senescent cell populations. At the appropriate time points, samples of PCC 7942 were collected from plate cultures by suspending cells in sterile water with the use of a thin glass spreader, pipetted off the plate and recovered by centrifugation before being flash-frozen in liquid nitrogen and stored at -80 °C prior to RNA extraction.

2.9.2. *Nostoc*

At the appropriate time points, samples of *Nostoc* PCC 9709 were collected from liquid culture, separated from growth medium by filtration and either immediately flash-frozen and placed into -80 °C storage or ground under liquid nitrogen using a mortar and pestle and resuspended in Trizol reagent (Invitrogen) and stored at -80 °C prior to RNA extraction. In the latter case, Trizol was added directly to the mortar while the cells were still frozen so that the cells thawed in Trizol before the mixture was transferred into tubes for -80 °C storage.

2.9.3. *P. membranace*a

Freshly collected specimens of *P. membranacea* were slightly moistened with tap water and kept in lidded plastic petri dishes under the same environmental conditions as used for culture of cyanobacterial strains (Section 2.1). Lichen thalli were cut so that all pieces contained roughly equal portions of the lobe tip. After a short period of acclimation (three days), the environmental settings were altered for time course studies as described in Section 2.9 above. For a each time point, one portion of each thallus was flash-frozen whole under liquid nitrogen and stored at -80 °C prior to RNA extraction.

2.10. Preparation of RNA

All purified RNA samples were prepared in DEPC-treated water. Vigorous vortex-mixing and brief (one to two minute) incubations at 50 °C were required to thoroughly redissolve precipitated RNA. RNAguard (RNase inhibitor isolated from human placenta, Amersham Biosciences) was added to protect purified RNA against enzymatic degradation and samples were stored at -80 °C. RNA concentration was determined by spectrophotometric measurement of absorbance at 260 nm (using water as diluent and blank). The quality of prepared RNA was assessed according to the integrity of rRNA bands as visualized by ethidium bromide staining of samples after electrophoresis in a 0.22 M formaldehyde/1.2 % agarose gel.

2.10.1. *S. elongatus*

RNA was extracted from *S. elongatus* using a hot-phenol procedure based on the method of Mohamed & Jansson (52). Cell pellets (20-30 μL volume) removed from -80 °C storage were subjected to one freeze-thaw cycle using liquid N_2 before being resuspended in 750 μL combined lysis solution (0.11 M sucrose, 9.4 mM sodium acetate pH 4.5, 28.3 mM EDTA pH 8.0, 1.1% SDS, 1% β-mercaptoethanol). Samples were mixed with 0.5 volume of hot phenol (water-saturated, Invitrogen) by vigorous vortexing and incubated for 3 minutes at 65 °C. The mixing and 65 °C incubation steps were repeated after the addition of 0.5 volume of chloroform. The aqueous phase was collected after centrifugation (15,000 x g for 20 minutes at 4 °C), extracted twice more with equal volumes of acid-phenol:chloroform (5:1, pH 4.7, Ambion), and extracted for a fourth time with an equal volume of chloroform. RNA was precipitated from the aqueous phase by overnight incubation at 0 °C (within an ice-water slurry) in the presence of 2 M LiCl (LiCl-mediated RNA precipitation is a standard procedure described in (4)). RNA was recovered by centrifugation (15,000 x g for 20 minutes at 4 °C), washed with 2 M LiCl and resuspended in water. The LiCl-mediated precipitation (minimum 4 hour incubation at 0 °C) and wash procedures were repeated once. After a final precipitation (overnight incubation at -20 °C) with 0.08 volume of 7.5 M ammonium acetate and 3 volumes of anhydrous ethanol, the purified RNA was recovered by centrifugation as before. The final RNA pellet was washed with 75 % ethanol before being resuspended in water and treated as described in Section 2.10 above.

2.10.2. *Nostoc* **method A**

RNA was obtained from cultured *Nostoc* in the same manner as for *S. elongatus* (Section 2.10.1) except that the filamentous *Nostoc* cell material was broken up by grinding it under liquid nitrogen with a mortar and pestle before hot phenol extraction.

2.10.3. *Nostoc* **method B**

After the filamentous cell material was ground into powder as described in Section 2.10.2, RNA was extracted from cultured *Nostoc* using Trizol reagent (Invitrogen) and the manufacturer's protocol for polysaccharide-rich tissue. RNA recovered from Trizol preparation was further purified by two LiCl-mediated precipitations and one ethanol-mediated precipitation as described above for *S. elongatus*.

2.10.4. *P. membranacea*

RNA was obtained in the same manner as described in *Nostoc* method B (Section 2.10.3) except that the standard Trizol protocol was utilized.

2.11. RT-PCR

Aliquots of RNA were incubated with DNase I (amplification grade, Invitrogen) for 15 minutes at room temperature to remove any contaminating DNA prior to RT-PCR analysis. Samples were then heated to 65 °C for 10 minutes to inactivate DNase I before proceeding to the reverse transcription (RT) step. Each 20 μL RT reaction mixture contained 1 μg cyanobacterial or 3 μg lichen RNA, 50 ng random hexamers, 625 μM dNTP mix, 10 mM dithiothreitol, 1× first-strand buffer, 13.6 units RNAguard (RNase inhibitor isolated from human placenta) and 200 units Superscript II RNase H⁻ reverse transcriptase. Hexamers and RNAguard were obtained from Amersham Biosciences. First-strand buffer, Superscript II, and dNTP mix were obtained from Invitrogen. RT reactions were carried out in PTC-100 thermal cyclers (MJ Research) according to the Superscript II manufacturer's protocol. Completed RT reactions were treated with 100 ng RNase (DNase-free, isolated from bovine pancreas, Roche Applied Science) for 20 minutes at 37 °C.

Each 20 μL PCR reaction contained 0.2 μM primers, 225 μM dNTP mix, 1.5 mM MgCl$_2$, 1 unit recombinant *Taq* DNA polymerase, 1× PCR buffer and 1 μL RT reaction as template. DNA polymerase, PCR buffer, MgCl$_2$ and dNTP mix were obtained from Invitrogen. All amplification programs shared a 45 second template denaturation step at 92 °C and a 45 second primer annealing step at 56 °C, but incubation times for the 72 °C polymerization step

varied between 45 seconds and 2 minutes depending on the length of the expected product. The number of amplification cycles varied between 14 and 36 cycles, according to the rarity of the target and the purpose of the experiment. Negative RT-PCR controls included RT reactions to which no reverse transcriptase was added and PCR reactions in which no template was added. Genomic DNA was used as a positive control for PCR amplification. Reaction products were visualized by staining with ethidium bromide after electrophoresis on 1.0 % agarose gels in 0.5× TBE buffer.

2.12. Northern blot analysis

2.12.1. RNA gel blots

Cyanobacterial (5 μg/lane) and lichen (15 μg/lane) RNAs were subjected to electrophoresis on denaturing 0.22 M formaldehyde/1.2 % agarose gels. RNA size markers (0.24 – 9.5 Kb RNA ladder, 1 μg/lane) were obtained from Invitrogen. Gels were stained with ethidium bromide and photographed before RNA was transferred onto Hybond-N+ positively charged nylon membranes (Amersham Biosciences) using a VacuGene XL vacuum blot apparatus (Amersham Biosciences) according to the manufacturer's protocol No. 4 for vacuum transfer of RNA. RNA was fixed to membranes by exposure to 120,000 μJ/cm^2 ultraviolet radiation in a UVC 500 UV Crosslinker (Amersham Biosciences).

2.12.2. Probes

Radioactively labeled single-strand antisense DNA probes were synthesized using a Strip-EZ PCR probe synthesis and removal kit (Ambion). These probes were generated by linear amplification of the following templates: PCC 7942 *kaiA* 0.6 kb O-672/O-673 PCR product, PCC 7942 *kaiBC* 0.8 kb O-674/O-675 PCR product, PCC 7942 *rnpB* 0.36 kb O-663/O-664 PCR product, PCC 7942 16S 0.4 kb 16S.1100.F16/16S.1511.R21 PCR product, PCC 9707 *kaiA* 0.5 kb O-642/O-662 PCR product, PCC 9709 *kaiC* 1.4 kb O-496/O-651 PCR product, PCC 9709 *rnpB* 0.4 kb O-663/O-664 PCR product, PCC 9709 16S 0.4 kb 16S.1100.F16/16S.1511.R21 PCR product. Primers used for single strand probe amplification and label incorporation were the antisense primers (listed second) of the each oligonucleotide pair used to generate the templates listed above. Oligonucleotide sequences are provided in Table 2.2 and Table 2.3. Location of PCC 7942 *kai* primer sites / probe segments are shown in Figure 5.3.

Each 20 μL probe generation reaction contained 1× Strip-EZ PCR buffer, 1× Strip-EZ PCR nucleotide mix, 2 μL [α-^{32}P]dATP (3000 Ci/mmol, 10 mCi/mL, Amersham Biosciences),

1 unit recombinant *Taq* DNA polymerase (Invitrogen), 1 μM antisense primer, and 5-10 ng PCR product as template DNA. Amplification comprising 35 cycles of template denaturation for 30 seconds at 94 °C, primer annealing for 20 seconds at 55 °C, and polymerization for 90 seconds at 72 °C was carried out in PTC-100 thermal cyclers (MJ Research). After removal of unincorporated isotope on a ProbeQuant G-50 Micro Column (Amersham Biosciences), the radioactivity of each labeled probe was measured using a MicroBeta TriLux (Wallac) and utilized for hybridization experiments immediately.

2.12.3. Hybridization and signal detection

Hybridization and wash steps were carried out at 50 °C in a HB-1D hybridization oven (Techne). Membranes were prehybridized for 2 hours in ULTRAhyb hybridization buffer (containing 50% formamide, Ambion) at 50 °C before freshly-labeled probe was added to a final concentration in the range of 1 x 10^5 CPM/mL to 1 x 10^6 CPM/mL. After overnight probe hybridization, membranes were subjected to two 10 minute washes in 2× SSC/0.1 % SDS followed by two 20 minute washes in 0.1× SSC/0.1 % SDS. In some cases, additional washes in 0.1× SSC/0.1 % SDS at 50 °C or 55 °C were required to reduce background signal. Membranes were probed sequentially for *kaiA*, *kaiBC*, *rnpB* and *16S* targets (in order of expected least to most abundant mRNA levels). In between hybridizations, blots were subjected to a relatively gentle probe removal procedure, involving three 10 minute washes at 68 °C, according to the Strip-EZ protocol (Ambion). Membranes were exposed to BioMax MR film (Kodak). Exposure times varied from several days with intensifying screens for the less abundant *kai* transcripts to a few hours for the more abundant mRNA controls. After development, films were illuminated against white light and spot densities were measured using a AlphaImager 2000 3.3b digital imaging system (Alpha Innotech). Specifically, images were captured using saturation analysis and then inverted (to white signals on black background). Integrated density values were calculated from areas of equal size for each sample lane as described in the imaging guidelines.

2.13. Relative RT-PCR assay

All samples for a given time course experiment were prepared and assayed together as a set. For each sample, a RT reaction was performed immediately prior to PCR analysis. The same RT reaction was used as template for PCR amplification with *kaiA*, *kaiC* and *rnpB* primer sets; these three PCR assays were performed simultaneously in the same thermal cycler (each set was manually removed after the desired number of amplification cycles had occurred). A no-

RT control was included for each sample in addition to a no-template negative control and a genomic template positive control for each PCR assay. The entire 20 μL volume of each PCR reaction was loaded onto a 1.2 % agarose gel containing ethidium bromide. After electrophoretic separation of PCR products in 0.5× TBE buffer, gels were destained in water to reduce background fluorescence and then visualized by ultraviolet illumination in an AlphaImager 2000 3.3b digital imaging system (Alpha Innotech) for densitometric analysis. Specifically, images were captured using saturation analysis and integrated density values were calculated (using areas fitted to each band) as described in the imaging guidelines.

2.13.1. Calibration of amplification cycles

To determine the appropriate number of PCR amplification cycles to use for each mRNA target and set of time course samples, those RNA samples that were anticipated (according to the *S. elongatus* circadian model) to contain the greatest and least amounts of *kai* mRNA were pre-tested to generate calibration curves. These calibration curves consisted of PCR products amplified for varying numbers of cycles: four to five calibration points ranging between 15 and 35 cycles of amplification. Test conditions for relative PCR were chosen such that enough PCR product would be amplified from low mRNA samples to be detected on agarose gels whereas product accumulation would not have reached saturation for the high mRNA samples. RT-PCR was performed as described in Section 2.11 above. *S. elongatus*: Calibration curves for the PCC 7942 time course were generated using Hour 24 and Hour 36 samples with *kaiA* primer pair O-672/O-673, *kaiC* primer pair O-682/O-675 and *rnpB* primer pair O-663/O-664. These oligonucleotide primers are described in Table 2.3 and Figure 5.3. *Nostoc*: Calibration curves for the PCC 9709 time course were generated using Hour 12 and Hour 24 samples with *kaiA* primer pair O-685/O-677, *kaiC* primer pair O-686/O-678 and *rnpB* primer pair O-663/O-664. These oligonucleotide primers are described in Table 2.2 and Figure 4. *P. membranacea*: Specimens AL#21 and AL#24 were analyzed using Hour 12 samples with the same *kaiA*, *kaiC* and *rnpB* primer sets as described for *Nostoc* PCC 9709 above. For AL#21, the number of PCR amplification cycles performed prior to agarose gel quantification was 27 cycles for *kaiA*, 27 cycles for *kaiC*, and 20 cycles for *rnpB*, based on calibration curves performed on the Hour 12 sample. For AL#24, the number of PCR amplification cycles performed prior to agarose gel quantification was 32 cycles for *kaiA*, 32 cycles for *kaiC*, and 20 cycles for *rnpB*, based on calibration curves performed on the Hour 12 sample.

<div align="center">**CHAPTER THREE**</div>

<div align="center">**Circadian clock genes in *Nostoc* PCC 9709**</div>

3.1. Introduction

3.1.1. The *S. elongatus kaiABC* model

In *S. elongatus* PCC 7942, the central components of the circadian clock are encoded on a 2.9 kb region of the chromosome that contains a three gene cluster named *kaiABC*. These three genes are oriented in the same direction and do not overlap each other. Through mutational analysis studies, the KaiA, KaiB and KaiC proteins have been proven to function as the essential elements of the circadian oscillator (32). Although circadian rhythms have been observed in various members of the Cyanophyta (reviewed in (24)) and potential *kai* homologues have been identified in numerous organisms ((19), (47)), *S. elongatus* PCC 7942 remains the model organism for studying the molecular biology of cyanobacterial circadian systems due in large part to its amenability to genetic manipulation and rapid automated screening methods.

3.1.2. Research objectives

With organisms that are not as tractable as *S. elongatus* for laboratory studies – such as *Nostoc* PCC 9709, which grows at a very slow rate in axenic culture (several months to obtain a dense culture, generation time estimated to be on the order of days versus hours) – other, perhaps less direct methodological approaches must be undertaken. This chapter describes an assessment of the genetic potential for a circadian clock in *Nostoc* PCC 9709. This goal was divided into the following research objectives: 1) develop a PCR-based assay for *kai* sequences with degenerate primers, 2) recover a *kai* sequence or sequences from PCC 9709 using the PCR screening method, 3) determine the sequence of the intact gene(s) corresponding to the *kai* amplicon(s), and 4) perform a detailed comparison of any identified potential *Nostoc* clock gene(s) against the *S. elongatus kaiABC* archetype.

3.2. PCR-based search for circadian clock genes in cyanobacteria

PCR with degenerate primers is a powerful tool for finding new genes or gene families. Proteins with like functions will often share structural similarities. Regions of known or suspected conserved protein sequence can be reverse-translated to make degenerate primers. These primers can be used to search for matching gene sequences in any organism of interest. The advantages of this so-called "degenerate PCR" over older methods such as low-stringency DNA hybridization are that it requires very small amounts of genetic material, a large number of

samples can be screened simultaneously with ease, and DNA sequence information from potential positive 'hits' can be obtained in a relatively short time.

3.2.1. Design of degenerate PCR primers for *kaiC*

When this study was begun, there were only two sets of *kaiABC* sequences available through the GenBank database: the gene cluster identified by functional analysis in *S. elongatus* PCC 7942, which is the model organism for cyanobacterial circadian rhythms, and a putative homologous gene cluster identified by sequence similarity analysis of the *Synechocystis* PCC 6803 genome (39). Within our research group, J. Scheffer had developed degenerate oligonucleotide primers O-336, O-337 and O-338 (46, 47) for a PCR-based approach to identifying novel circadian clock genes in other cyanobacteria. The design of these PCR primers had been based upon an alignment of deduced KaiB and KaiC protein sequences from PCC 7942 and PCC 6803 as well as from putative *kaiC*-like genes identified in *Methanobacterium thermoautotrophicum* (66), *Pyrococcus horikoshii* (40) and *Archaeoglobus fulgidus* (43) genomes by sequence similarity analysis. Work done by J. Scheffer and J. Lorne had resulted in the generation of amplified sequences from seven cyanobacterial strains that had a high degree of similarity to each other as well as to *kaiC* of *S. elongatus* PCC 7942. However, this original primer set had not been able to produce amplicons of the expected size from one-fourth of the cyanobacterial strains tested (46), including *Nostoc* PCC 9709.

I designed a second set of oligonucleotide primers for amplification of circadian clock genes. The objective was to make primers that were capable of amplifying *kai* sequences from a more extensive range of cyanobacterial species. It was thought that the initial PCR-generated data set would identify conserved regions that could be targeted by the next-generation primers. Attention was focused on *kaiC* because there was more sequence information available for that gene than for *kaiA* or *kaiB*. The second generation PCR primers, named O-488 and O-489, were designed by reverse-translation of conserved regions inferred from an alignment of deduced KaiC or KaiC-like amino acid sequences (Figure 3.1) belonging to seven cyanobacteria: *Synechococcus* PCC 7942, *Synechocystis* PCC 6803, *Nodularia* PCC 73104, *Scytonema* PCC 7110, *Chroococcidiopsis* PCC 6712, *Xenococcus* PCC 7305, and *Geitlerinema* PCC 7105. These seven strains include members of four of the five taxonomic sections of cyanobacteria defined by Rippka et al. (58, 59) and thus were considered to be a diverse group of organisms suitable for the purpose of identifying protein regions conserved throughout the Cyanophyta.

```
                 15              30              45              60              75              90
PCC 7942   NNNSEHQAIAKMRTM IEGFDDISHGGLPIG RSTLVSGTSGTGKTL FSIQFLYNGIIEFDE PGVFVTFEETPQDII KNARSFGWDIAKLVD
PCC 6803   RPDVPRKGYQKIRTV IEGFDEITHGGLPIG RTTLVSGTSGTGKTL LAVQFLYQGIHHFDY PGLFIFEESPSDII  ENAYSFGWDLQQLID
PCC 73104  --NTQMAGVEKIRTM IEGFDDISHGGLPVG RTTLVSGTSGTGKTL FSLQFLYNGITRXDE P-------------  --------------
PCC 7110   --------------- --------------- --------------- --LQFFXNGISYFDE AGVFVTFEESPSDII KNACIFGWNIQRLID
PCC 6712   --------------- --------------- --------------- --------------- -VFVTFEESPKDII  TNARSFGWDIQSLID
PCC 7305   --------------- --------------- --------------- --------------- --------------- -AYSFGWDLQNLVE
PCC 7105   --------------- --------------- --------------- --------------- --------------- ------QGLFD

                                                ** ********  * *******

                 105             120             135             150             165             180
PCC 7942   EGKLFILDASPDPEG QEVVGGFDLSALIER INYAIQKYRARRVSI DSVTSVFQQYDASSV VRRELFRLVARLKQI GATTVMTTERIEEYG
PCC 6803   DGKLFILDASPDPEG QEVVGTFDLSALIER IQYAVRKYKAKLVSI DSVTAVFQQYDAASV VRREIFRLVARLKQL QVTSIMTTERVEEYG
PCC 73104  --------------- --------------- --------------- --------------- --------------- ---------------
PCC 7110   DGKLFILDASPDPEG QDVVGNFDLSALIER LQYAIXKYKAKRVSI DSITAVFQQYEAVGV VRREIFXLVARLKQI SVTTLITTERTDEYG
PCC 6712   QGKMFILDASPDPEG QEVVGNFDFSALIER IQYAINKYKAKLVSI DSVTAIFQQYDAASV VRREIFRLVARLKLI NVTSIMTTERVEEYG
PCC 7305   RGKLFILDASPDPEG QEVVGNFDLSALIER IQYAINKYKAKLVSI DSVIAVFQQYDAASV VRREIFRLVARLKQL KVTSIMTTERIEEYG
PCC 7105   EGKIVILEASPDPEG QDVVGNFDLSALIER IQYGIRKYKAKRVSI DSVTAVFQQQDAAPV IRREIERLVARLKQL GATTVMTTERIEEYG

                 195             210             225             240             255             265
PCC 7942   PIARYGVEEFVSDNV VILRNVLEGERRRRT LELLKLRGTSHMKGE YPFTITDHGINIFPL GAMRLTQRSSNVRVS SGVVRLDEMC
PCC 6803   PIARFGVEEFVSDNV VVLRNVLEGERRRRT VEILKLRGTTHMKGE YPFTITHDGINIFPL GAMRLTQRSSNARIS SGVQTLDEMC
PCC 73104  --------------- ------EGRRRRT   IEILKLRGTTHMKGE YPFTITNEGVNIFPL FAMRLTQRSSNIR-  --------------
PCC 7110   PVACFGVEEFVSDNV VIVRNVLEGERRRRT IELLKLRGTTHMKGE YPFTITNEGVNIFPL GAMRLTQRSS----  --------------
PCC 6712   PIARFGVEEFVSDNV IIIRNVLEGERRRRT IELLKLRGTTHMKGE YPFTIREDGINIFPL GAMRLTQRSSNVRPL LELR----
PCC 7305   PVARFGVEEFVSDNV VIVRNVIEERRRR-I IETLKLRGTTHMKGE YPFTMTDDGINIFPL GAMRLTQRSSNIRVS SGVKNLDEMC
PCC 7105   PVARFGVEEFVSDNV VIVRNVLEGERRRRT IELLKLRGTTHMKGE YPFTMPNNGINIFPL GAMRLTQRSK----  --------------

                                ****** *****  *****
```

Figure 3.1. Translated amino acid alignment used to design PCR primers O-448 and O-449. The aligned sequences were determined by J. Scheffer and J. Lorne as described in (46). Conserved residues s are highlighted in grey. Primer O-448 was designed by reverse-translation of the consensus sequence at positions 14 through 22 and primer O-449 was designed by reverse-translation of the consensus sequence at positions 221 through 229 (denoted by asterisks). This alignment corresponds to amino acid residues 10 through 274 of the *S. elongatus* PCC 7942 deduced KaiC protein sequence. Other organisms are identified as follows: PCC 6803 = *Synechocystis*, PCC 73104 = *Nodularia*, PCC 7110 = *Scytonema*, PCC 6712 = *Chroococcidiopsis*, PCC 7305 = *Xenococcus*, PCC 7105 = *Geitlerinema*.

Primer locations and sequences were chosen such that 1) primer sequence degeneracy is minimized and oligonucleotide lengths are sufficiently extended to provide specificity, and 2) amplicon size is maximized to provide the largest amount of sequence information. O-488 has 2048-fold sequence degeneracy and O-489 has 512-fold sequence degeneracy; both primers are 26 nucleotides in length (Table 2.3) and together they were designed to amplify a 650 bp region located within the first half of the *kaiC* gene (Figure 3.2a). This primer pair has subsequently been demonstrated to successfully amplify *kaiC*-like PCR products from 36 of 36 cyanobacterial strains tested, including *Nostoc* PCC 9709 (47).

3.2.2. Detection of a single *kaiC* amplicon from *Nostoc* PCC 9709

Primers O-488 and O-489 were used to initiate this study of circadian clock genes in *Nostoc* PCC 9709. PCR amplification of PCC 9709 genomic DNA with the O-488/O-489 primer set resulted in a single major product band of the expected size upon electrophoretic gel analysis (Figure 3.2b). Although not visible in Figure 3.2b, minor products of smaller sizes (usually 550 bp) were occasionally observed after amplification of PCC 9709 and other cyanobacteria. It is possible that these products may arise from poorly-efficient cross-priming events with related DNA sequences such as the structurally-related second domain of the *kaiC* gene or even *kaiC* gene duplications that may exist within the PCC 9709 genome (19); it is also possible that these products arise from accidental cross-priming events with unrelated DNA sequences. Because these side products appeared inconsistently and were relatively low in amount, they were considered spurious and not further investigated.

Careful of the possibility that the O-488/O-489 PCR product might contain amplicons from multiple *kaiC* genes that may exist within the PCC 9709 genome, the DNA sequence identity of the 650 bp product band was assessed using several approaches. The size-isolated population of amplicons was directly subjected to DNA sequencing. In addition, PCR products were ligated into a plasmid vector and one of the cloned inserts was subjected to DNA sequencing. The direct sequence data does not show evidence of a mixed template population and is identical to that of the cloned amplicon named pKN10A1. Sixteen other clones were examined by means of a PCR amplification test with primers O-495 and O-496 (which are specific to the pKN10A1 insert) and all were positively determined to carry the same insert sequence. Based on these results, it is concluded that the 650 bp product band amplified from *Nostoc* PCC 9709 by the O-488/O-489 primer pair corresponds to a single amplicon sequence.

A

B

L: 1 Kb Plus DNA Ladder

1: pJS1

2: *Nostoc* PCC 9709 genomic DNA

3: Genomic library λZAP:773

4: *P. membranacea* TDI#AR95 genomic DNA

5: No-template control

C

```
PCC 9709   SHGGLPIGRTTLISGTSGTGKTLFSLQFLYNGITYFDEAGVFVTFEESPS   50
PCC 7942   SHGGLPIGRSTLVSGTSGTGKTLFSIQFLYNGIIEFDEPGVFVTFEETPQ   50
           ********.**.************.*******  ***.********.*

PCC 9709   DIIKNAHVFGWNLPRLIEEGKLFILDASPDPEGQDIVGNFDLSALIERLQ   100
PCC 7942   DIIKNARSFGWDLAKLVDEGKLFILDASPDPEGQEVVGGFDLSALIERIN   100
           ******.  ***  *  .*..*************..**.*********..

PCC 9709   YAIRKYKAKRVSIDSITAVFQQYEAMGVVRREIFRLVARLKLLNVTTVIT   150
PCC 7942   YAIQKYRARRVSIDSVTSVFQQYDASSVVRRELFRLVARLKQIGATTVMT   150
           ***.**.*.******.*.*****.*   *****.********  .***.*

PCC 9709   TERGEEYGPVASFGVEEFVSDNVVIVRNVLEGERRRRTIEILKLRGTT   198
PCC 7942   TERIEEYGPIARYGVEEFVSDNVVILRNVLEGERRRRTLEILKLRGTS   198
           ***.*****.*  .************.***********.*******.
```

Figure 3.2. Amplification of a *kaiC*-like sequence from *Nostoc* PCC 9709. A) Location and orientation of primer O-488 and O-489 binding sites. B) Electrophoretic separation of completed O-488/O-489 PCR reactions on a 1 % agarose gel shows 650 bp products amplified from all three samples (lanes 2-4) containing PCC 9709 genetic material. Plasmid pJS1 was used as a positive control for amplification. C) Deduced amino acid sequence of the PCC 9709 amplicon aligns to residues 32 through 229 of the *S. elongatus* KaiC protein. Asterisks mark residues that are invariant. Dots mark residues that are chemically similar. A conserved P-loop motif (GXXXXGKT/S) is indicated in reversed font.

In a similar fashion, I addressed the possibility that the PCR product sequence had not been amplified from the target genome but from sample contaminants. The use of PCR methodology is more vulnerable to this type of false positive result than other techniques (such as Southern hybridization) which do not depend on exponential signal amplification. The sequence of the *Nostoc* PCC 9709 *kaiC* PCR amplicon was determined by cross-confirmation from three related DNA sources: 1) genomic DNA from the cyanobacterium itself, 2) a genomic DNA sample of the lichen specimen TDI#AR95 from which the PCC 9709 strain had originally been isolated (51), and 3) lysate of a bacteriophage genomic library of *Nostoc* PCC 9709. As shown in Figure 3.2b, the expected 650 bp PCR product was obtained from each of the templates upon PCR amplification with the *kaiC* primer set. The products of all three amplifications were directly subjected to DNA sequence analysis and found to be identical. Given that the DNA sources had been prepared by three different workers on separate days, this result strongly suggests that the amplified sequence originates from the *Nostoc* PCC 9709 genome (a component shared by all three materials) and not from a chance contaminant that entered into a DNA sample.

BLASTX analysis of the *Nostoc* PCC 9709 *kaiC* amplicon against the GenBank database established that it was a novel sequence. Comparison of the PCR product to *kaiC* of *S. elongatus* PCC 7942 revealed a high degree of sequence similarity. Nucleotide alignment of the *Nostoc kaiC* PCR amplicon against the segment of the *S. elongatus kaiC* gene bracketed by the O-488/O-489 primer set exhibits 70 % sequence identity, with (mis)matches distributed evenly throughout the 595 bp region (Appendix A). The GC content of the PCC 9709 amplicon is 41.2 %, which is consistent with its *Nostoc* origin. Through genome sequencing, two closely related organisms, *Anabaena/Nostoc* sp. PCC 7120 (38) and *Nostoc punctiforme* ATCC 29133 (50), are known to have a GC content of 41.3 % and 41.5 %, respectively. The GC content of the family *Nostocaceae*, which includes the *Nostoc* and *Anabaena* genera, ranges between 38 % and 47 % (30). Translation of the *Nostoc kaiC* amplicon sequence in the +2 reading frame produces a 198 residue deduced amino acid sequence that is not interrupted by termination codons. Alignment of this partial putative protein sequence against the PCR-targeted segment of the *S. elongatus* KaiC polypeptide demonstrates 79 % sequence identity (Figure 3.2c). The higher percentage of amino acid (versus 70 % nucleotide) matches is consistent with conservation at the level of functional protein sequence. An ATP/GTP-binding P-loop motif (GXXXXGKT/S) that has been

demonstrated to be important to KaiC function in *S. elongatus* (55) is conserved in the *Nostoc* amplicon sequence.

In summary, a PCR-based screen for cyanobacterial circadian clock genes was successfully designed and application of this method produced a single amplicon from *Nostoc* PCC 9709. Analysis of the novel product sequence demonstrated that it is very similar to a portion of *S. elongatus kaiC*. These results were sufficiently encouraging to support further exploration for *kai* genes in the *Nostoc* genome.

3.3. Identification of a *kaiABC* gene cluster in *Nostoc* PCC 9709

The following research objective was to assess whether the Nostoc *kaiC* amplicon sequence was derived from an intact *kaiC* gene and whether it lay next to additional circadian clock genes on the chromosome. DNA sequence information for genomic regions flanking a known segment can be obtained using a number of approaches, which may in principle include the use of inverse PCR and/or genomic libraries. For this thesis project, the application of DNA hybridization methodology complements the PCR-based approach used in Section 3.2.

3.3.1. Screening a genomic library

The novel *kaiC* amplicon was used to probe a lambda bacteriophage genomic library of *Nostoc* PCC 9709 by DNA hybridization as described in Section 2.4. A primary screen of 100,000 library clones resulted in the recovery of 14 clones that appeared positive. The hybridization assay was capable of detecting less than 1 pg probe DNA, which was found to be an adequate level of sensitivity. Positive hybridization signals were strong, clearly defined ("background noise" from non-specific binding of probe to membranes was negligible) and reproduced on duplicate plaque lifts. As a supplementary means of assessment, hybridization-positive plaques were tested by PCR amplification with the *kaiC* O-488/O-489 primer pair after both the primary and secondary (purification) screens. Although four plaques (kai2B-3, kai2C-1, kai2D-5 and kai2K-1) were not positive by the PCR test, phagemids were generated for all 14 clones to facilitate a more detailed map analysis of the inserts.

3.3.2. Analysis of recovered library clones

Mapping analyses showed that the 14 library clones represented 11 overlapping inserts ranging from 1.5 to 7 kb in length. As diagrammed in Appendix B, the position of the probe region within each insert, size of insert and insert orientation within the phagemid vector was determined by a combination of restriction analysis (Section 2.5) and PCR using various
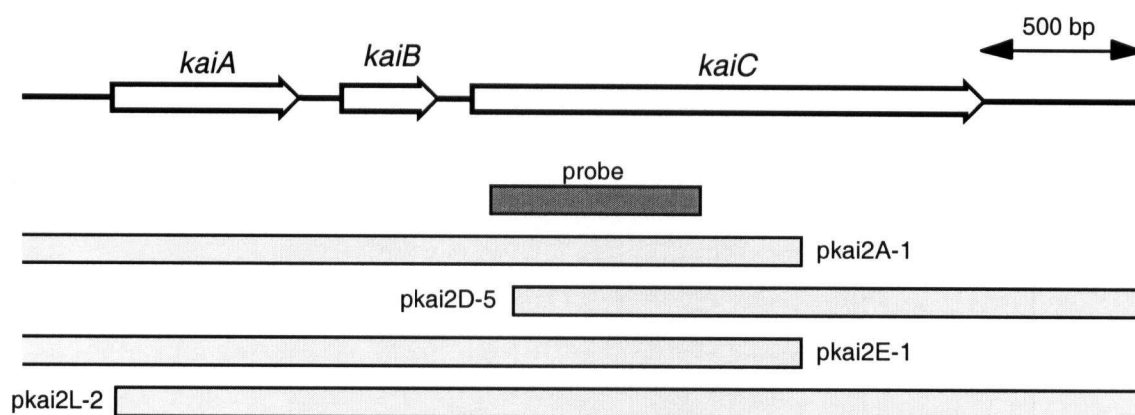
Figure 3.3. Alignment of four library clones recovered by hybridization to *kaiC* amplicon. Overlapping library clones pkai2A-1, pkai2D-5, pkai2E-1 and pkai2L-2 were recovered from a bacteriophage library by hybridization to a PCR-amplified *kaiC*-like sequence (probe). The relative position of the *kaiABC* gene cluster (identified through sequence analysis of the clones) is also shown.

pairings of vector- and probe-specific primers (Section 2.6.2). These methods confirmed that all isolated clones contained at least a portion of the *kaiC* amplicon sequence used to probe the library. One of the 11 inserts (represented by clones kai2B-3 and kai2C-1) was chimeric, being the combination of a *kai* sequence and one other small *Sau3A* I-digested fragment ligated together into the lambda bacteriophage vector (map analysis of chimeric clone confirmed by DNA sequence analysis).

Clones kai2A-1, kai2D-5, kai2E-1, and kai2L-2 were chosen for DNA sequence analysis because the overlap of these four inserts provided the greatest coverage of the genomic region surrounding the probe sequence. The clones provided multiple independent verifications of the *Nostoc* genomic sequence so that artifacts – such as the chimeric insert found in clone kai2B-3 – could be recognized and disregarded. Two rounds of primer-walking outwards from the probe segment in both directions served to define the bounds of a locus on the *Nostoc* chromosome containing three open reading frames comparable to the *kaiABC* cluster of *S. elongatus* with respect to sequence and organization (Figure 3.3). The DNA sequence provided in Figure 3.4, representing a 3.4 kb region of the *Nostoc* PCC 9709 genome, is compiled from the initial primer-walking data as well as from additional confirmatory analyses on coding and noncoding strands of all four library clones (Section 2.8.2).

3.4. Characterization of the *Nostoc* PCC 9709 *kaiABC* gene cluster

The 3.4 kb genomic sequence surrounding the 0.65 kb amplicon/probe segment contains three consecutive open reading frames in the same orientation. These open reading frames have been named *kaiA*, *kaiB* and *kaiC* based on sequence similarity to the *S. elongatus kai* genes (Figure 3.5a). The *kai* gene cluster in *Nostoc* PCC 9709 is arranged in the same pattern as in *S. elongatus*.

3.4.1. *kaiA*

The open reading frame designated *kaiA* is 582 nucleotides in length, encoding a predicted 193 amino acid protein with a calculated molecular mass of 22.4 kD. This gene is two-thirds the length of the 855 bp *S. elongatus kaiA*. No protein motifs were identified in the putative *Nostoc* KaiA sequence (or *S. elongatus* KaiA) when compared to Swiss-Prot databases. Unlike the other two genes in the cluster, polypeptide sequence similarity to *S. elongatus* KaiA is distributed unevenly and is observed only at the C-terminal region. At shown in Figure 3.5b, the last 98 amino acids of the putative *Nostoc* protein shares 46 % identity and 72 % similarity

Figure 3.4 (two pages)

```
TCTATCTATTTATATACATAAATATGCAAATAATAAATTAGCTGCTGTGTCTTGCTCTGCCGACAGATCC        70
CTCATCATCAAAGACCTACTAAGTTCTACTACTCAGTTGCTGCACTCTTGTTGTTGAGTGGCTATCATTA       140
TCTGCAAACACAGAGCGTTATTTTGGGGTGGAATCACCAGTTTACTGCACACCCAGTATTTCCCTAGAAG       210
CGTTAGCCAAACTTATCAGCCCACCTTTCTTGGTCAGGGCGATCGAGGAGATATATTCAAGAGTAAATGT       280
                                                                  M  L
```

```
TATTACCGATCTTGATTCTGCCACCTGATGTTAATAAAAGTTTAAATAATCTAGTTGACTTAGCCAATCA       350
  L  P  I  L  I  L  P  P  D  V  N  K  S  L  N  N  L  V  D  L  A  N  Q
```

```
AAAAGCTTTGATCCCGTGGGTGTGGAAATTGATTGACGCAACTATTTCCCAATTTTCATCCCTACCTGTT       420
  K  A  L  I  P  W  V  W  K  L  I  D  A  T  I  S  Q  F  S  S  L  P  V
```

```
GCTGCAACTACCTCTGAAAAAATCAACTATTTACCAAATTGGCTGCAACAAAGTCCAAGTAAGGCATACA       490
  A  A  T  T  S  E  K  I  N  Y  L  P  N  W  L  Q  Q  S  P  S  K  A  Y  T
```

```
CTGGCAAATACGTCTATGTGTTTGCCAGTCAGATGCAAAAAAGCCAACAGCATTTACAGGAGATGACTCC       540
  G  K  Y  V  Y  V  F  A  S  Q  M  Q  K  S  Q  Q  H  L  Q  E  M  T  P
```

```
AGCCGAAAGGCAAGGATTATTAAGACAGCTTAAATTAGATTATAGCCTGATTCTTATAGATTATTTTACC       630
  A  E  R  Q  G  L  L  R  Q  L  K  L  D  Y  S  L  I  L  I  D  Y  F  T
```

```
ACAGATAAAACACTCAAAGATAAAATTGATAAATTTATCAATACTATATTTTATGCTAATATTCCTGTGC       700
  T  D  K  T  L  K  D  K  I  D  K  F  I  N  T  I  F  Y  A  N  I  P  V  P
```

```
CCCAAATAATCGAAATTCACATGGAGGTAATTGAAGATTTTTCTAACCAGCTAAAATTAGAAGGAAGGAG       770
  Q  I  I  E  I  H  M  E  V  I  E  D  F  S  N  Q  L  K  L  E  G  R  S
```

```
CAATGAAACGTTACTGGATTACCGTCTAACGTTAATAGATATCCTGGCTCACCTGTGCGAACTCTATCGG       840
  N  E  T  L  L  D  Y  R  L  T  L  I  D  I  L  A  H  L  C  E  L  Y  R
```

```
AGTTCGATTTCTAAATAATTAAATTTATCGATGTATATTCAACTGGGAACAAGCTACTGCTGTTAGTGAC       910
  S  S  I  S  K  *
```

```
TACCAATGCGCGAAAGGTCTTGTATTTCTATGATCGTGACTAGCCATTAAATATATGAATAAAGCCAGAA       980
                                                        M  N  K  A  R  K
```

```
AAACCTACGTTCTCAAGCTTTACGTAGCAGGGAACACACCTAATTCAGTCCGGGCATTAAAAACACTCAA      1050
  T  Y  V  L  K  L  Y  V  A  G  N  T  P  N  S  V  R  A  L  K  T  L  K
```

```
GGATATTTTAGAACAGGAGTTTGAAGGTGTTTATGCTTTAAAAGTGATCGATGTCCTGAAAAGCCCGCAA      1120
  D  I  L  E  Q  E  F  E  G  V  Y  A  L  K  V  I  D  V  L  K  S  P  Q
```

```
CTGGCAGAAGAAGATAAAATATTGGCGACGCCAACATTATCTAAAATTTTGCCTCCACCCGTTCGCAAAA      1190
  L  A  E  E  D  K  I  L  A  T  P  T  L  S  K  I  L  P  P  P  V  R  K  I
TTATCGGGGATCTTTCAGATAGAGAAAGAGTATTGATTGGATTAGATTTGCTCTATGAAGAACTGAGTGA      1260
  I  G  D  L  S  D  R  E  R  V  L  I  G  L  D  L  L  Y  E  E  L  S  E
AGAAGATTTTGAATAATCCGATTTTAATCATAAAAAATTGAGTACAAACTTTAGTAATAAAAAAACAGGT      1330
  E  D  F  E  *
```

```
TTTAATACCAATTGATTATCAAGCAATGATTGAAAACGAGCAAGTAGAACCAAAGCAAACACCGATAATT      1400
                              M  I  E  N  E  Q  V  E  P  K  Q  T  P  I  I
```

```
AGGGGTGTAGAAAAAAATTCGTACGATGATCGAAGGGTTTGACGATATTAGTCATGGTGGTTTACCAATTG      1470
  R  G  V  E  K  I  R  T  M  I  E  G  F  D  D  I  S  H  G  G  L  P  I  G
```

```
GTAGAACTACCTTGATCAGTGGTACCTCCGGCACAGGCAAAACTTTATTCTCTCTTCAGTTTCTCTATAA      1540
  R  T  T  L  I  S  G  T  S  G  T  G  K  T  L  F  S  L  Q  F  L  Y  N
```

```
CGGTATCACCTACTTTGATGAAGCAGGAGTATTTGTTACCTTTGAAGAATCACCCAGTGATATTATTAAA      1610
  G  I  T  Y  F  D  E  A  G  V  F  V  T  F  E  E  S  P  S  D  I  I  K
```

```
AATGCCCATGTTTTTGGTTGGAACTTGCCACGCCTAATTGAAGAAGGCAAGTTATTTATTCTTGATGCAT      1680
  N  A  H  V  F  G  W  N  L  P  R  L  I  E  E  G  K  L  F  I  L  D  A  S
```

```
CTCCCGATCCAGAAGGTCAAGATATCGTTGGTAATTTTGACCTTTCTGCACTCATTGAACGCTTGCAATA      1750
  P  D  P  E  G  Q  D  I  V  G  N  F  D  L  S  A  L  I  E  R  L  Q  Y
```

```
TGCCATCCGTAAATACAAAGCTAAACGAGTTTCAATCGACTCAATAACAGCAGTATTTCAGCAGTATGAA      1820
  A  I  R  K  Y  K  A  K  R  V  S  I  D  S  I  T  A  V  F  Q  Q  Y  E
```

Figure 3.5 (two pages)

```
A   M   G   V   V   R   R   E   I   F   R   L   V   A   R   L   K   L   L   N   V   T   T   V
TAATTACCACTGAACGTGGTGAAGAATATGGGCCTGTTGCCTCTTTCGGAGTAGAAGAATTTGTTTCTGA   1960
    I   T   T   E   R   G   E   E   Y   G   P   V   A   S   F   G   V   E   E   F   V   S   D
TAATGTAGTAATTGTTCGCAACGTTTTAGAAGGAGAACGTCGCCGTCGCACAATTGAAATTCTCAAGTTG   2030
    N   V   V   I   V   R   N   V   L   E   G   E   R   R   R   R   T   I   E   I   L   K   L
CGCGGGACAACTCACATGAAAGGCGAATATCCCTTCACGATTACTAACGAAGGAGTTAACATCTTCCCAC   2100
    R   G   T   T   H   M   K   G   E   Y   P   F   T   I   T   N   E   G   V   N   I   F   P   L
TGGGAGCAATGCGCTTGACTCAACGATCTTCTAATGTCAGGGTATCTTCTGGTGTCAAAACCTTAGATGA   2170
    G   A   M   R   L   T   Q   R   S   S   N   V   R   V   S   S   G   V   K   T   L   D   E
AATGTGCGGTGGTGGTTTCTTTAAAGATTCAATTATCTTGGCAACAGGAGCCACCGGTACTGGCAAAACC   2240
    M   C   G   G   G   F   F   K   D   S   I   I   L   A   T   G   A   T   G   T   G   K   T
TTGCTAGTCAGTAAGTTTATTCAAGATGGCTGCCTGAATGGAGAACAGGCAATATTATTTGCTTATGAAG   2310
    L   L   V   S   K   F   I   Q   D   G   C   L   N   G   E   Q   A   I   L   F   A   Y   E   E
AATCACGCGCTCAACTATCTCGTAATGCTTCTTCTTGGGGAATTGATTTTGAAGAATTAGAAGATCAAGG   2380
    S   R   A   Q   L   S   R   N   A   S   S   W   G   I   D   F   E   E   L   E   D   Q   G
TTTACTCAAAATAATCTGTACCTATCCCGAATCAACAGGTTTAGAAGACCACTTACAAATTATTAAATCA   2450
    L   L   K   I   I   C   T   Y   P   E   S   T   G   L   E   D   H   L   Q   I   I   K   S
GAAATTGCTATCTTCAAACCAGCTCGCATTGCCATTGATTCCCTTTCAGCACTAGCTAGAGGAGTAAGTA   2520
    E   I   A   I   F   K   P   A   R   I   A   I   D   S   L   S   A   L   A   R   G   V   S   N
ATAATGCATTTCGGCAGTTTGTAATTGGTGTAACGGGTTATGCTAAACAAGAAGAAATTACTGGTTTCTT   2590
    N   A   F   R   Q   F   V   I   G   V   T   G   Y   A   K   Q   E   E   I   T   G   F   F
TACCAACACAACTGACCAATTTCTAGGAGCGCATTCGATTACTGACTCTCATATCTCCACGATTACCGAT   2660
    T   N   T   T   D   Q   F   L   G   A   H   S   I   T   D   S   H   I   S   T   I   T   D
ACCATTCTAATGTTACAGTACGTAGAAATTCGCGGAGAAATGTCGCGGGCAATTAACGTGTTCAAAATGA   2730
    T   I   L   M   L   Q   Y   V   E   I   R   G   E   M   S   R   A   I   N   V   F   K   M   R
GAGGCTCTTGGCATGATAAGGGGATTCGTGAGTATAATATTACTGCTGACGGGCCCGATATTAAAGATTC   2800
    G   S   W   H   D   K   G   I   R   E   Y   N   I   T   A   D   G   P   D   I   K   D   S
TTTCCGAAACTACGAACGGATTATCAGCGGTGCTCCTACTCGCGTTAGTATCGATGAAAAGGCGGAACTT   2870
    F   R   N   Y   E   R   I   I   S   G   A   P   T   R   V   S   I   D   E   K   A   E   L
TCTCGCATTGTTAGACGTTTTGAAGACAAACAGAGTTCCGAACCC**TAA**ATTTAGTATCGTAGTATATTCT   2940
    S   R   I   V   R   R   F   E   D   K   Q   S   S   E   P   *
TTCCTAAGTTTAGTATCGTGCTATATTCTGTGGTGAAGAAAGGTTTTCTGCCGCTAGATTGATTTTCGTG   3010
TGAGGGATGCGGTGAAAGGACAGCAATTATTTTATAGCTTCTTACCTGGTGTAACGGCAGCGGTCTTAAC   3080
AACTCAGCCTGCTTGGGCTGGTACTGCAAAACTAACTGAGGTACAACTGGCGTCTTCTCCTAGTGTTTTG   3150
ACTTCTACTTATGGTCAAGACTTGGTTGTGGACACCATGAATACGCAACTGCCTAAGGGTGCAGATGTTA   3220
GTGTTCCAACTCTAGTACCTGCTTTTGGTTTTACTAAACTTAGTGTGAAGCCTTTAAGTAATAACAGTAT   3290
TCCAGTATTCTCGACTGAAAATACTGTTGTGCCAATAAAACAAGTACTTAAGAAAGATGAAGGTAGATTT   3360
TTCAGTTTGACACCTACCTCTAATCCTTCCCAACAGCTTG   3400
```

Figure 3.4. Annotated DNA sequence of the Nostoc PCC 9709 *kaiABC* gene cluster.
Deduced KaiA, KaiB and KaiC polypeptide sequences lie under the corresponding nucleotide segments. Putative start codons are highlighted in grey. Putative stop codons are shown in bold and indicated by asterisks. Possible purine-rich Shine-Dalgarno motifs are underlined. The GTG start codon of a partial open reading frame downstream of *kaiC* is highlighted in grey. The *Nostoc kaiABC* sequence was deposited in the GenBank database (accession number AY373437).

A

Figure 3.5 (two pages)

A

PCC 7942

kaiA   kaiB   kaiC

PCC 9709

kaiA   kaiB   kaiC

500 bp
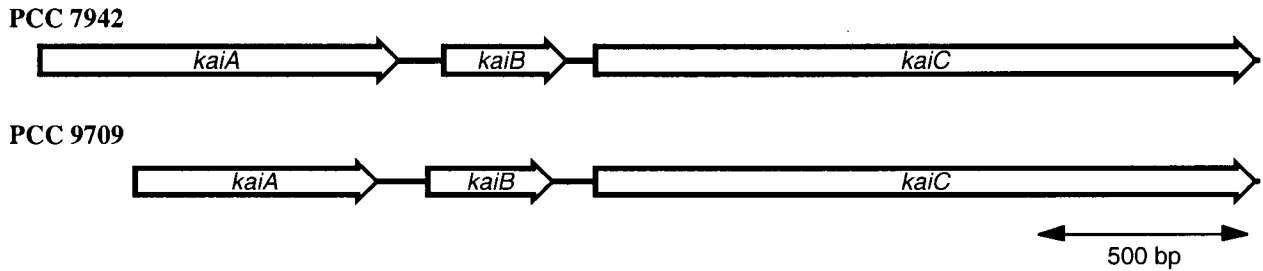
B

```
                    10        20        30        40        50        60
PCC 9709 kaiA
PCC 7942 kaiA   MLSQIAICIWVESTAILQDCQRALSADRYQLQVCESGEMLLEYAQTHRDQIDCLILVAAN


                    70        80        90        100       110       120
PCC 9709 kaiA                          MLLPILILPPDVNKSLNNLVDLANQKALIPWVWKLIDA
PCC 7942 kaiA   PSFRAVVQQLCFEGVVVPAIVVGDRDSEDPDEPAKEQLYHSAELHLGIHQLEQLPYQVDA
                                                    *         *        .  .**


                    130       140       150       160       170       180
PCC 9709 kaiA   TISQFSSLPVAATTSEKI-----NYLPNW--LQQSPSKAYTGKYVYVFASQMQKSQQHLQ
PCC 7942 kaiA   ALAEFLRLAPVETMADHIMLMGANHDPELSSQQRDLAQRLQERLGYLGVYYKRDPDRFLR
                ...*     *     *  .. *     *. *     *. ..    *.      .   .*.


                    190       200       210       220       230       240
PCC 9709 kaiA   EMTPAERQGLLRQLKLDYSLILIDYFTTDKTLKDKIDKFINTIFYANIPVPQIIEIHMEV
PCC 7942 kaiA   NLPAYESQKLHQAMQTSYREIVLSYFSPNSNLNQSIDNFVNMAFFADVPVTKVVEIHMEL
                 .    * * * . ..   *   *.. **.    .*   ** *.*  *.* .** ...*****.


                    250       260       270       280
PCC 9709 kaiA   IEDFSNQLKLEGRSNETLLDYRLTLIDILAHLCELYRSSISK
PCC 7942 kaiA   MDEFAKKLRVEGRSEDILLDYRLTLIDVIAHLCEMYRRSIPRET
                ...*.  .*..****  . *********..*****.**  **  .
```

C

```
                    10        20        30        40        50        60
PCC 9709 kaiB   MNKARKTYVLKLYVAGNTPNSVRALKTLKDILEQEFEGVYALKVIDVLKSPQLAEEDKIL
PCC 7942 kaiB   MSP-RKTYILKLYVAGNTPNSVRALKTLKNILEVEFQGVYALKVIDVLKNPQLAEEDKIL
                *    ****.******************* *** **.************ *********


                    70        80        90        100
PCC 9709 kaiB   ATPTLSKILPPPVRKIIGDLSDRERVLIGLDLLYEELSEEDFE
PCC 7942 kaiB   ATPTLAKVLPLPVRRIIGDLSDREKVLIGLDLLYGELQDSDDF
                *****.*.** ***.*********.********* **  .  *
```

Figure 3.5 (two pages)

D

```
                    10        20        30        40        50        60
PCC 9709 kaiC   MIENEQVEPKQTPIIRGVEKIRTMIEGFDDISHGGLPIGRTTLISGTSGTGKTLFSLQFL
PCC 7942 kaiC   MTSAEMTSPNNNSEHQAIAKMRTMIEGFDDISHGGLPIGRSTLVSGTSGTGKTLFSIQFL
                *    *   *  ..   . . *.****************** **.*********** .***

                    70        80        90       100       110       120
PCC 9709 kaiC   YNGITYFDEAGVFVTFEESPSDIIKNAHVFGWNLPRLIEEGKLFILDASPDPEGQDIVGN
PCC 7942 kaiC   YNGIIEFDEPGVFVTFEETPQDIIKNARSFGWDLAKLVDEGKLFILDASPDPEGQEVVGG
                ****   *** *******.* ******. *** *  .*..***************..**.

                   130       140       150       160       170       180
PCC 9709 kaiC   FDLSALIERLQYAIRKYKAKRVSIDSITAVFQQYEAMGVVRREIFRLVARLKLLNVTTVI
PCC 7942 kaiC   FDLSALIERINYAIQKYRARRVSIDSVTSVFQQYDASSVVRRELFRLVARLKQIGATTVM
                *********..***.**.*.******.*.*****.*    *****.********  .***.

                   190       200       210       220       230       240
PCC 9709 kaiC   TTERGEEYGPVASFGVEEFVSDNVVIVRNVLEGERRRRTIEILKLRGTTHMKGEYPFTIT
PCC 7942 kaiC   TTERIEEYGPIARYGVEEFVSDNVVILRNVLEGERRRRTLEILKLRGTSHMKGEYPFTIT
                ****.*****.* .***********.************.*******.***********

                   250       260       270       280       290       300
PCC 9709 kaiC   NEGVNIFPLGAMRLTQRSSNVRVSSGVKTLDEMCGGGFFKDSIILATGATGTGKTLLVSK
PCC 7942 kaiC   DHGINIFPLGAMRLTQRSSNVRVSSGVVRLDEMCGGGFFKDSIILATGATGTGKTLLVSR
                *.*******************************************************.

                   310       320       330       340       350       360
PCC 9709 kaiC   FIQDGCLNGEQAILFAYEESRAQLSRNASSWGIDFEELEDQGLLKIICTYPESTGLEDHL
PCC 7942 kaiC   FVENACANKERAILFAYEESRAQLLRNAYSWGMDFEEMERQNLLKIVCAYPESAGLEDHL
                *..   *  * *.**************. *** ***.****.* *.****.*.****.******

                   370       380       390       400       410       420
PCC 9709 kaiC   QIIKSEIAIFKPARIAIDSLSALARGVSNNAFRQFVIGVTGYAKQEEITGFFTNTTDQFL
PCC 7942 kaiC   QIIKSEINDFKPARIAIDSLSALARGVSNNAFRQFVIGVTGYAKQEEITGLFTNTSDQFM
                *******  ********************************************* ****.***.

                   430       440       450       460       470       480
PCC 9709 kaiC   GAHSITDSHISTITDTILMLQYVEIRGEMSRAINVFKMRGSWHDKGIREYNITADGPDIK
PCC 7942 kaiC   GAHSITDSHISTITDTIILLQYVEIRGEMSRAINVFKMRGSWHDKAIREFMISDKGPDIK
                *****************..*********************** ***. *.  *****

                   490       500       510       520
PCC 9709 kaiC   DSFRNYERIISGAPTRVSIDEKAELSRIVRRFEDKQSSEP
PCC 7942 kaiC   DSFRNFERIISGSPTRITVDEKSELSRIVRGVQEK-GPES
                *****.******.***...***.******  ..*    *
```

Figure 3.5. Alignment of *Nostoc* and *S. elongatus* KaiA, KaiB and KaiC. A) Diagrammatic alignment of the gene clusters, B) alignment of deduced KaiA polypeptide sequences, C) alignment of deduced KaiB polypeptide sequences, D) alignment of deduced KaiC polypeptide sequences. Identical and similar amino acids are indicated by asterisks and periods, respectively. Two conserved P-loop motifs (GXXXXGKT/S) located within KaiC are displayed in reverse font, two conserved putative imperfect Walker's motif Bs (ZZZZD) are boxed and two conserved putative catalytic carboxylate glutamate residues (positions 77 and 318) are highlighted in grey.

with the 100 amino acids at the C-terminus of *S. elongatus* KaiA, but sequence conservation is not apparent between the first 95 amino acids of *Nostoc* KaiA and the first 184 amino acids of *S. elongatus* KaiA. The possibility that the lack of N-terminal sequence similarity is due to a sequencing error or mutation in a cloned insert is unlikely for a number of reasons. First, the *kaiA* region of the *Nostoc* genomic sequence was determined from three independently obtained clones (Figure 3.3) and sequence analysis of coding and non-coding template strands from all three clones were identical. Secondly, BLASTX analysis of the first 700 bases of the compiled sequence in Figure 3.4 did not reveal protein sequence similarity to the N-terminal region of *S. elongatus* KaiA in any of the three possible reading frames. Finally, the putative protein sequence of the *Nostoc* KaiB is highly similar to a hypothetical protein identified from genome sequence analysis of the closely-related organism *Nostoc punctiforme* ATCC 29133 (GenBank accession ZP_00107016, see also (50)), being of the same length and sharing 85 % identity and 90 % sequence similarity. Further discussion of KaiA sequence conservation and function will be continued in Section 3.5.

3.4.2. *kaiB*

The open reading frame designated *kaiB* is 312 nucleotides in length, encoding a predicted 103 amino acid protein with a calculated molecular mass of 11.8 kD. As shown in Figure 3.5c, KaiB is well conserved between *Nostoc* and *S. elongatus*, both in length (*S. elongatus* KaiB has 102 amino acids) and in protein sequence (81 % identity and 88 % similarity). No protein motifs were identified in the putative *Nostoc* KaiB sequence (or *S. elongatus* KaiB) when compared to Swiss-Prot databases. BLASTP analysis of *Nostoc* KaiB against GenBank databases revealed that its closest matches are the 103 amino acid KaiB protein identified from genome sequence analysis of *Nostoc* sp. PCC 7120 (94 % identity and 100 % similarity; GenBank accession NP_486925, see also (38)) and a 104 amino acid hypothetic protein identified from genome sequence analysis of *N. punctiforme* ATCC 29133 (96 % identity and 98 % sequence similarity; GenBank accession ZP_00101015, see also (50)). Further discussion of identified *kaiB* sequences will be continued in Section 3.5.

3.4.3. *kaiC*

The open reading frame designated *kaiC* is 1563 nucleotides in length, encoding a predicted 520 amino acid protein with a calculated molecular mass of 58.2 kD. As shown in Figure 3.5d, KaiC is well conserved between *Nostoc* and *S. elongatus*, both in length (*S.*

elongatus KaiC has 519 amino acids) and in protein sequence (80 % identity and 91 % similarity). Putative protein motifs originally identified in *S. elongatus* KaiC (32) are conserved in the putative *Nostoc* KaiC sequence. This includes two P-loop motifs (GXXXXGKT/S, also known as Walker's A motif) and two imperfect Walker's B motifs (ZZZZD) as well as two possible catalytic glutamate residues that are conserved in ATP/GTP-binding proteins and are involved in interactions with the target nucleotide (75). The overall composition of KaiC as tandem duplicate domains (34) is also conserved between *Nostoc* and *S. elongatus*. The two domains of *Nostoc* KaiC are shown as KaiCI and KaiCII in Figure 3.6 and share 22 % identity and 43 % amino acid similarity. BLASTP analysis of *Nostoc* KaiC against GenBank databases identified its closest matches as a 520 amino acid hypothetical protein of *N. punctiforme* ATCC 29133 (96 % identity and 98 % similarity; GenBank accession ZP_00107014, see also (50)) and the 519 amino acid KaiC of *Nostoc* sp. PCC 7120 (91 % identity and 95% similarity; GenBank accession NP_486926, see also (38)). Further discussion of *kaiC* genes and protein function will be continued in Section 3.5.

3.4.4. Other genes, transcriptional and translational signals

Aside from *kaiA*, *kaiB* and *kaiC*, there do not appear to be any other intact genes located on either strand of the 3.4 kb genomic sequence shown in Figure 3.4. There are three small open reading frames encoding 36-40 amino acid polypeptides (one on the sense strand, two on the anti-sense strand) but these did not match any known or putative protein sequences in the GenBank databases. Based on BLASTX analysis, a large portion of the 482 bp coding strand sequence downstream of *kaiC* appears to be highly similar to a hypothetical protein (with no known function) identified by genome sequence analysis in *N. punctiforme* ATCC 29133. This region of deduced protein sequence similarity matches to a partial open reading frame initiated with a GTG start codon (encoding valine instead of methionine, Figure 3.4). As *kaiABC* in *S. elongatus* is flanked upstream by genes encoding ribosomal proteins and downstream by a putative dehydrogenase (GenBank accession AB010691, also (32)), it seems that conservation of genomic sequence arrangement between *Nostoc* and *S. elongatus* does not extend beyond the *kai* gene cluster itself.

Scrutiny of the 3.4 kb *Nostoc* genomic sequence (Figure 3.4) for transcriptional and translational motifs did not result in strong identification of putative signals. Cyanobacterial transcriptional motifs are not well characterized and the collective dataset for this group of

```
KaiCI     1 MIENEQVEPKQTPIIRGVEKIRTMIEGFDDISHGGLPIGRTTLISGTSGTGKTLFSLQFL  60
KaiCII  261 ----------------V-RVSSGVKTLDEMCGGGFFKDSIILATGATGTGKTLLVSKFI  302
               ‘               *  ..  ..  *..  **        *  .*  .******   .*.

KaiCI    61 YNGITYFDEAGVFVTFEESPSDIIKNAHVFGWNLPRLIEEGKLFILDASPDPEGQDIVGN  120
KaiCII  303 QDGCLNGEQAILF-AYEESRAQLSRNASSWGIDFEELEDQGLLKIICTYPESTG------  355
               .*      ..*  .*   .*** . . .** .* .    * ..* * *.   *.  *

KaiCI   121 FDLSALIERLQYAIRKYKAK RVSID SITAVFQQYEAMGVVRREIFRLVARLKLLNVTTVI  180
KaiCII  356 --LEDHLQIIKSEIAIFKPA RIAID SLSALARGV-SNNAFRQFVIGVTGYAKQEEITGFF  412
               *     .. ..  *   .* *..***..*.  .    *...   *    .*

KaiCI   181 TTERGEEYGPVASFGVE-EFVSDNVVIVRNVLEGERRRRTIEILKLRGTTHMKGEYPFTI  239
KaiCII  413 TNTTDQFLGAHSITDSHISTITDTILMLQYVEIRGEMSRAINVFKMRGSWHDKGIREYNI  472
               *    . *  .       ..* ..... *      * *  . *.**. * **   . *

KaiCI   240 TNEGVNIFPLGAMRLTQRSSN--------------------------- 260
KaiCII  473 TADGPDIKDSFRNYERIISGAPTRVSIDEKAELSRIVRRFEDKQSSEP 520
               *  .*  .*              *.
```

Figure 3.6. Alignment of the tandem duplicate domains of *Nostoc* KaiC. The predicted amino acid sequence of the first half of KaiC (residues 1-260) is aligned with that of the second half (residues 261-520). Identical and similar amino acids are indicated by asterisks and periods, respectively. Hyphens were introduced to optimize alignment. P-loop motifs are shown in reversed font. Imperfect Walker's B motifs are boxed, and possible catalytic glutamate residues are highlighted in grey.

ancient and diverse genera is still fairly small. As reviewed by Curtis and Martin (13), a majority of putative cyanobacterial promoters contain a conserved TANNNT motif that conforms in sequence and position to the –10 hexamer consensus (TATAAT) of *E. coli* σ^{70} promoters. A minority of promoters contain motifs that resemble the –35 consensus hexamer (TTGACA) of *E. coli* σ^{70} promoters; when present, the resemblance is relatively weak (less than 25 % conservation at each of the first three positions). Given that the genus *Nostoc* is known to have a low GC content, it is not surprising that a large number of TANNNT motifs can be located on the 3.4 kb *Nostoc kaiABC* sequence (motifs are not indicated in Figure 3.4). Very little is known about transcription termination in cyanobacteria; in the absence of genera-specific knowledge, the Rho-independent and Rho-dependent systems elucidated in the classic eubacterial paradigm, *E. coli,* are usually applied. No canonical Rho-independent sequences were identified on the coding strand of the *Nostoc kaiABC* genomic region. Similarly, little is known about cyanobacterial translational systems and the use of the purine-rich hexameric Shine-Dalgarno motif (AGGAGG in *E. coli*) which is generally positioned within 20 bp of a start codon. Recent analysis of the published *Synechocystis* sp. PCC 6803 genome indicated that only 26 % of genes in this cyanobacterium were associated with Shine-Dalgarno sequences (48). Potential Shine-Dalgarno motifs for *Nostoc kaiA, kaiB* and *kaiC* are marked in Figure 3.4, but those identified for *Nostoc kaiB* and *kaiC* are not ideal in terms of either sequence or location. In the *S. elongatus* model, the *kaiABC* gene cluster is expressed as two transcriptional units, with *kaiA* on a separate mRNA transcript from *kaiBC*. These data had been obtained with the use of a promoter trap assay and Northern blot analysis (32) and do not rely on sequence information. Because sequence-based evidence for gene expression signals within the *Nostoc kaiABC* region is ambiguous, further discussion of this topic will be deferred to Section 5.4, wherein the results from RT-PCR analyses of RNA transcriptions are described.

3.5. Discussion

3.5.1. Occurrence of *kai* genes in prokaryotes

When this work was initiated in June of 1999, very few *kaiA, kaiB* and *kaiC* gene sequences were available: those from the *S. elongatus* PCC 7942 circadian model, one *kaiA,* three *kaiB* and three potential *kaiC* homologues in the *Synechocystis* sp. PCC 6803 genome sequence, and a handful of *kaiC*-like conserved hypothetical proteins discovered in three archaebacterial genome sequences. Since then, the advent of the genome sequencing era has

produced a massive inundation of genetic information. As described by Dvornyk (19), putative *kai* homologues have been identified in almost all the major taxa of the Archaea as well as four major taxa of the Eubacteria. In the Cyanobacteria, *kaiC* genes appear to be universal (46), but they are irregularly distributed within the Proteobacteria, Thermotogae and Chloroflexi as well as in the Archaea. While putative *kaiB* and *kaiC* homologues occur in other taxa, *kaiA* genes appear to be unique to the Cyanobacteria. Dvornyk uses the term "homologue" loosely to indicate sequence similarity and/or an evolutionary relationship between genes rather than protein function. Currently, the only organism in which *kai* genes have been linked to function is *S. elongatus* PCC 7942. It is as yet unclear what function(s) may be served by multiple copies of *kai* genes in an organism. Those cyanobacteria which are known to carry multiple copies of *kai*-like genes have one copy of each gene which is clearly more similar to *S. elongatus kaiA*, *kaiB* or *kaiC* than the other copies. It seems likely that the most highly conserved copies serve the original function of a circadian pacemaker, whereas the more distantly related copies may be functionally associated or have completely distinct roles.

It is possible that *Nostoc* PCC 9709 contains more *kai*-like genes in addition to the *kaiABC* gene cluster described in this report. Although the genome of the closely-related *N. punctiforme* ATCC 29133 does not contain extra *kai* genes in addition to its single *kaiABC* cluster (50), the genome of the somewhat more distantly-related *Nostoc* PCC 7120 contains two *kaiB* genes (38). If additional *kai* sequences exist in *Nostoc* PCC 9709, it is unlikely that these related genes will be discovered using the methods that identified the *kaiABC* gene cluster given the various experimental controls and checks that have already been performed.

3.5.2. *Nostoc* KaiA, KaiB and KaiC – sequence and function

Of the three genes, *kaiA* appears the least conserved amongst the cyanobacteria, both in length and in polypeptide sequence similarity. The findings described in Section 3.4.1 are congruent with more conclusions drawn from a wider spectrum of data (19): 1) gene lengths in unicellular cyanobacteria are 1.5 to 2 times as long as those in filamentous cyanobacteria, and 2) deduced polypeptide sequence homology is located at the C-terminal region. The C-terminal domain of *S. elongatus* KaiA has been demonstrated to interact directly with KaiC and stimulate KaiC autophosphorylation, whereas the N-terminal region appears to be a pseudo-receiver domain that is important for receiving input from environmental signaling (76). These structure-function relationships are consistent with conservation of the central circadian oscillation

mechanism in cyanobacteria (mediated by the conserved KaiA C-terminal region) and variation in the signal input pathways (channeled through the non-conserved N-terminal region) that reflect the diversity of cyanobacterial lifestyles.

Very little is known about the function of KaiB, except that (in *S. elongatus*) it is a component of heteromultimeric complexes formed with KaiA, KaiC and SasA, a clock-associated protein that shares homology with KaiB at its C-terminal end (34, 35, 37). Dvornyk grouped *kaiB*-like genes into three evolutionary clades; one of which (B1) contains only cyanobacterial sequences and includes *S. elongatus kaiB* (19). Interestingly, for those cyanobacteria that carry multiple copies of *kaiB*, only one copy grouped into clade B1. This clade displays a very low number of nonsynonymous nucleotide substitutions (approximately 9 %) compared to the other two clades (48 % and 69 %). Because KaiB acts in close contact with other clock proteins, it seems reasonable to: a) associate a necessarily high degree of sequence conservation with this function; and b) conclude that the B1 clade contains those genes most likely to actually be directly involved in the circadian pacemaker mechanism. Given the findings in Section 3.4.2, this lends credence to the suggestion that *Nostoc* KaiB is a functional homologue of S. *elongatus* KaiB.

Of the three clock components, the KaiC protein of *S. elongatus* has been the most intensively studied. It contains one P-loop motif, one imperfect Walker's B motif and one putative catalytic glutamate residue on each of its two tandemly duplicated domains. These protein motifs, suggestive of ATP/GTP-binding activity, are conserved in *Nostoc* KaiC (Section 3.4.3). *S. elongatus* KaiC has been demonstrated to have ATPase and autophosphorylation activities (55) and is now classified with in the RecA superfamily of ATPases (44). The duplicate domains of KaiC both bind KaiA (67), but may have distinct functions as *in vitro* assays have determined ATP-binding activity in only the N-terminal domain. Direct associations between KaiC and other clock proteins are required for generation of circadian rhythms (34). Although no DNA-binding motifs have been recognized in this protein, recent reports indicate that KaiC forms an ATP-induced hexameric structure that can bind forked DNA substrates (29, 53). Given the high degree of sequence conservation between *S. elongatus* and *Nostoc* KaiC polypeptides, particularly regarding the functionally-characterized protein motifs and the overall structure of the gene, there is no obvious evidence to deny the suggestion that *Nostoc* KaiC is a homologue of *S. elongatus* KaiC.

In this chapter, I have shown that *Nostoc* PCC 9709 contains a gene cluster that is very like *kaiABC* of *S. elongatus* PCC 7942. The putative *Nostoc kaiA*, *kaiB* and *kaiC* homologues are arranged in the same relative order on the chromosome and share sequence similarities that can be related to conservation of protein functions. Following chapters show that these open reading frames are genes (i.e., are transcribed) and describe investigation of the transcriptional activities of *kaiABC* in axenic and symbiotically-associated *Nostoc* with the aim of determining whether RNA patterns support the hypothesis that *Nostoc kaiABC* is the genetic basis of a circadian mechanism homologous to the *S. elongatus* clock oscillation model.

CHAPTER FOUR

Characterization of *Peltigera membranacea* lichen specimens for comparative studies

4.1. Introduction

4.1.1. Origin of Nostoc PCC 9709

Nostoc sp. PCC 9709 originated from a study on the genetic identity of photobionts associated with *Peltigera membranacea* colormorphs (57). This strain was isolated from lichen specimen TDI#AR95 collected at Eagle Ridge in the district of North Vancouver, British Columbia (51). Molecular analysis indicated that *Nostoc* PCC 9709 was the primary cyanobiont of lichen TDI#AR95. Based on 16S rDNA sequence analysis, *Nostoc* PCC 9709 was determined to be identical to one other strain isolated from lichen TDI#AR95 and two strains isolated from another *P. membranacea* specimen TDI#94, but not the primary cyanobiont of TDI#94 (1.86 % difference over 1396 nucleotides). Of the four cyanobacterial isolates, *Nostoc* PCC 9709 was deposited into the Pasteur Cyanobacterial Culture Collection because it was the only one purified of bacterial and fungal contaminants. Phylogenetic analysis grouped Nostoc PCC 9709 and the primary photobiont of specimen TDI#AR94 closely with *N. punctiforme* ATCC 29133 (synonymous to PCC 73102), which is a species known for symbiotic associations with plants and fungi (50).

4.1.2. Research objectives

The goal of the project described in this chapter was to extend the comparative study of the cyanobacterial circadian clock genes to *Nostoc* in the state of symbiotic association. This goal was divided into the following parts: 1) Collect fresh specimens of *P. membranacea*, 2) establish the genetic relatedness of the lichen specimen(s) to *Nostoc* PCC 9709 and to each other, and 3) evaluate the genetic similarity in *Nostoc* and lichen clock genes before progressing to transcriptional analyses of isolated and symbiotically-associated *Nostoc* in Chapter Five.

4.2. *Peltigera membranacea* – Collection & morphological identification

Fresh specimens of *P. membranacea* (designated AL#20, AL#21, AL#22, AL#23 and AL#24) were collected from a single (1 m by 1 m) site on the coastal side of Cypress Mountain (49°22'N, 123°16'W), located northwest of Vancouver, British Columbia (Appendix C). This is essentially the same geographic locale from which previous specimens TDI#AR94 and TDI#AR95 were collected. *P. membranacea* is easily recognized in the field due to its

characteristic "membranous" texture, which is in contrast to the thicker and stiffer thalli of similar lichens. Specimens were directly transported to the laboratory where they were briefly washed in tap water to remove field debris. After the species identification was confirmed with the use of a dissecting microscope to verify additional taxonomic features (e.g., raised veins covered in minute erect hairs, (25)), *P. membranacea* specimens were photographed on black and white Polaroid film (Figure 4.1) and genomic DNA was extracted from a small thallus portion as described in Section 2.2.1.

4.3. Analysis of 16S rRNA gene sequences as a measure of relatedness

For a measure of relatedness, I used the 16S rRNA gene, which is a common marker for evolutionary studies of cyanobacteria (31, 71). PCR amplification of lichen specimens AL#21 and AL#24 for DNA sequence analysis is described in Section 2.6.3. Information for a 1408 bp region was collected from *P. membranacea* specimens. 16S rDNA sequences from AL#21 and AL#24 were found to be identical. Alignment to the PCC 9709 16S rDNA sequence revealed 29 differences (2.06 %) distributed across the region (data not shown, an alignment of identical sequences was provided in (51)). Interestingly, the AL#21/AL#24 16S rDNA sequence is the same as that reported for the primary photobiont of lichen specimen TDI#AR94 (51).

4.4. Comparison of circadian clock sequences

4.4.1. *kaiABC* genomic region

PCR amplification of *P. membranacea* genomic DNA with *kaiA* and *kaiC*-specific primers designed for Nostoc PCC 9709 (Section 2.6.4) resulted in a single amplicon that was 0.2 kb larger than the 1.7 kb product predicted from the *Nostoc* PCC 9709 *kaiABC* genomic sequence. Complete DNA sequencing of the 1.9 kb product gave identical sequences from lichens AL#21 and AL#24, which is a result congruent with 16S rDNA analysis of those two specimens. Eighty percent sequence identity is observed across a 1871 nucleotide alignment of the lichen and Nostoc *kaiABC* sequences (Figure 4.2). Most of the mismatch can be attributed to an insertion of 175 bp sequence into the intergenic region between *kaiB* and *kaiC* of the lichen sequence (Figure 4.3a). The extended *kaiBC* intergenic region contains a small open reading frame that encodes a 60 amino acid polypeptide (Figure 4.3b). BLASTX analysis of the intergenic region returns a few weak matches to short sequences found in ssDNA-binding and
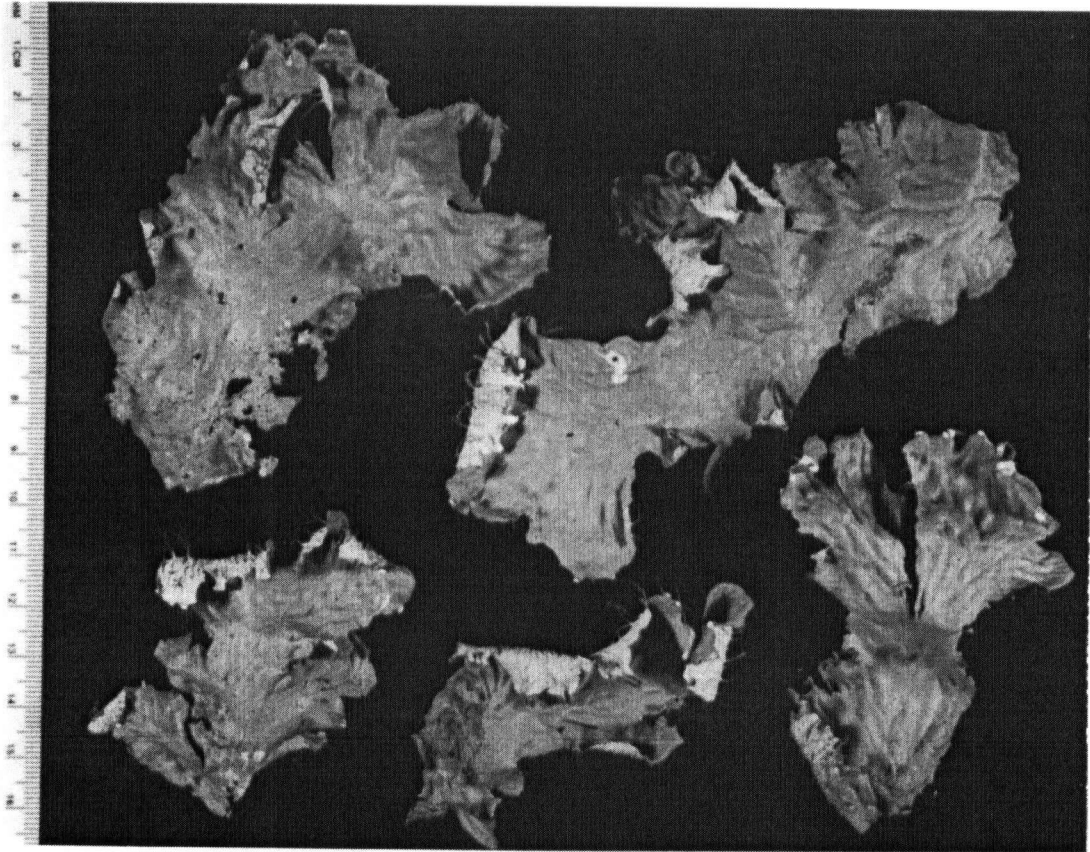
Figure 4.1. *Peltigera membranacea* **specimens.** Lichen thalli were collected from a single location on Cypress Mountain in the district of North Vancouver (British Columbia, Canada). Specimens are identified as (top row, left to right) AL#20, AL#21, (bottom row, left to right) AL#22, AL#23 and AL#24.

Figure 4.2 (three pages)


```
PCC 9709    AAAAGTTTAAATAATCTAGTTGACTTAGCCAATCAAAAAGCTTTGATCCCGTGGGTGTGG 60
AL#21/AL#24 AAATGTTTAGAAAATCTAGTTGGCTTGGCCAATCAAAGGATTTTTGGTGTGTGGGCGTGG 60
            *** ***** * ********* *** *********    ***       ***** ****


PCC 9709    AAATTGATTGACGCAACTATTTCCCAATTTTCATCCCTACCTGTTGCTGCAACTACCTCT 120
AL#21/AL#24 AAATTTATTGACGCAATTATTGCCCAATTTTGCTACTTACCTGTTGCTGCAACTACCACT 120
            ***** ********* **** ********  * * ****************** **


PCC 9709    GAAAAAATCAACTATTTACCAAATTGGCTGCAACAAAGTCCAAGTAAGGCATACACTGGC 180
AL#21/AL#24 GGGGAAATCAAATATTTACCAAATTGGCCGCAAAAAAATCTAAGTAGCGCCTACACTGGC 180
            *      ****** **************** **** *** ** *****  ** ********


PCC 9709    AAATACGTCTATGTGTTTGCCAGTCAGATGCAAAAAAGCCAACAGCATTTACAGGAGATG 240
AL#21/AL#24 CAATACGTCTATGTATTTGCCAGTCAGATGCAAAAAAGCCAACAGCATTTCCAGGAGATG 240
            ************* ************************************* ********


PCC 9709    ACTCCAGCCGAAAGGCAAGGATTATTAAGACAGCTTAAATTAGATTATAGCCTGATTCTT 300
AL#21/AL#24 ACACCAGCAGAAAGGCAAGGATTATTAAGACAGCTTAAATCAGATTACACCCTAATCCTT 300
            ** ***** ****************************** ****** * *** ** ***


PCC 9709    ATAGATTATTTTACCACAGATAAAACACTCAAAGATAAAATTGATAAATTTATCAATACT 360
AL#21/AL#24 ATAAACTATTTTACTACAGACAAAACACTAAAAGAAAAAATTGATAAATTTATCAATACT 360
            *** * ******** ***** ******** **** ** *********************


PCC 9709    ATATTTTATGCTAATATTCCTGTGCCCCAAATAATCGAAATTCACATGGAGGTAATTGAA 420
AL#21/AL#24 ATATTTTATGCTAATATTCCTGTGCCCCAAATCATCGAAATTCACATGGAAATAATTGAA 420
            ******************************** ***************** *******


PCC 9709    GATTTTTCTAACCAGCTAAAATTAGAAGGAAGGAGCAATGAAACGTTACTGGATTACCGT 480
AL#21/AL#24 GAATTTTCTATCCAGCTAAGAATAGAAGGAAGGAGCAATGAAACCTTACTTGATTATCGC 480
            ** ******* ******* * ********************* **** ***** **


PCC 9709    CTAACGTTAATAGATATCCTGGCTCACCTGTGCGAACTCTATCGGAGTTCGATTTCTAAA 540
AL#21/AL#24 CTGACGTTGATAGATATCTTGGCTCACCTGTGCGAAGTCTATAGAAGTTCGAGTCCTAGA 540
            ** ***** ********* ***************** ***** * ******* * *** *


PCC 9709    TAA-TTAAATTTATCGATGTATATTCAACTGGGAACAAGCTACTGCTGTTAGTGACTACC 599
AL#21/AL#24 TAACTTAAATTTATCGATGT----TCAATTGGGAATAAACTAACGTTGTTAGTTATTCTT 596
            *** ***** **********    **** ****** ** * ** *  * ******* * *


PCC 9709    AATGCGCGAA-AGGTCTTGTATTTCTATGATCGTGACTAGCCATTAAATATATGAATAAA 658
AL#21/AL#24 AACGTGCGAACACATCCTGTATTTCTATCTGTATAACTAGCCTTTAAATATATGATTAAA 656
            ** * ***** *  ** **********   * ****** ********** *****  ****


PCC 9709    GCCAGAAAAACCTACGTTCTCAAGCTTTACGTAGCAGGGAACACACCTAATTCAGTCCGG 718
AL#21/AL#24 GCCAAAAAAACCTACGTTCTCAAGCTTTACGTAGCAGGCAACACCCCTAATTCCGTTCGG 716
            ****  ******************************* ***** **  ***** ** ***


PCC 9709    GCATTAAAAACACTCAAGGATATTTTAGAACAGGAGTTTGAAGGTGTTTATGCTTTAAAA 778
AL#21/AL#24 GCATTAAAAACACTCAAAGATATTTTAGAACAGGAGTTTGAAGGTGTTTATGCTTTAAAA 776
            ***************** ******************************************
```

Figure 4.2 (three pages)

```
PCC 9709      GTGATCGATGTCCTGAAAAGCCCGCAACTGGCAGAAGAAGATAAAATATTGGCGACGCCA 838
AL#21/AL#24   GTGATCGATGTACTAAAAAGCCCGCAACTGGCGGAAGAAGATAAAATATTGGCAACGCCA 836
              * * * * * * * * * *   * *   * * * * * * * * * * * * * * *   * * * * * * * * * * * * * * * * * * *   * * * * * *


PCC 9709      ACATTATCTAAAATTTTGCCTCCACCCGTTCGCAAAATTATCGGGGATCTTTCAGATAGA 898
AL#21/AL#24   ACATTATCTAAAATTTTACCTCCACCTGTTCGCAAAATTATCGGGGATCTTTCAGATCGA 896
              * * * * * * * * * * * * * * * *   * * * * * * * *   * * * * * * * * * * * * * * * * * * * * * * * * * * * *   * *


PCC 9709      GAAAGAGTATTGATTGGATTAGATTTGCTCTATGAAGAACTGAGTGAAGAAGATTTTGAA 958
AL#21/AL#24   GAAAGAGTATTAATTGGATTAGATTTGCTCTATGAAGAACTGAGTGAAGGAGATTTTGAA 956
              * * * * * * * * * *   * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *   * * * * * * * * * *


PCC 9709      ---TAATCCGAT--------------------------------------TTTAATCAT--- 976
AL#21/AL#24   GAGTAAAATAATTAATGAATGTGGCAAGAGAGAATTTATAAAGGAAATTTTAACTGTGTA 1016
                *   *     * *                                          * * * * *     *


PCC 9709      ------------------------------------------------------------ 976
AL#21/AL#24   GGGAAGCCAGTGTATTGCTCTAATCAAGATAGTTCGATCGCTATTTATCTGGATGGGTTA 1076


PCC 9709      ------------------------------------------------------------ 976
AL#21/AL#24   CGTCAAAAATATTTATTTTGGCGATCTTTTAGTTCAGATTGTTACCTAACACACCCTACT 1136


PCC 9709      ---------------------------AAAAAATTGAGTACAAACTT----TAGTAAT 1003
AL#21/AL#24   TGCTTAAGATTTGGTTTATAAATATCCTAAAAAAATTGAGTCCAAACTTACTTTAGTAAT 1196
                                         * * * * * * * * * * * *   * * * * * * *       * * * * * *


PCC 9709      AAAA-AAACAGGTTTTAATACCAATTGATTATCAAGCAATGATTGAAAACGAGCAAGTAG 1062
AL#21/AL#24   AAAACAAACAGTTTTTAATACCCTTTTATTATCAAGCAATGAGTCAAAACGAGCAAATAG 1256
              * * * *   * * * * * *   * * * * * * * * *     * *   * * * * * * * * * * * * * *   *   * * * * * * * * * *   * * *


PCC 9709      AACCAAAGCAAACACCGATAATTAGGGGTGTAGAAAAAATTCGTACGATGATCGAAGGGT 1122
AL#21/AL#24   AACCAAATAAAACACCTAAAAATGGAGGTGTAGAAAAAATTCGCACGATGATCGAAGGAT 1316
              * * * * * * *     * * * * * * *   *   * *   *   *   * * * * * * * * * * * * * * * * *   * * * * * * * * * * * * * *   *


PCC 9709      TTGACGATATTAGTCATGGTGGTTTACCAATTGGTAGAACTACCTTGATCAGTGGTACCT 1182
AL#21/AL#24   TTGAYGATATTAGTCATGGTGGTTTACCAATTGGTAGAACTACTTTGGTTAGTGGTACTT 1376
              * * * *   * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *   * * *   *   * * * * * * * *   *


PCC 9709      CCGGCACAGGCAAAACTTTATTCTCTCTTCAGTTTCTCTATAACGGTATCACCTACTTTG 1242
AL#21/AL#24   CCGGTACAGGTAAAACTTTATTATCTCTTCAGTTTCTCTATAACGGTATCACCTACTTTG 1436
              * * * *   * * * * *   * * * * * * * * * *   * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *


PCC 9709      ATGAAGCAGGAGTATTTGTTACCTTTGAAGAATCACCCAGTGATATTATTAAAAATGCCC 1302
AL#21/AL#24   ATGAAGCCGGAGTATTTGTCACCTTTGAAGAATCACCCAGTGATATCATTAAAAATGCCC 1496
              * * * * * * *   * * * * * * * * * *   * * * * * * * * * * * * * * * * * * * * * * * * * * *   * * * * * * * * * * *


PCC 9709      ATGTTTTTGGTTGGAACTTGCCACGCCTAATTGAAGAAGGCAAGTTATTTATTCTTGATG 1362
AL#21/AL#24   ATGTTTTTGGTTGGAACTTGCCACGTCTAATTGAAGAAGGCAAGTTGTTTATTCTTGATG 1556
              * * * * * * * * * * * * * * * * * * * * * * * *   * * * * * * * * * * * * * * * * * * * *   * * * * * * * * * * * *
```

Figure 4.2 (three pages)

```
PCC 9709     CATCTCCCGATCCAGAAGGTCAAGATATCGTTGGTAATTTTGACCTTTCTGCACTCATTG 1422
AL#21/AL#24  CATCTCCCGATCCAGAAGGTCAAGATATCGTTGGGAATTTTGACCTTTCAGCACTTATTG 1616
             ********************************** ************** ***** ****


PCC 9709     AACGCTTGCAATATGCCATCCGTAAATACAAAGCTAAACGAGTTTCAATCGACTCAATAA 1482
AL#21/AL#24  AACGCCTGCAATATGCCATTCGCAAATACAAAGCTAAACGAGTTTCAATCGACTCAATGA 1676
             ***** ************* ** *********************************** *


PCC 9709     CAGCAGTATTTCAGCAGTATGAAGCGATGGGAGTAGTGCGACGTGAGATTTTTCGCCTGG 1542
AL#21/AL#24  CAGCCGTATTTCAGCAGTACGAAGCGATGGGAGTAGTGAGACGCGAGATTTTTCGCTTGG 1736
             **** ************** ****************** **** *********** ***


PCC 9709     TAGCACGTCTGAAATTATTGAATGTCACCACTGTAATTACCACTGAACGTGGTGAAGAAT 1602
AL#21/AL#24  TAGCACGGCTCAAAATACTGAATGTCACCACTGTAATTACCACTGAACGGGGTGAAGAAT 1796
             ******* ** *** ** ******************************* *********


PCC 9709     ATGGGCCTGTTGCCTCTTTCGGAGTAGAAGAATTTGTTTCTGATAATGTAGTAATTGTTC 1662
AL#21/AL#24  ATGGCCCCGTTGCTTCTTTCGGTGTGGAAGAATTTGTTTCTGATAATGTAGTAATTGTTC 1856
             **** ** ***** ******* ** **************************************


PCC 9709     GCAACGTTTTA ·                                                1673
AL#21/AL#24  GCAACGTTTTA                                                  1867
             ***********
```

Figure 4.2. Alignment of *kaiABC* sequences from lichen specimens and Nostoc PCC 9709.
This alignment corresponds to nucleotides 316 through 1988 of the *Nostoc kaiABC* genomic
sequence as shown in Figure 3.4. Asterisks indicate positions where the two sequences are
identical. Putative *kaiB* and *kaiC* start codons are highlighted in grey. Putative *kaiA* and *kaiB*
stop codons are shown in boldface. The lichen *kaiABC* sequence was deposited in the GenBank
database (accession number AY373438).

A
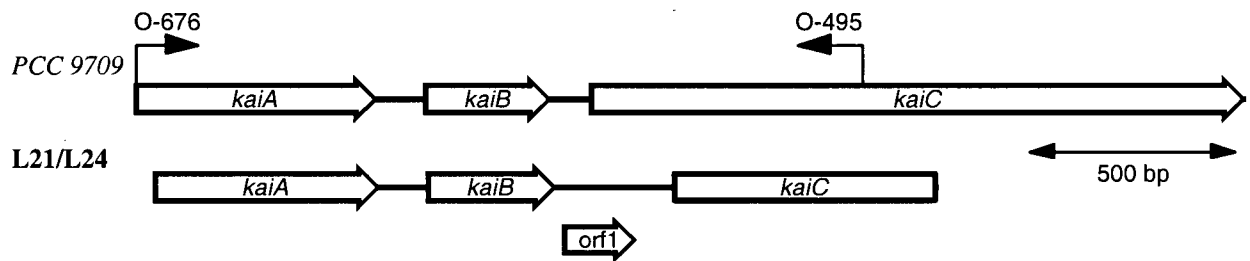
B

```
                      ************************ *********    ***
PCC 9709      GATTTGCTCTATGAAGAACTGAGTGAAGAAGATTTTGAA---TAA-------------- 1276
AL#21/AL#24   GATTTGCTCTATGAAGAACTGAGTGAAGGAGATTTTGAAGAGTAAAATAATTAATGAATG

PCC 9709      ----------------------------------------------------------- 1276
AL#21/AL#24   TGGCAAGAGAGAATTTATAAAGGAAATTTTAACTGTGTAGGGAAGCCAGTGTATTGCTCT

PCC 9709      ----------------------------------------------------------- 1276
AL#21/AL#24   AATCAAGATAGTTCGATCGCTATTTATCTGGATGGGTTACGTCAAAAATATTTATTTTGG

                                                           *   *  ** ***
PCC 9709      ---------------------------------------------------TCCGATTTTAA 1287
AL#21/AL#24   CGATCTTTTAGTTCAGATTGTTACCTAACACACCCTACTTGCTTAAGATTTGGTTTATAA

              **  * *********** ******      ********** ***** ********
PCC 9709      ---TCAT-AAAAAATTGAGTACAAACTT----TAGTAATAAAA-AAACAGGTTTTAATAC 1338
AL#21/AL#24   ATATCCTAAAAAAATTGAGTCCAAACTTACTTTAGTAATAAAACAAACAGTTTTAATAC

              *   ** ************** * ********** ********* ******* * **
PCC 9709      CAATTGATTATCAAGCAATGATTGAAAACGAGCAAGTAGAACCAAAGCAAACACCGATAA 1398
AL#21/AL#24   CCTTTTATTATCAAGCAATGAGTCAAAACGAGCAAATAGAACCAAATAAAACACCTAAAA
```

Figure 4.3. Alignment of intergenic region between *kaiB* and *kaiC* of lichen specimens and *Nostoc* PCC 9709. A) Diagram showing the alignment of the gene clusters. Arrows indicate the location and orientation of primers used to amplify the partial *kaiABC* sequence from *P. membranacea* specimens AL#21 and AL#24. B) Nucleotide alignment of the region between *kaiB* and *kaiC*. Numbers on the right side correspond to nucleotide positions of PCC 9709 *kaiABC* the sequence as shown in Figure 3.4. The TAA stop codon in boldface indicates the 3' end of *kaiB*, and the ATG start codon shaded in grey indicates the 5' end of *kaiC*. Underlined ATG and TAA indicate the ends of the short open reading frame in between *kaiB* and *kaiC* in *P. membranacea* AL#21 and AL#24.

A

```
Nostoc KaiA   MLLPILILPPDVNKSLNNLVDLANQKALIPWVWKLIDATISQFSSLPVAATTSEKINYLP  60
Lichen KaiA            // KCLENLVGLANQRIFGVWAWKFIDAIIAQFCYLPVAATTTGEIKYLP  47
                          *.* *** ****.    * ** *** *.**. ******.  * ***


Nostoc KaiA   NWLQQSPSKAYTGKYVYVFASQMQKSQQHLQEMTPAERQGLLRQLKLDYSLILIDYFTTD 120
Lichen KaiA   NWPQKNLSSAYTGQYVYVFASQMQKSQQHFQEMTPAERQGLLRQLKSDYTLILINYFTTD 107
              ** *.   * ****.************** *************** **.**** *****


Nostoc KaiA   KTLKDKIDKFINTIFYANIPVPQIIEIHMEVIEDFSNQLKLEGRSNETLLDYRLTLIDIL 180
Lichen KaiA   KTLKEKIDKFINTIFYANIPVPQIIEIHMEIIEEFSIQLRIEGRSNETLLDYRLTLIDIL 167
              ****.**********************.**.** **..*****************


Nostoc KaiA   AHLCELYRSSISK                                                193
Lichen KaiA   AHLCEVYRSSSPR                                                180
              *****.****  .
```


B

```
Nostoc KaiB   MNKARKTYVLKLYVAGNTPNSVRALKTLKDILEQEFEGVYALKVIDVLKSPQLAEEDKIL  60
Lichen KaiB   MIKAKKTYVLKLYVAGNTPNSVRALKTLKDILEQEFEGVYALKVIDVLKSPQLAEEDKIL  60
              * **.******************************************************


Nostoc KaiB   ATPTLSKILPPPVRKIIGDLSDRERVLIGLDLLYEELSEEDFE                 103
Lichen KaiB   ATPTLSKILPPPVRKIIGDLSDRERVLIGLDLLYEELSEGDFEE                104
              *************************************** ***
```


C

```
Nostoc KaiC   MIENEQVEPKQTPIIRGVEKIRTMIEGFDDISHGGLPIGRTTLISGTSGTGKTLFSLQFL  60
Lichen KaiC   MSQNEQIEPNKTPKNGGVEKIRTMIEGFDDISHGGLPIGRTTLVSGTSGTGKTLLSLQFL  60
              * .***.** .**    **************************.********** *****


Nostoc KaiC   YNGITYFDEAGVFVTFEESPSDIIKNAHVFGWNLPRLIEEGKLFILDASPDPEGQDIVGN 120
Lichen KaiC   YNGITYFDEAGVFVTFEESPSDIIKNAHVFGWNLPRLIEEGKLFILDASPDPEGQDIVGN 120
              ************************************************************


Nostoc KaiC   FDLSALIERLQYAIRKYKAKRVSIDSITAVFQQYEAMGVVRREIFRLVARLKLLNVTTVI 180
Lichen KaiC   FDLSALIERLQYAIRKYKAKRVSIDSMTAVFQQYEAMGVVRREIFRLVARLKILNVTTVI 180
              ********************.*****.*************************.*******


Nostoc KaiC   TTERGEEYGPVASFGVEEFVSDNVVIVRNVLEGERRRTIEILKLRGTTHMKGEYPF // 237
Lichen KaiC   TTERGEEYGPVASFGVEEFVSDNVVIVRNVL //                          211
              ******************************
```


**Figure 4.4. Alignment of deduced KaiA, KaiB and KaiC protein sequences from lichen
specimens and *Nostoc* PCC 9709.** A) Alignment of deduced KaiA sequences. B) Alignment of
deduced KaiB sequences. C) Alignment of deduced KaiC sequences. // indicates a partial
protein sequence. P-loop motifs are shown in reversed font. Imperfect Walker's B motifs are
boxed, and possible catalytic glutamate residues are highlighted in grey.

capsid proteins of *Saccharomyces cerevisiae* virus L-A. It is unclear whether this open reading frame may be translated into protein.

4.4.2. *kaiA*, *kaiB* and *kaiC* genes

The 543 nucleotide sequence available for *P. membranacea kaiA* (the 5' end is missing) aligns reasonably well (87 % identity, no gaps) to the corresponding region of *Nostoc kaiA*. Similarly, there is 86 % identity and 94 % sequence similarity in the deduced 180 residue polypeptide region (Figure 4.4a). There is perhaps a slight decrease in similarity outside of the 100 amino acid C-terminal region (69 % identity and 78 % similarity 5' of the C-terminal region, versus 87 % identity and 95 % similarity in the last 100 residues), which is a domain associated with conservation of clock protein function. The *kaiB* gene found in *P. membranacea* is one codon longer than *Nostoc* PCC 9709 but otherwise displays high sequence conservation: 94 % identity in the nucleotide sequence, 96 % identity and 97 % similarity in the deduced polypeptide sequence (Figure 4.4b). The 633 nucleotide sequence available for *P. membranacea kaiC* (the 3' half is missing) is also well-matched (93 % invariant positions, no gaps) to the corresponding region of *Nostoc kaiC*. The KaiC alignment (Figure 4.4c) reveals 94 % identity and 97 % similarity across 211 amino acid residues. The P-loop motif, imperfect Walker's B motif and the putative catalytic glutamate residue with the 3' domain of KaiC are conserved.

4.5. Discussion

4.5.1. Photobionts of lichen specimens are closely related to *Nostoc* PCC 9709

The oligonucleotides used for PCR amplification of 16S rDNA sequences from *P. membranacea* were designed to be universal eubacterial primers. Therefore, the 16S sequences obtained from specimens AL#21 and AL#24 most likely represent the majority eubacterial member (primary photobiont) of those lichens. Although the 16S rDNA sequence found in lichen specimens AL#21/AL#24 is not identical to that of *Nostoc* PCC 9709, the close sequence similarity (98 % invariant positions across a 1408 bp region) is indicative of a high degree of relatedness. The primary photobionts within the collected lichen specimens appear similar enough to *Nostoc* PCC 9709 for inclusion in the *S. elongatus* comparative study.

4.5.2. Use of *P. membranacea* AL#21 and AL#24 as *Nostoc* PCC 9709 surrogates

In order to evaluate the suitability of *P. membranacea* AL#21 and AL#24 as surrogates for *kai* gene expression of symbiotically-associated *Nostoc* PCC 9709, genomic sequences for roughly two-thirds of the *kaiABC* locus on these lichens were examined. Although there was

some variability outside of the *kaiA*, *kaiB* and *kaiC* coding regions (particularly between *kaiB* and *kaiC*), deduced protein sequences were highly conserved (above 85 % for KaiA, above 95% for KaiB and KaiC) between *Nostoc* PCC 9709 and *P. membranacea* AL#21 and AL#24. As discussed in Chapter Three, KaiA appears to be the least conserved protein of the three circadian clock components, with the greatest sequence identity in the C-terminal region.

Given that the lichen symbiosis from which Nostoc PCC 9709 was originally isolated cannot be reconstituted under laboratory conditions, a proxy must be utilized for comparisons of this organism in isolated culture conditions versus symbiotic association. The two *P. membranacea* specimens evaluated as substitutes are identical to each other in the 16S rDNA and kaiABC sequences examined and they are highly similar to the corresponding genetic regions in Nostoc PCC 9709. Thus, they are appropriate surrogates for the investigation of symbiotically-associated *Nostoc* PCC 9709 *kaiABC* transcription patterns described in Chapter Five.

CHAPTER FIVE

Investigation into *Nostoc kai* gene expression

5.1. Introduction

5.1.1. The *S. elongatus kaiABC* model

In this prokaryotic system, as well as in eukaryotic model organisms such as *Drosophila* and *Neurospora*, the daily oscillations of the circadian clock are postulated to be driven by autoregulatory feedback loop(s). The rhythm of the cyanobacterial clock can be monitored by observing the relative change in accumulated *kai* gene transcripts during the course of a daily cycle. In *S. elongatus* PCC 7942, clock genes are transcribed on two mRNAs, one encoding KaiA and the other encoding KaiB as well as KaiC. Promoter activities and mRNA levels of these two *kai* transcripts are known to fluctuate rhythmically on a 24 cycle, peaking just before or at dusk and reaching minimum levels just before dawn (32). Although KaiB and KaiC proteins are also rhythmically abundant, KaiA protein abundance does not seem to oscillate (77).

The nature of the autoregulatory feedback loop in the *S. elongatus* clock is not well understood. Initially, it was thought that the KaiC protein acts negatively on the *kaiBC* promoter to generate an oscillation pattern (32). Such a mechanism is analogous to what is known for the *Neurospora* clock protein FRQ as well as the *Drosophila* clock protein PER (28), and it implies that rhythmicity is dependent on *kaiBC* promoter function. However, according to Xu et al. (78), circadian oscillation in *S. elongatus* does not specifically require the *kaiBC* promoter sequence. Instead, the authors assert that rhythmicity is dependent on the phosphorylation state and degradation rate of KaiC. Greater understanding of the *S. elongatus* circadian mechanism will clearly require further elucidation of the protein interactions of KaiA, KaiB and KaiC.

5.1.2. Research objectives

This chapter describes comparative studies undertaken to evaluate the similarity of *Nostoc kaiABC* to the *S. elongatus* circadian oscillator at the level of expression (versus genomic sequence comparisons described in Chapter Three). The most fundamental characteristic of circadian systems is the fact of a daily oscillation. I chose to detect clock oscillation signals at the RNA rather than the protein level, using the techniques of Northern blot hybridization and relative RT-PCR. The comparative analysis was organized into the following research objectives: 1) develop appropriate methods in support of the chosen investigative studies, 2) determine the transcriptional organization of the *kaiABC* gene cluster in *Nostoc, P.*

membranacea and *S. elongatus*, 3) determine the temporal *kai* gene expression profile of *Nostoc*, *P. membranacea* and *S. elongatus* by Northern blot analysis, and 4) determine the temporal *kai* gene expression profile of *Nostoc*, *P. membranacea* and *S. elongatus* by relative RT-PCR analysis.

5.2. Structure of circadian time course studies

Two fundamental characteristics of circadian systems are the ability to maintain rhythmicity in the absence of external cues and the ability to reset to the local environment (referred to as entrainment) according to key signals such as light and temperature. Studies to demonstrate the circadian profile of *Nostoc kai* gene expression were designed with these characteristics in mind. The term circadian time (CT) is used to indicate the subjective time of day in a 24 hour cycle. By convention, circadian time zero (CT 0) corresponds to subjective dawn (light appears) and subjective dusk is the time at which lighting is removed. Biological samples (cyanobacterial cultures and lichen field samples) were first pre-treated with at least 12 hours of darkness, as described in (32), to synchronize cells. Other than the alternation in lighting, environmental conditions were kept constant during the days leading up to as well as for the duration of the studies. Constant environmental conditions are required in order to detect the inherent rhythmicitiy of the circadian clock (as opposed to the response of the clock to entrainment signals). Upon assumption of constant light conditions, samples were removed at various time points to monitor changes in mRNA levels over at least a 24 hour period. Ideally, a time course experiment would occur over a period of time greater than one (preferably two) 24 hour cycles in order to demonstrate repetition of any daily pattern, but in this case the amount of available biological material was a limiting factor.

5.3. Preparation of RNA from cyanobacterial cultures and lichens

In order to perform the intended studies on *kai* gene expression, appropriate amounts of suitable quality RNA had to be obtained. For Northern blot studies, clear positive signals require relative large amounts (microgram quantities) of minimally degraded RNA. RNA extracted from unicellular *S. elongatus* using an established hot phenol method (52) was found to be satisfactory, but it was more difficult to obtain intact purified RNA from cultured *Nostoc* or *P. membranacea* thalli (methods described in Section 2.10). The critical factor was found to be the fragmentation of filamentous Nostoc mats and lichen thalli fragments and efficient cell lysis without activation of endogenous RNases. This was accomplished by grinding cell material in a mortar and pestle under liquid nitrogen and adding Trizol extraction reagent directly into the

cold mortar to resuspend samples. Other extraction methods (for example, beadbeater homogenization or Qiagen RNeasy RNA extraction) were observed to result in greater degrees of RNA degradation and/or gave less reproducible results (see hot phenol method, Section 2.10.2). A secondary factor which required modification of extraction procedures was the presence of polysaccharide in extracted samples. In axenic culture, *Nostoc* PCC 9709 filaments are encased in a mucilaginous sheath. Sheath material is composed primarily of polysaccharides and is known to co-purify with nucleic acids (14). Trizol-extracted RNA was precipitated using LiCl in order to separate it from the large amount of polysaccharide material found in *Nostoc* PCC 9709 cultures. Total RNA yields from axenic cultures of *Nostoc* were poor (0.2 µg per mg wet cell weight) compared to cultured *S. elongatus* (3.5 µg RNA per mg wet cell weight) or even fresh specimens of *P. membranacea* (0.4 µg RNA per mg damp thallus).

Gel electrophoretic analysis of cyanobacterial RNA preparations revealed one prominent band in addition to the expected number of bands corresponding to the major rRNA molecular species (Figure 5.1). This additional band, which migrated below the uppermost band of 23S rRNA, is routinely observed in RNA preparations from *S. elongatus* (S. Golden, personal communication) and *Anabaena* (R. Haselkorn, personal communication) and has been previously identified to be the result of specific postmaturational cleavage of the 23S rRNA species *in vivo* (15). Ribosomal RNA lability may be considered somewhat uncommon in prokaryotes but has been observed in numerous cyanobacterial species (26, 56). The 23S rRNA cleavage product was also observed in RNA preparations from the lichen *P. membranacea* (Figure 5.2). A rough visual comparison of the intensities of the eukaryotic and prokaryotic rRNA bands shown in Figure 5.2 indicates that approximately one-third of the RNA extracted from the lichen originated from the cyanobiont. This approximation was used to adjust the amounts of total RNA analyzed by Northern blot hybridization so that the amount of cyanobacterial RNA would be (roughly) equivalent.

5.4 *Nostoc kaiABC* is expressed as two transcriptional units

As shown in Figure 5.3, RT-PCR analysis confirmed published reports that *S. elongatus* clock genes are transcribed as two mRNA molecules, one encoding KaiA and the other encoding KaiB and KaiC. There was a faint signal for *kaiAB* (lane 2), but this is not considered
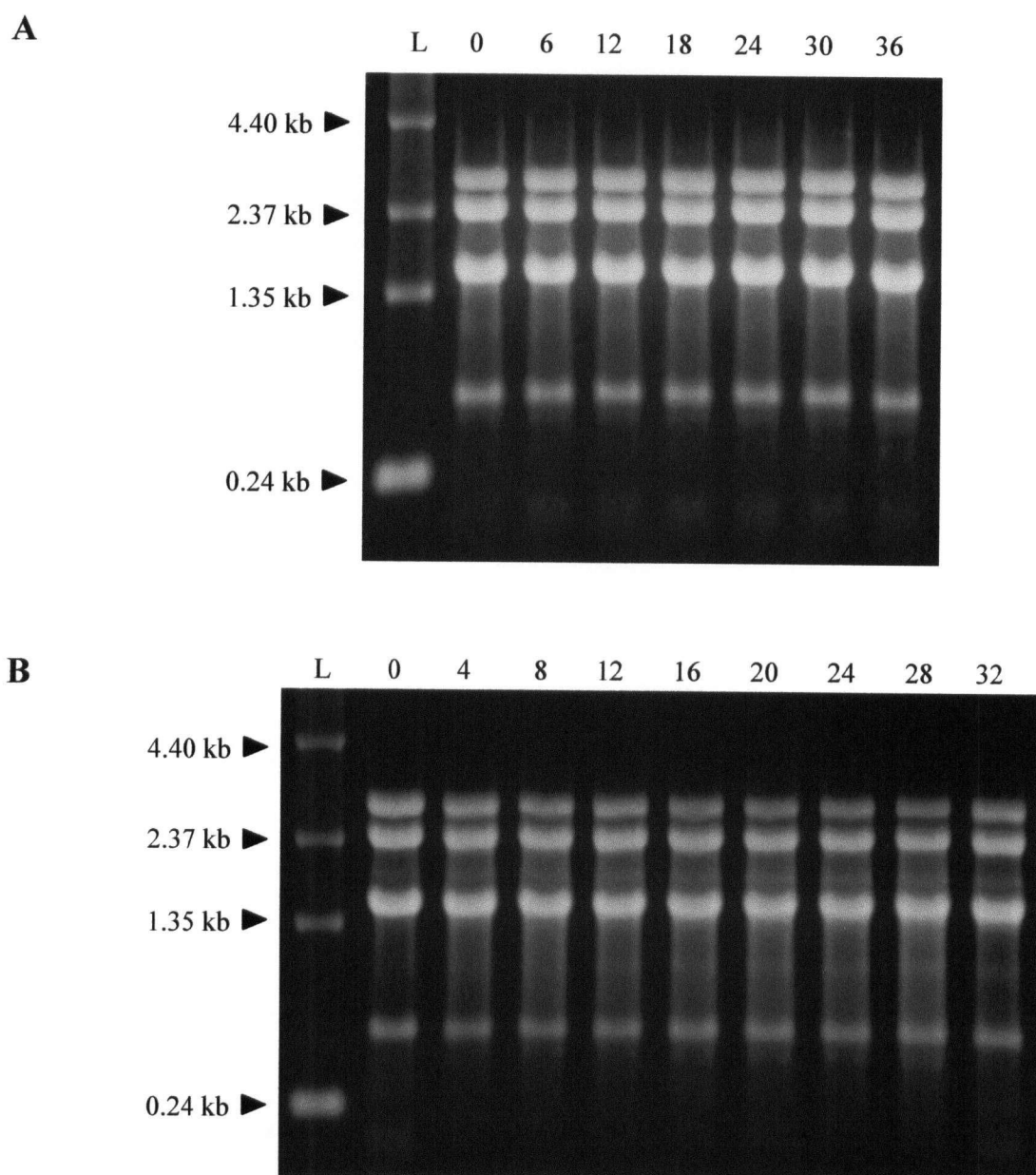
Figure 5.1. Total RNA isolated from *S. elongatus* and *Nostoc*. A) *S. elongatus* PCC 7942
RNA prepared using a hot-phenol extraction method as described in Section 2.10.1. B) *Nostoc*
PCC 9709 RNA prepared using Trizol as described in Section 2.10.3. Total RNA (5 μg) was
separated by electrophoresis on a 1.2 % agarose-formaldehyde gel and transferred to membranes
in preparation for Northern blot analysis. Numbers above each lane indicate the hour at which
cell samples were collected. L = 1 μg 0.24 – 9.5 Kb RNA Ladder. The quality of RNA
preparations is demonstrated as discrete bands of intact rRNA species. Cleavage of the 23S
rRNA (as shown by the additional presence of a 2.4 kb band) is common in cyanobacterial
species.
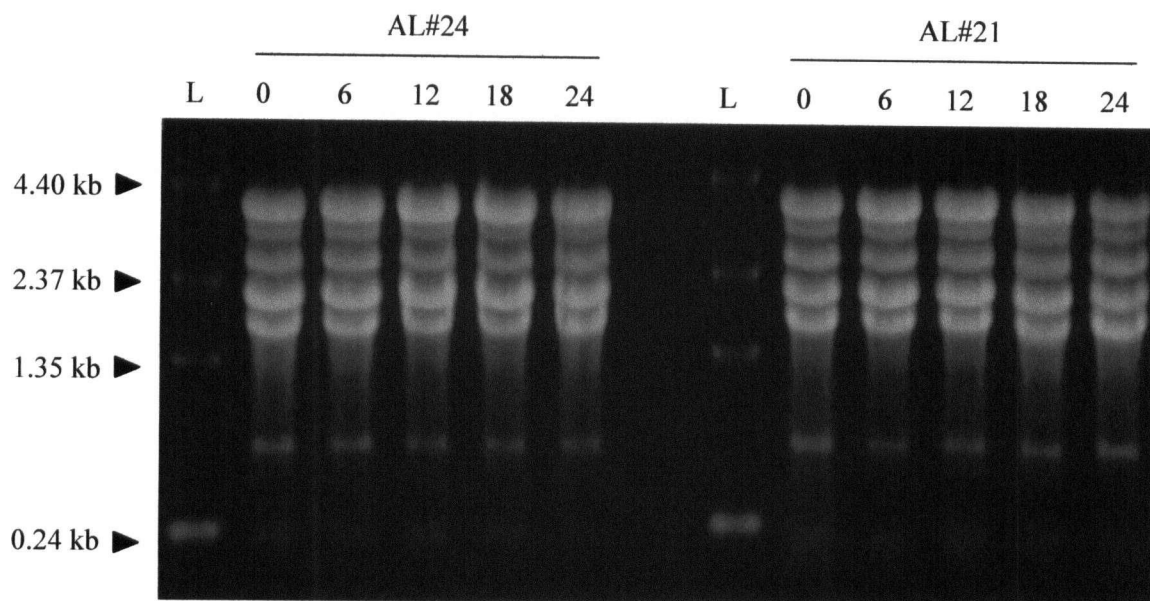
Figure 5.2. Total RNA isolated from *P. membranacea*. *P. membranacea* AL#24 and AL#21 RNA prepared using Trizol as described in Section 2.10.4. Total RNA (15 µg) was separated by electrophoresis on a 1.2 % agarose-formaldehyde gel and transferred to membranes in preparation for Northern blot analysis. Numbers above each lane indicate the hour at which cell samples were collected. L = 1 µg 0.24 – 9.5 Kb RNA Ladder. The quality of RNA preparations is demonstrated as discrete bands of intact eukaryotic and prokaryotic rRNA species. Cleavage of the 23S rRNA (as shown by the additional presence of a 2.4 kb band) is common in cyanobacterial species.

A
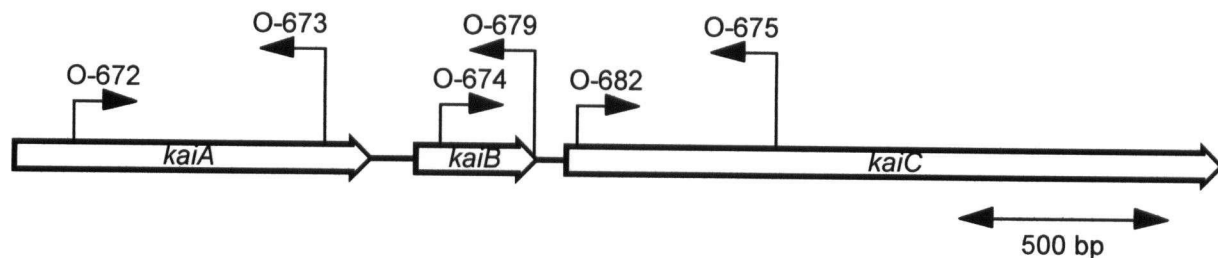
B
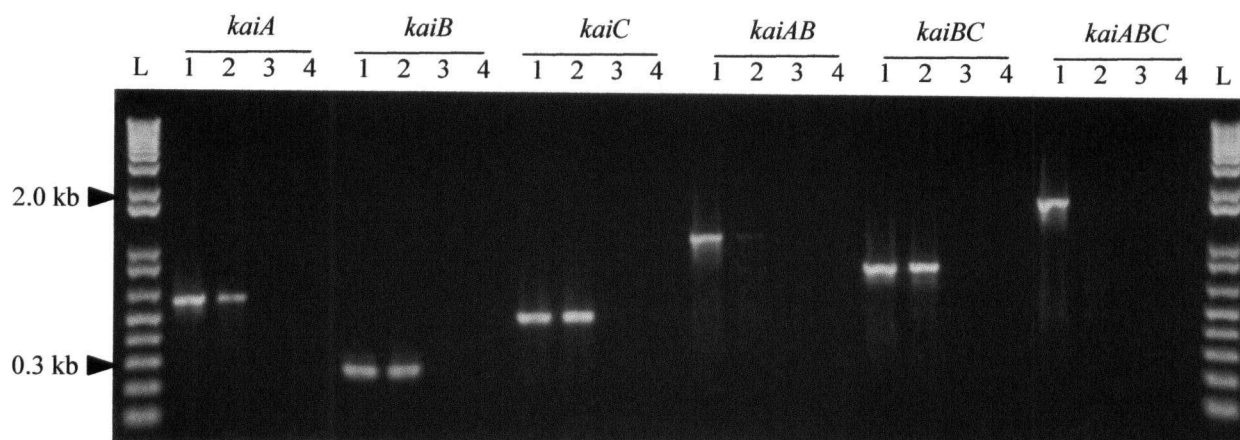
Figure 5.3. RT-PCR mapping of *kaiABC* transcripts in *S. elongatus*. A) Diagram of the gene cluster with locations of primers used for RT-PCR. B) 1 % agarose gel of RT-PCR products. 1 = PCC 7942 genomic DNA, 2 = PCC 7942 total RNA (+RT), 3 = PCC 7942 total RNA (-RT), 4 = water. L = 300 ng 1 Kb Plus DNA Ladder (Invitrogen). *S. elongatus* RNA template was purified from a 36 hour time point sample using the method described in Section 2.10.1. Reverse-transcribed products were subjected to 28 cycles of PCR amplification.
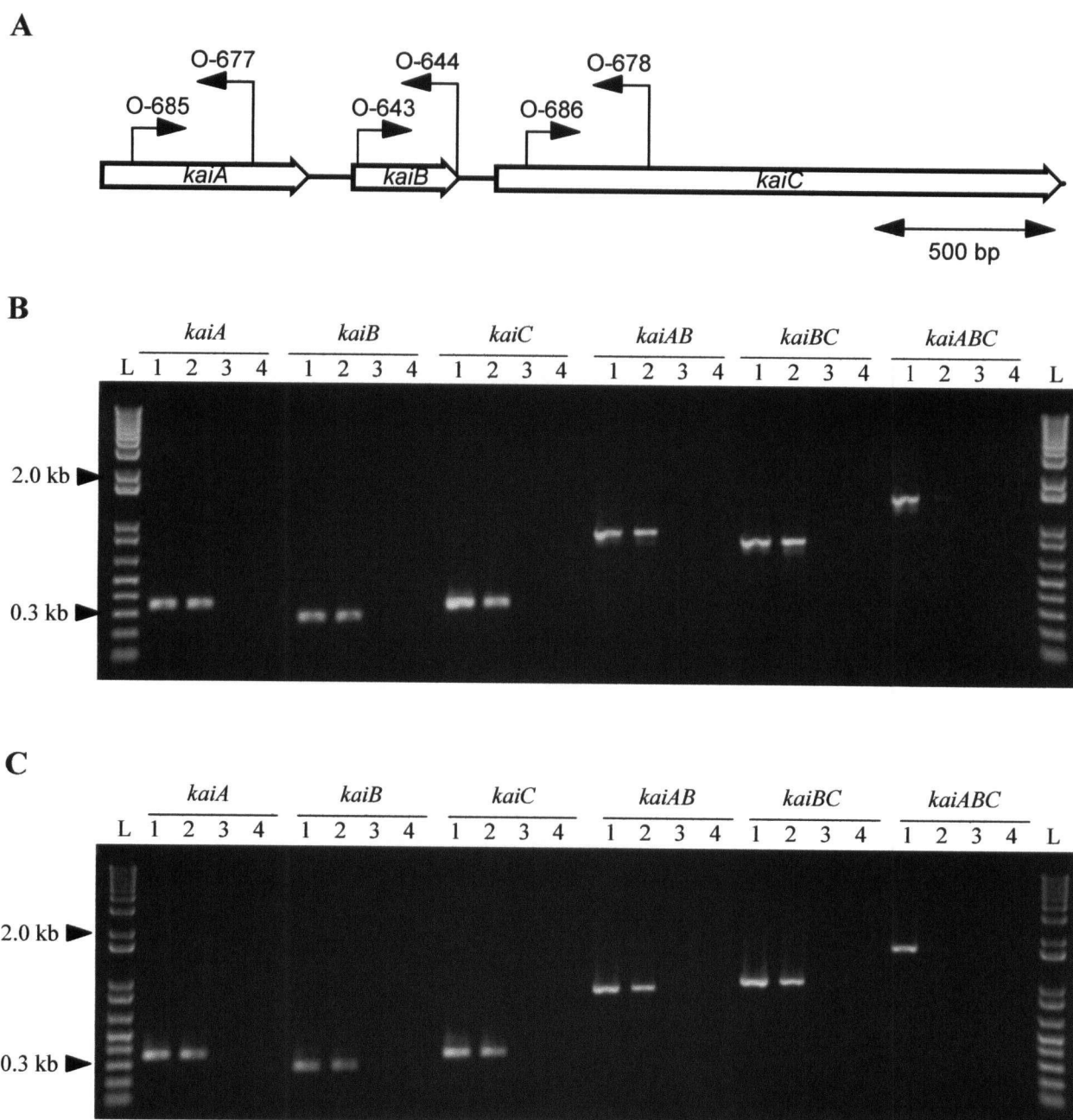
Figure 5.4. RT-PCR mapping of *kaiABC* transcripts in *Nostoc* and *P. membranacea*.
A) Diagram of the gene cluster with locations of primers used for RT-PCR. B) 1 % agarose gel of RT-PCR products for *Nostoc* PCC 9709. C) 1 % agarose gel of RT-PCR products for *P. membranacea* AL#21. RNA template was purified from a 1 = genomic DNA, 2 = total RNA (+RT), 3 = total RNA (-RT), 4 = water. L = 300 ng 1 Kb Plus DNA Ladder (Invitrogen). *Nostoc* and *P. membranacea* AL#24 RNA templates were purified from 12 hour time point samples using the method described in Section 2.10.2. and Section 2.10.4, respectively. Reverse-transcribed products were subjected to 30-34 cycles of PCR amplification.

to be significant for data interpretation because it is much weaker than the other positive signals for individual and/or combinations of gene targets. This minor product could possibly have arisen as a result of an unusual circumstance such as 'overlap' PCR in which the 3' end of the *kaiBC* cDNA anneals to the 3' end of the *kaiA* cDNA and acts as a primer for *Taq*-mediated polymerization, or it could reflect a minor subpopulation of longer *kaiA* transcripts (note that the location of the 3' *kaiB* primer allows for the possibility that the *kaiB* sequence is incomplete on these mRNA molecules, see Figure 5.3a).

RT-PCR mapping of *kai* transcripts in *Nostoc* and *P. membranacea* also indicated that *kaiABC* is transcribed as two mRNA molecules (Figure 5.4). However, it appears that *kaiB* is encoded by both transcripts in the filamentous cyanobacterium and in the lichen rather than solely with *kaiC* (which is the case for the *S. elongatus* model). In other words, the *Nostoc kaiABC* gene cluster appears to have one promoter upstream of *kaiA* and a second promoter upstream of *kaiB*, which is the same as what is known for *S. elongatus*. However, it appears that the first transcription stop site is further downstream for *Nostoc* than for *S. elongatus*. There was a very faint signal for *kaiABC* (lane 2) from *Nostoc* but not from the lichen, but this is not considered to be significant for data interpretation for reasons similar to those described in the previous paragraph. These RT-PCR results suggest that Nostoc KaiB is translated from two separate mRNA transcripts. Note that, as the locations of both *kaiB* primer sites are internal to the boundaries of the coding region, it is possible that either one or both transcripts lack a completely intact *kaiB* sequence (see also Section 7.2). Potential implications of multiple *kaiB* transcripts will be discussed in Section 5.9.

5.5. Methods for normalizing gene expression

Detection of daily oscillation patterns in *kai* mRNAs requires a means by which signals from individual samples can be compared to each other. Many different normalization methods have been utilized for research in diverse fields, and the relative merits of each type of standardization vary with the application (68, 72). Commonly used controls include: cell number, DNA mass quantity, RNA mass quantity, rRNA quantity, and mRNA quantity of constitutively expressed 'housekeeping' genes.

The choice of appropriate normalization controls depends on the nature of the sample population and the degree of variation to be measured. For these comparative studies on circadian rhythm, detection of a daily oscillation pattern did not require absolute quantification of *kai* mRNAs or even precise measurement of small differences, as changes in mRNA levels

were expected to be relatively large (visually estimated at five- to ten-fold in *S. elongatus* Northern blots (32)). The most appropriate normalization method for these studies would have been cell number, but this approach was impractical for lichen samples. Because all the samples in a given time course study were treated the same way, and the differences to be detected were large, the stringency of the normalization method was not required to be high. This allowed a greater range of choice in controls.

Housekeeping genes have the advantage over rRNA genes for normalization in that they tend to be present at levels similar to the gene of interest. The major pitfall of their use, and one that is shared with rRNA, is that constitutive expression does not necessarily entail a constant level of expression, particularly when a sample set includes different types of cells or cells that have been subjected to widely varying treatments. In contrast to eukaryotic systems for which many such controls have been developed (for example, β-actin or glyceraldehyde-3-phosphate dehydrogenase), there are no well-established housekeeping gene controls for cyanobacterial research. However, there have been previously published reports which used the *rnpB* gene (encoding the RNA component of the RNase P holoenzyme) as an internal control for expression in the filamentous heterocyst-forming cyanobacterium *Anabaena* (9, 10). *rnpB* is an evolutionarily conserved gene (found in Archaea, Bacteria and Eucarya) required during all stages of growth for tRNA processing (21), and its concentration in *Anabaena* appears to be fairly constant even when the organism is subjected to major environmental shocks such as shifts in temperature and light (10).

Oligonucleotide primers O-663 and O-664 (Table 2.3) for PCR amplification of cyanobacterial *rnpB* were designed based on a nucleotide alignment of sequences available from the Ribonuclease P Database (http://www.mbio.ncsu.edu/RNaseP/) for *S. elongatus* PCC 7942, *Nostoc* PCC 6705, *Nostoc* PCC 7107, *Nostoc* PCC 7413, *Anabaena* PCC 7120 and *Anabaena* ATCC 29413 (7, 27, 73, 74). Primers sites are located close to the 5' and 3' ends of the coding sequence and produce a 0.36 kb amplicon from *S. elongatus* (as expected) and a 0.39 kb amplicon from *Nostoc*. DNA sequence analysis of the *Nostoc* PCC 9709 product confirmed that it was highly similar to known *Nostoc* and *Anabaena rnpB* sequences. These *rnpB* primers and amplicons were used as the normalization control for relative RT-PCR analyses and as one of several normalization controls for Northern blot analyses described in the following sections.

5.6. Northern blot analyses of temporal gene expression patterns

As discussed in Section 5.5, each type of normalization control has its advantages and disadvantages. Rather than depend on any single control and its particular attendant weaknesses, Northern blot hybridization data was normalized between time course samples using three independent methods: a) against total RNA loaded (calculated from A_{260} measurements, b) against the *rnpB* hybridization signal, or c) against the 16S rRNA hybridization signal.

5.6.1. *S. elongatus*

Time course studies using Northern blot analysis to detect *S. elongatus kaiA* and *kaiC* gene expression have been previously published by Ishiura et al. (32). I performed a similar experiment on *S. elongatus* in order to confirm that the chosen technical procedures gave suitable levels of sensitivity and selectivity. The results matched previous reports. *S. elongatus kaiA* and *kaiBC* mRNAs appeared to be relatively labile and patterns of abundance were more easily observed in degradation products rather than the intact transcripts (Figure 5.5). Rhythmic abundance of expressed *kaiC* (but not *kaiA*) was observed over the 36 hour time period (1.5 circadian cycles), peaking at CT 12 and falling to a low at CT 0. Difficulty in detection of oscillation in *kaiA* mRNA using Northern hybridization methodology has also been experienced by others (S. Golden, personal communication). 16S rRNA and *rnpB* hybridization results support their utility as internal controls in that signals did not demonstrate rhythmic abundance over the course of time studied but appeared to remain relatively constant. From a purely technical point of view, *rnpB* appeared to be a better normalization control than 16S rRNA because the former was expressed at levels comparable to target genes and because densitometric analysis of the intense, sharply-defined signals was easier. Quantification of *S. elongatus* Northern blots was not performed as visual inspection was sufficient to meet the research objective of this control component of the time course studies (validation of method by reproducing results previously published in (32)).

5.6.2. *Nostoc*

Northern hybridization signals for *Nostoc kaiA* and *kaiC* gene expression were detected (Figure 5.6). The presence of intact *kai* transcripts was greater compared to *S. elongatus* results. Surprisingly, the *rnpB* probe hybridized to two RNA bands at 0.45 kb and 0.23 kb. In *Anabaena* (G. Owttrim, personal communication, see also (10)), *Synechocystis* (70) as well as *S. elongatus* (Figure 5.5), *rnpB* gene expression is visualized as a single band. It is possible that the *Nostoc rnpB* transcript undergoes cleavage into two equal fragments during the RNA extraction

process. Densitometric analysis was performed on the bands representing intact 1.0 kb *kaiA* and 2.7 kb *kaiC* transcripts, intact 1.7 kb 16S rRNA transcripts as well as the *rnpB* doublet (Figure 5.7). All three methods of quantitative analysis, as well as visual inspection of blots, support circadian cycling of *kaiC* in *Nostoc* PC 9709 over a 32 hour time course study. The peak of *kaiC* expression occurred at CT 12, the trough at 28 hours under constant illumination, and cycles back to increasing expression beyond that time point. It was more difficult to discern circadian cycling in *Nostoc kaiA*. Numerical and visual analyses suggest that there was a weak peak-and-trough pattern over the first 24 hour period of the time course study, but based on the last time point sample at 32 hours, there was no indication that the pattern would repeat during the subsequent 24 hour period. According to densitometric analyses, the amplitude of *kaiC* rhythm was 2.5:1 (peak to trough ratio) and the amplitude of *kaiA* rhythm was 2:1. The temporal graphs generated for *Nostoc kai* expression are not smooth curves but this may be due to technical reasons. Elevated background values and indistinct positive bands result in a low signal-to-noise ratio, which can increase the variability of measured values. The techniques utilized in this experiment are not perfectly appropriate for accurate and precise measurement of small differences in expression levels between one time point and the next. Discussion of these results will be continued in Section 5.9.

5.6.3. *P. membranacea*

Northern blots of lichen specimens AL#21 and AL#24 are shown in Figure 5.8. Because of the high sequence similarity between *Nostoc* PCC 9709 and lichen cyanobionts (Sections 4.3 and 4.4), PCC 9709-specific probes were used for analysis of lichen samples. No signal was detected for *kaiA* (data not shown). Because the *Nostoc* and cyanobiont genomic sequences share 88 % identity for this gene, it is likely that a lack of signal reflected a lack of transcription and not a lack of probe cross-reactivity. A strong *kaiC* signal was detected in both lichens, indicating an intact mRNA of the same size as for *Nostoc*. RNase P RNA and 16S rRNA signals for the lichen were also identical to *Nostoc* results, as expected. Both lichen specimens demonstrated a peak-and-trough pattern of kaiC expression, with the peak at CT 12 for AL#21 and CT 6 for AL#24. The changes in *kaiC* mRNA abundance were dramatic over the 24 hour period of this time course study. The amplitude of this pattern is greater than observed for *Nostoc* or *S. elongatus*, with a peak-to-trough ratio of 5:1 (AL#21) or 10:1 (AL#24) based on densitometry. These data are consistent with circadian cycling. Demonstration of pattern repetition was precluded by specimen mass limitations.
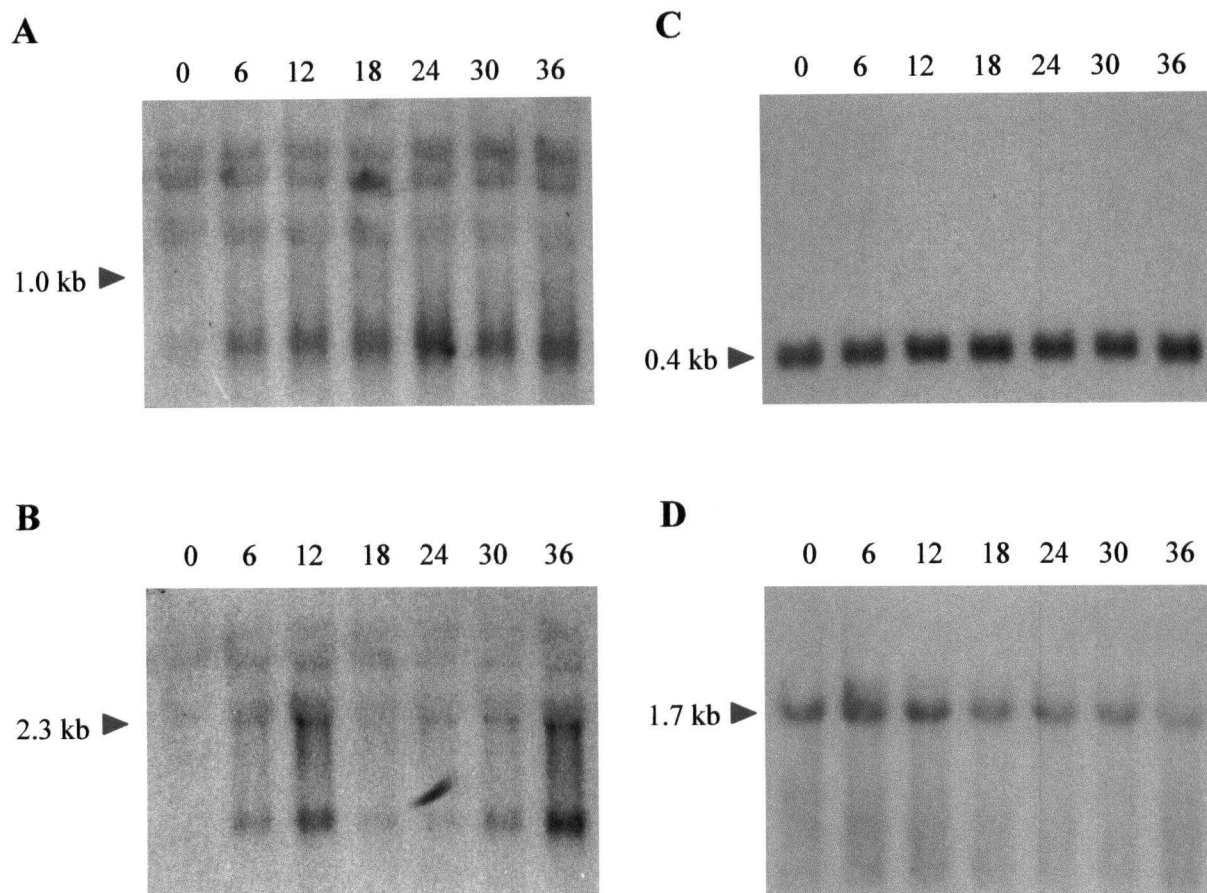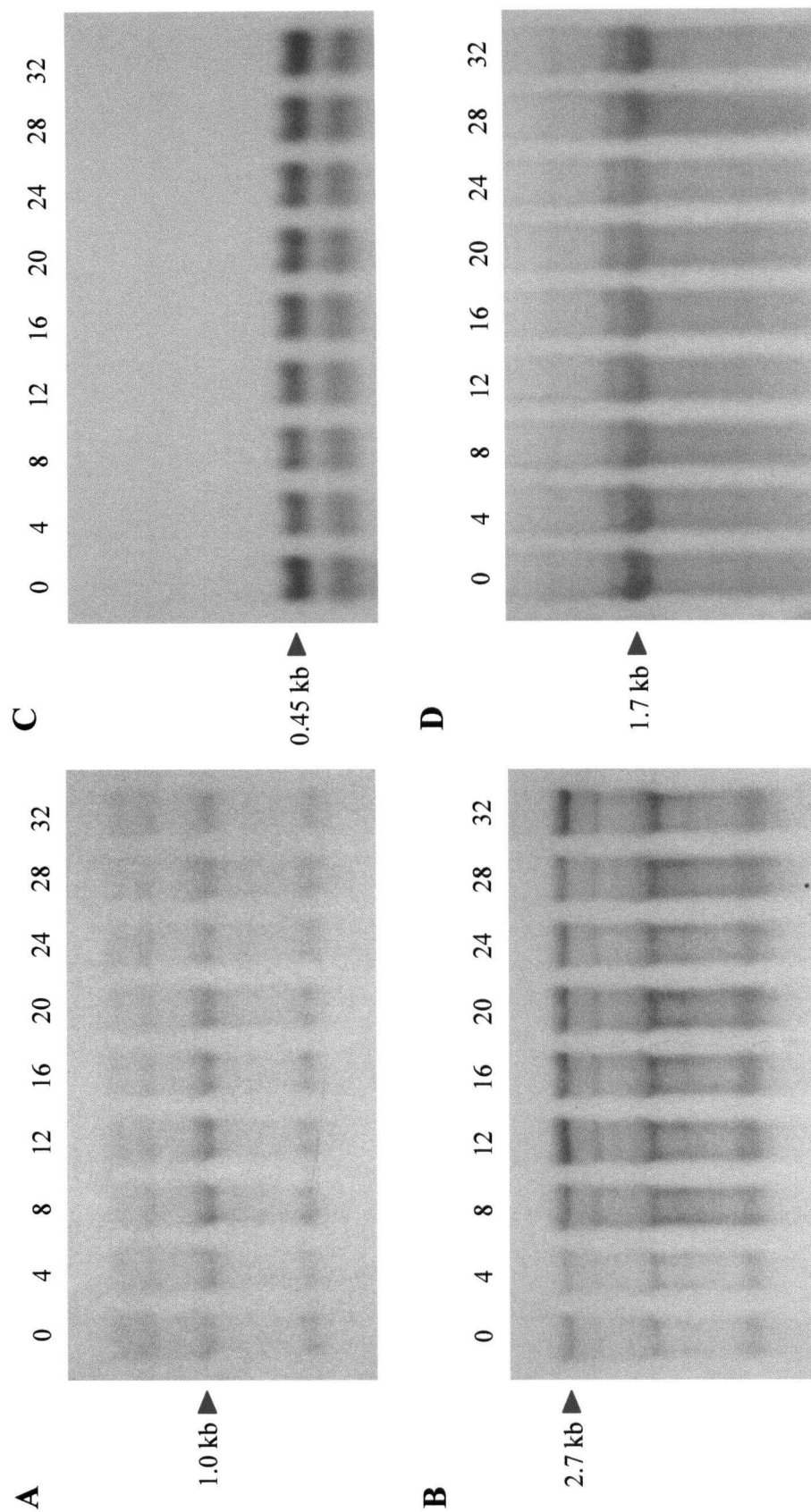
Figure 5.5. Northern blot analysis of *S. elongatus kai* gene expression. A) *kaiA*, B) *kaiBC*, C) *rnpB*, D) *16S rRNA*. Numbers above each lane indicate the number of hours each cell sample was exposed to constant light during a time course study. Arrows indicate the band of expected size. 5 µg total RNA was analyzed per sample. As previously observed ((32), also S. Golden, personal communication), the *kaiA* and *kaiBC* mRNAs appear to be relatively labile and patterns of abundance are more easily observed in degradation products rather than the intact transcripts.

Figure 5.6. Northern blot analysis of *Nostoc* PCC 9709 *kai* gene expression. A) *kaiA*, B) *kaiC*, C) *rnpB*, D) 16S rRNA. Numbers above each lane indicate the number of hours each cell sample was exposed to constant light during a time course study. Arrows indicate bands of expected size. 5 μg total RNA was analyzed per sample. See Figure 5.7 for densitometric analysis.
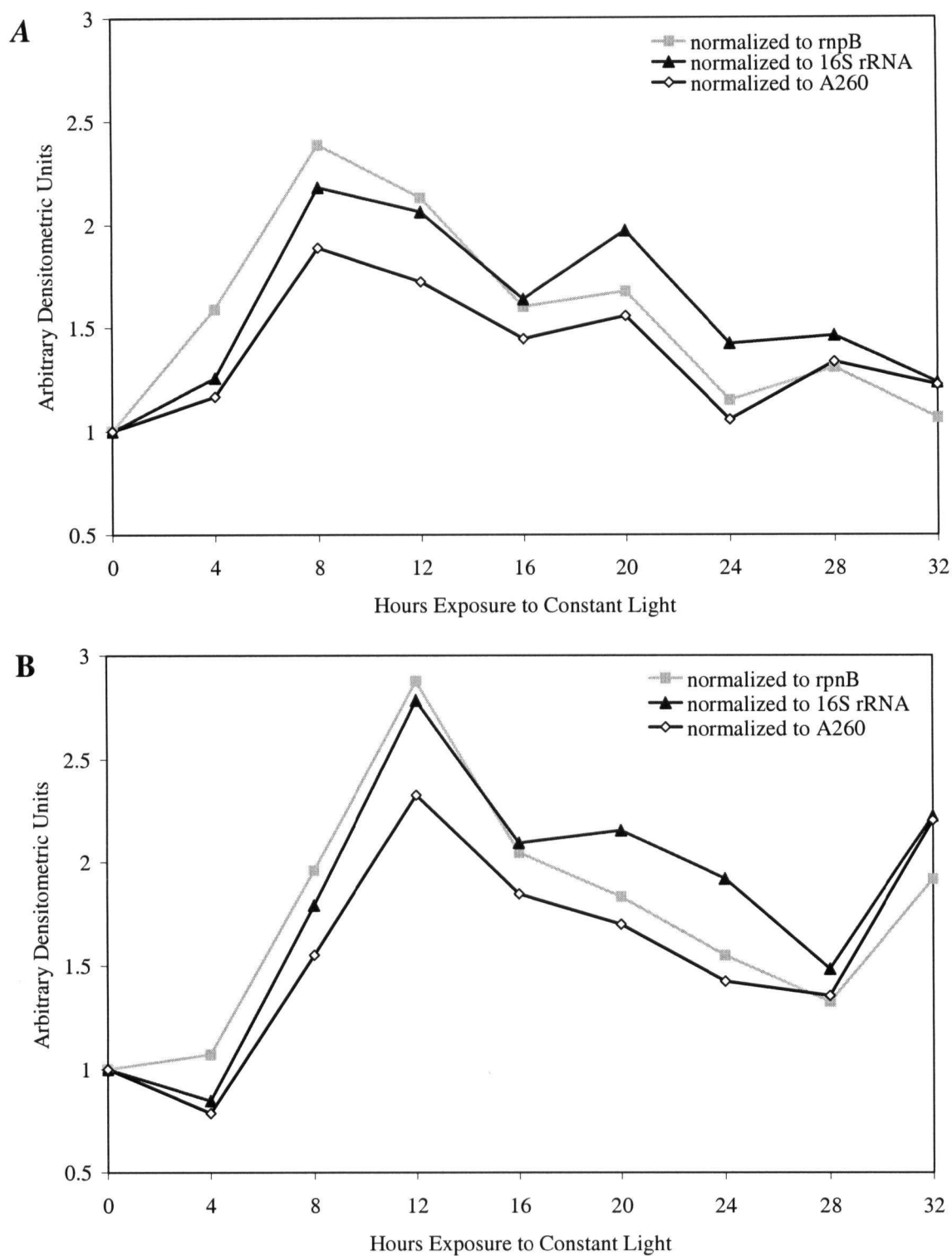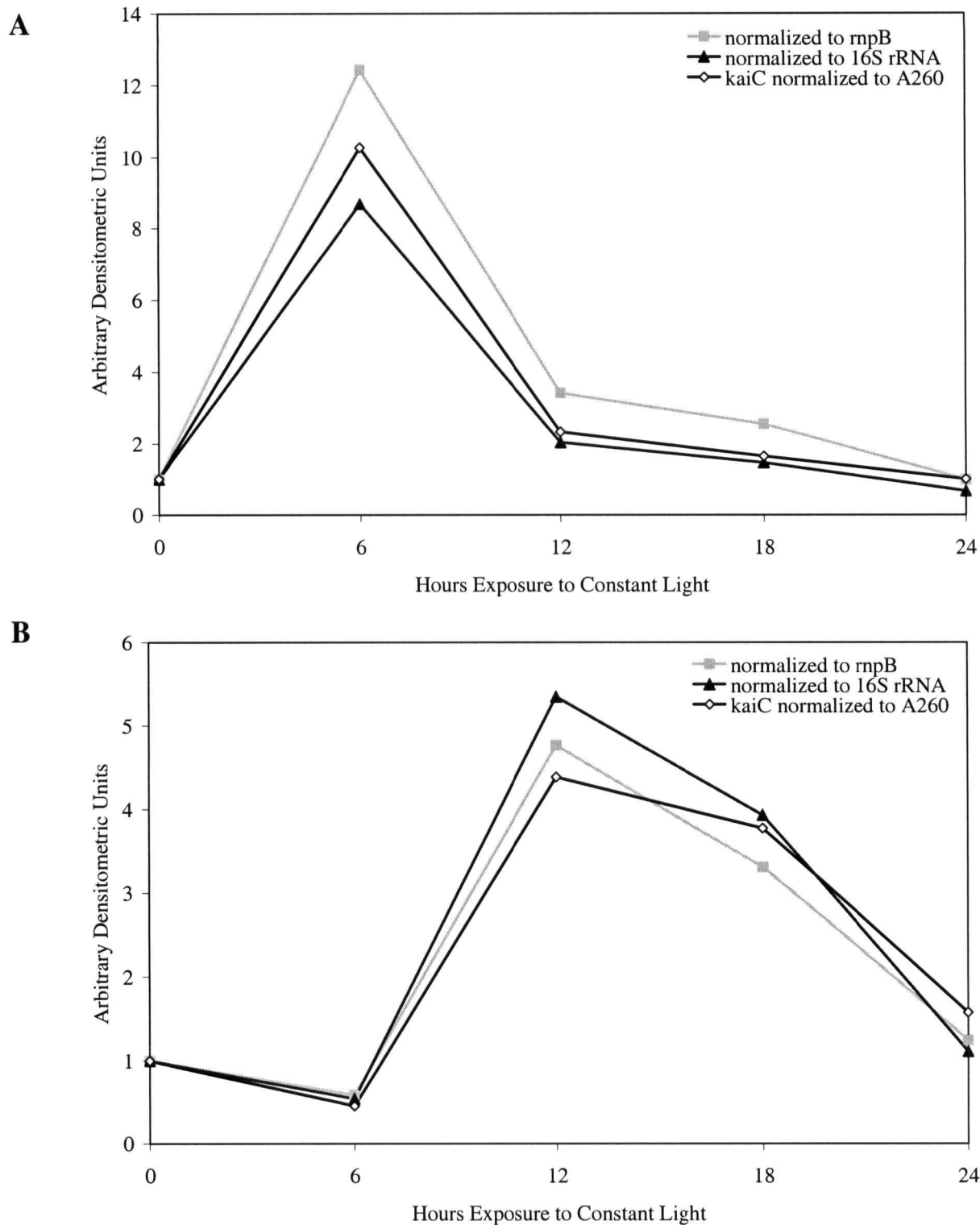
Figure 5.7. Graphical representation of *Nostoc kaiA* and *kaiC* gene expression as detected by Northern blot analysis. A) *kaiA*, B) *kaiC*. Clock gene hybridization signal (Figure 5.6) was normalized against either the *rnpB* signal, the 16S rRNA signal or the total amount of RNA loaded and then compared against the internal zero time point (arbitrarily set as one unit).
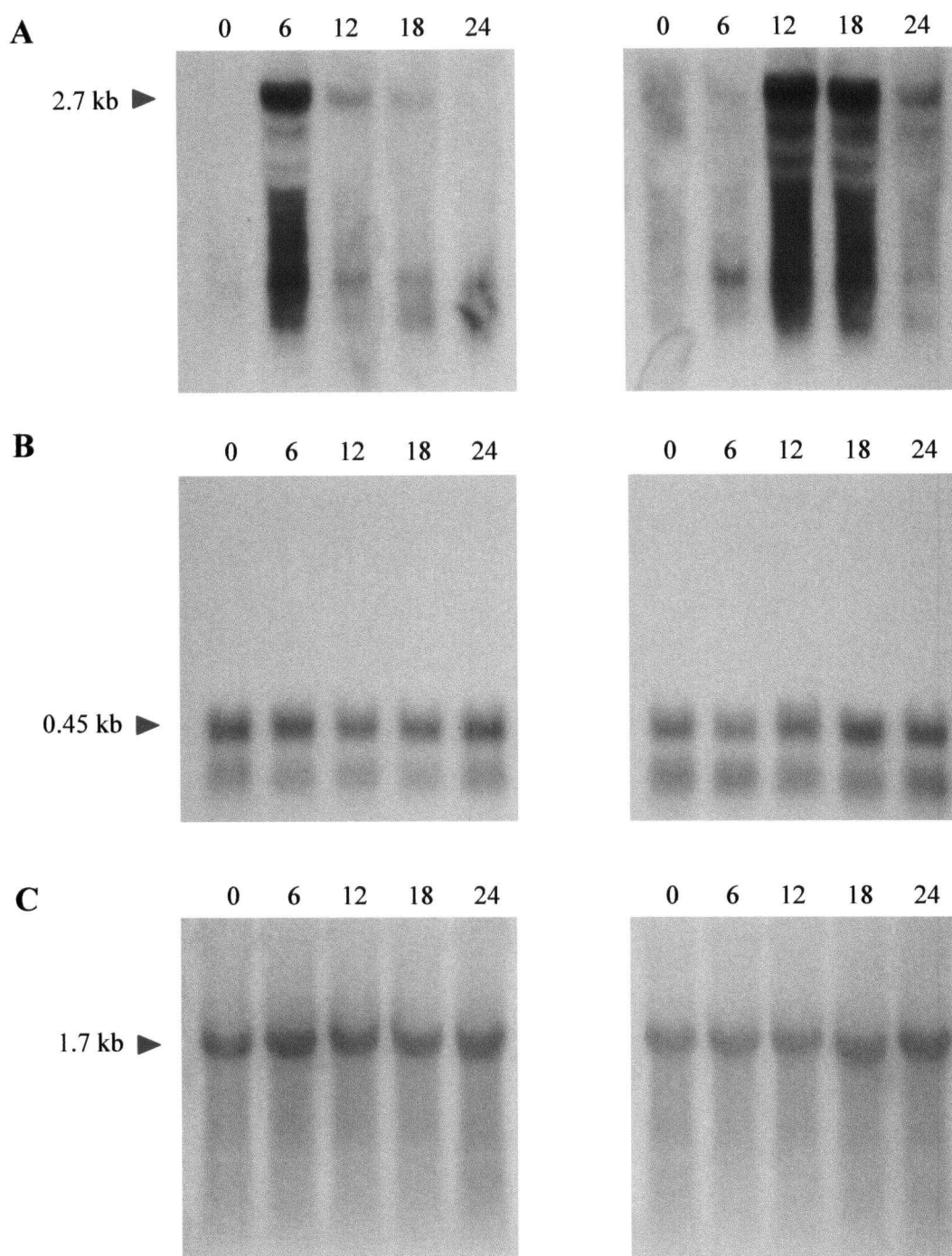
Figure 5.8. Northern blot analysis of *P. membranacea kaiC* expression. A) *kaiC* B) *rnpB*, C) 16S rRNA. PCC 9709-specific probe for *Nostoc kaiC* produced a strong hybridization signal whereas no signal was detected with the *Nostoc kaiA* probe (data not shown). Specimen AL#24 is shown on the left panel, and AL#21 is shown on the right panel. Numbers above each lane indicate the number of hours each cell sample was exposed to constant light during a time course study. Arrows indicate bands of the expected size. 15 µg total RNA was analyzed per sample. See Figure 5.9 for densitometric analysis.

Figure 5.9. Graphical representation of *P. membranacea kaiC* gene expression as detected by Northern blot analysis. A) AL#24, B) AL#21. Clock gene hybridization signal (Figure 5.8) was normalized against either the *rnpB* signal, the 16S rRNA signal or the total amount of RNA loaded and then compared against the internal zero time point (arbitrarily set as one unit).

5.7. Relative RT-PCR analyses of temporal gene expression patterns

The sets of *S. elongatus*, *Nostoc* and *P. membranacea* time course samples used for Northern hybridization experiments (Section 5.6) were also analyzed by relative RT-PCR assay.

5.7.1. Assay principles

There exist many methods utilizing the technique of RT-PCR for quantification of mRNA levels (reviewed in (22)). The goal of these *kai* expression time course studies was to detect oscillation patterns that, based on previous reports (32, 77), were anticipated to have peak-to-trough ratios of roughly 10:1. Therefore, because the research objective was to measure fairly large changes in target populations, neither absolute quantification nor stringent precision were necessary assay properties. RT-PCR-based relative quantification methods involve the use of an internal control for normalization of input materials and values must be collected during the exponential phase of PCR amplification. I chose to utilize a technically simple variant of an endpoint RT-PCR assay, called relative RT-PCR, for relative quantification which does not utilize competimers or require comparison to known quantities (17, 79). For this method, the exponential phase must be predetermined such that all given amounts of target mRNA display linear accumulation. Usually, the cycle number in the exponential phase ranges between 10 and 25; a lag phase is encountered with fewer cycles, and onset of the plateau phase comes with greater cycle numbers. Within the exponential phase, lower cycle numbers are chosen to avoid possible entry into plateau phase in samples that contain abundant target mRNA. Because these lower cycle numbers tend to result in amounts of PCR products that are difficult to discern on agarose gels stained with ethidium bromide, alternate visualization techniques such as Southern hybridization or PCR incorporation of markers are often required. For the *kai* gene expression studies described below, the *rnpB* gene was utilized as an constantly expressed internal control to normalize for variation in the amount of starting materials as well as efficiency of RT and PCR reactions.

5.7.2. S. elongatus

Figure 5.10 demonstrates the means by which the exponential phase for each organism/primer pair combination was determined. To represent the range of mRNA levels likely to be present within a given sample set, the amount of product amplified from CT 0 (minimum) and CT 12 (maximum) samples was evaluated after various numbers of PCR amplification cycles. The appropriate cycle number for each target was chosen such that it was below the plateau phase and resulted in enough product for ethidium -based gel quantification.

Similar to the approach taken for Northern hybridization experiments described in Section 5.6, a time course study of *S. elongatus* was undertaken in order to validate the relative RT-PCR quantification methodology. As shown in Figure 5.11, the data clearly demonstrate circadian cycling of *kaiC* mRNA with a strong peak signal (large amplitude) but no obvious rhythm for *kaiA* mRNA. These results are qualitatively consistent with results previously published by Ishiura et al. (32) as well as the Northern blot data described in Section 5.6.

To address the question of whether the invariant *kaiA* signal was representative of actual mRNA abundance or merely due to PCR reactions entering plateau phase, relative RT-PCR assays for *kaiA*, *kaiC* and *rnpB* were repeated using lower numbers of PCR cycles (22, 20 and 14 cycles, respectively). Because product amounts were too low for ethidium bromide visualization, gels were analyzed by radioactive Southern hybridization (data not shown). Results were consistent with the data shown in Figures 5.10 and 5.11, demonstrating that the ethidium bromide-based method is suitable for detection of circadian oscillation in *kai* gene expression.

5.7.3. *Nostoc*

Relative RT-PCR analysis of *Nostoc kai* gene expression is shown in Figure 5.12. Densitometric analysis of PCR products revealed data consistent with a circadian oscillation pattern (peak at CT 12, trough at CT 0) for both *kaiA* and *kaiC* over the first 24 hour period. However, repetition of this oscillation pattern was not demonstrated over the 32 hour length of the study. It is possible that this failure was simply due to technical difficulties with a single time point sample (at 32 hours); the entire time course study would have to be repeated with fresh culture material to investigate this possibility. These results were similar to the Northern blot data shown in Figure 5.7 over the initial 24 hour period, except that the peak-to-trough ratio was greater as measured by relative RT-PCR. The problematic 32 hour time point sample showed conflictingly results for *kaiC* (Northern blot showed an increase, relative RT-PCR did not. Further discussion is deferred to Section 5.9.

5.7.4. *P. membranacea*

Figures 5.13 and 5.14 show results for relative RT-PCR assays on lichen specimens AL#21 and AL#24. The *kaiC* expression patterns are very similar to those resulting from Northern blot analysis (Figure 5.9). As seen in *Nostoc* but not *S. elongatus*, *kaiA* expression patterns matched *kaiC* expression patterns in both lichen specimens. For both clock genes,
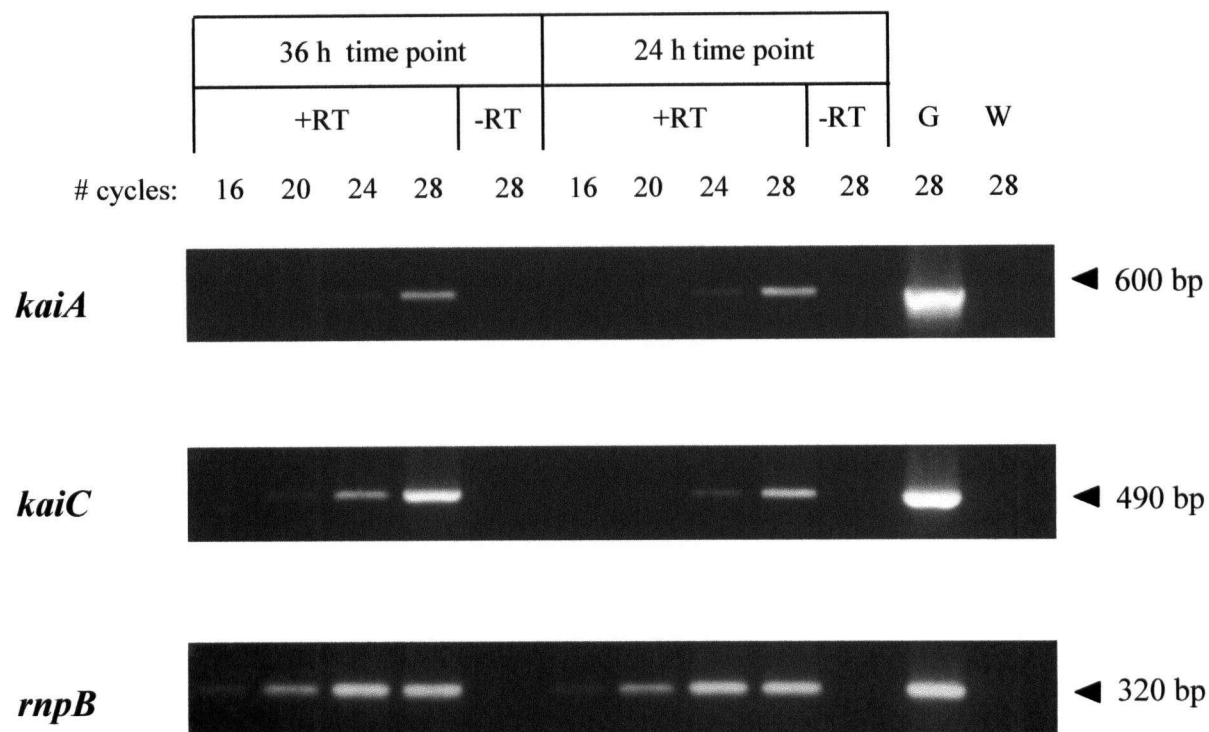
Figure 5.10. Calibration of amplification cycles for relative RT-PCR assay. *S. elongatus* RNA was subjected to a range of PCR amplification cycles for each primer pair. The 24 h (CT 0) and 36 h (CT 12) time points were chosen in the expectation that the two samples would represent the minimum and maximum expression levels within the full set of time course samples. Based on the gel results, an appropriate number of amplification cycles was chosen for each primer pair (in this case: 30 cyles for *kaiA*, 26 cycles for *kaiC*, and 20 cycles for *rnpB*). +RT = reverse transcriptase added to RNA. -RT = reverse transcriptase omitted. G = PCC 7942 genomic DNA. W = water.
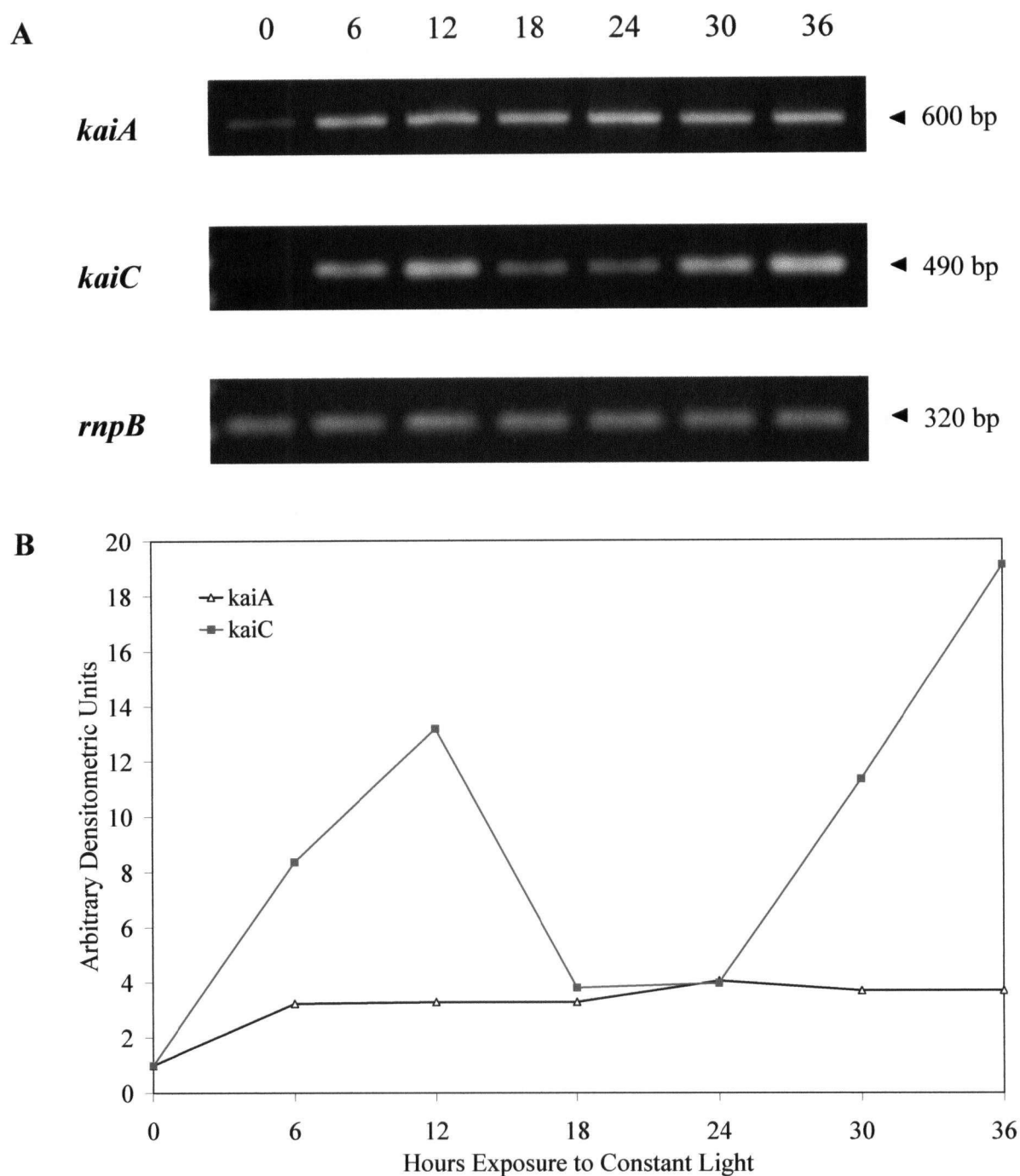
Figure 5.11. Assessment of the relative RT-PCR assay using model organism *S. elongatus*. A) RT-PCR products after electrophoresis on 1 % agarose gels. Numbers above each lane indicate the length of time (h) each cell sample had been exposed to constant light. Individual bands were quantified by densitometric analysis. B) Each *kaiA* and *kaiC* value was normalized against the corresponding *rnpB* value before being compared against the internal zero time point (arbitrarily set as one unit).
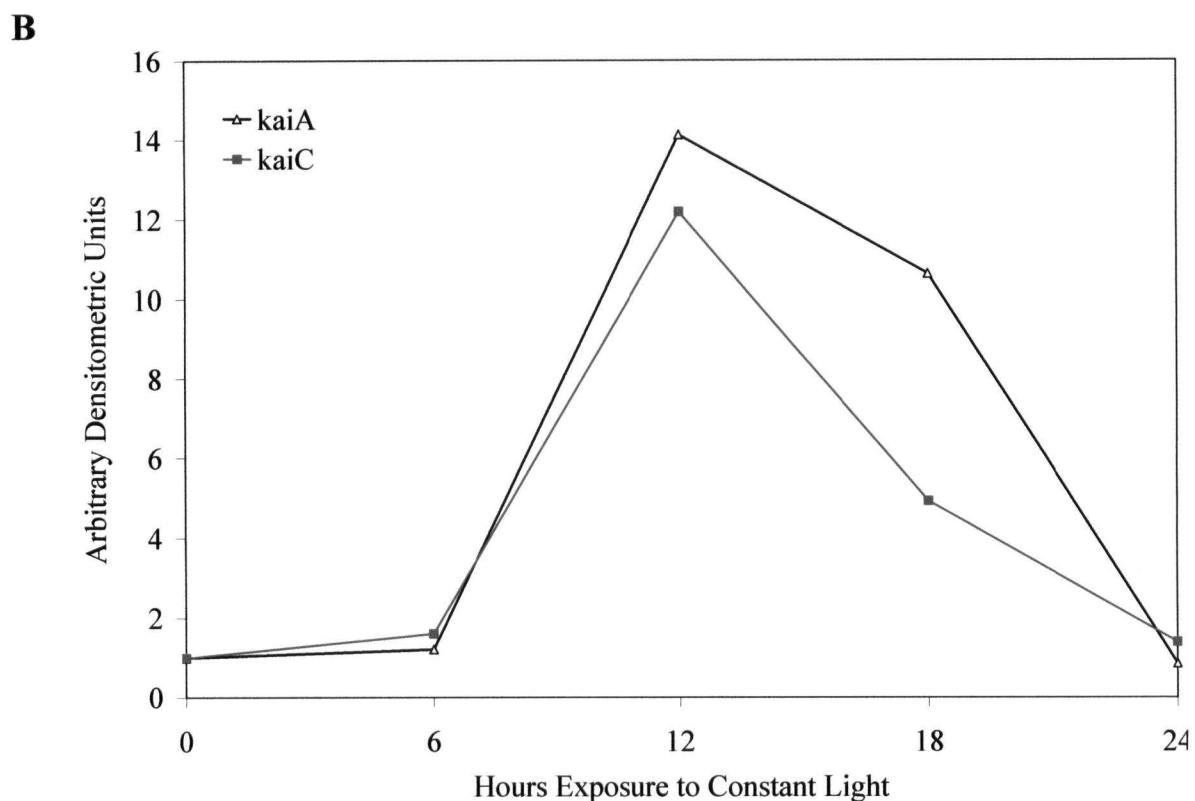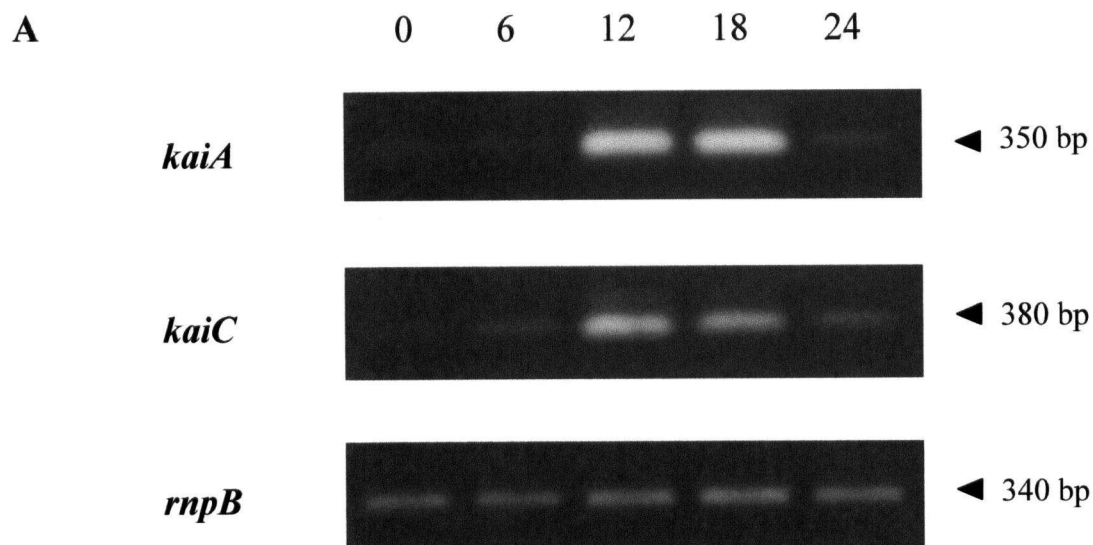
A

B

Hours Exposure to Constant Light
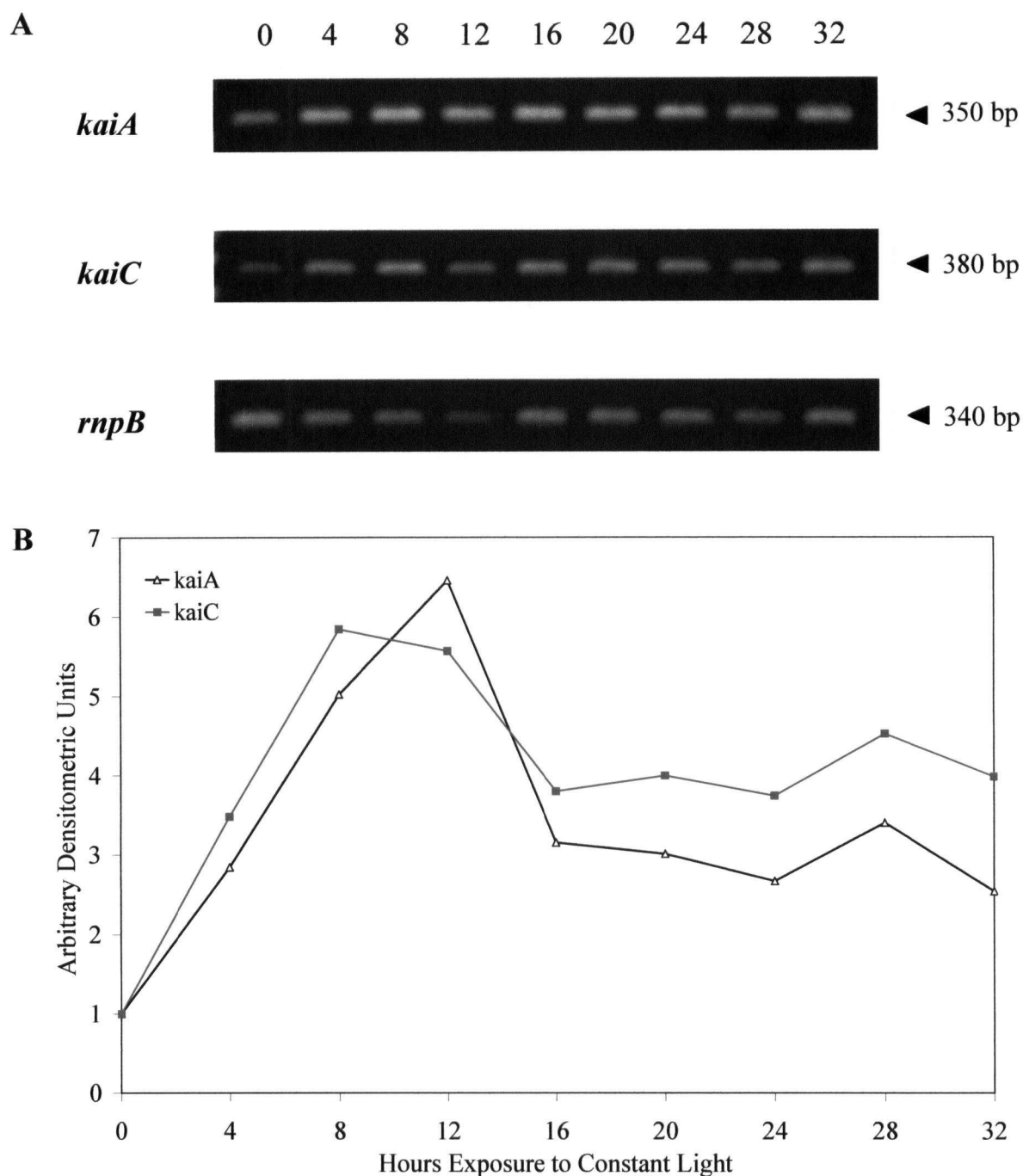
Figure 5.12. Analysis of *Nostoc kai* gene expression by relative RT-PCR. A) RT-PCR products after electrophoresis on 1 % agarose gels. Numbers above each lane indicate the length of time (h) each cell sample had been exposed to constant light. Individual bands were quantified by densitometric analysis. B) Each *kaiA* and *kaiC* value was normalized against the corresponding *rnpB* value before being compared against the internal zero time point (arbitrarily set as one unit).
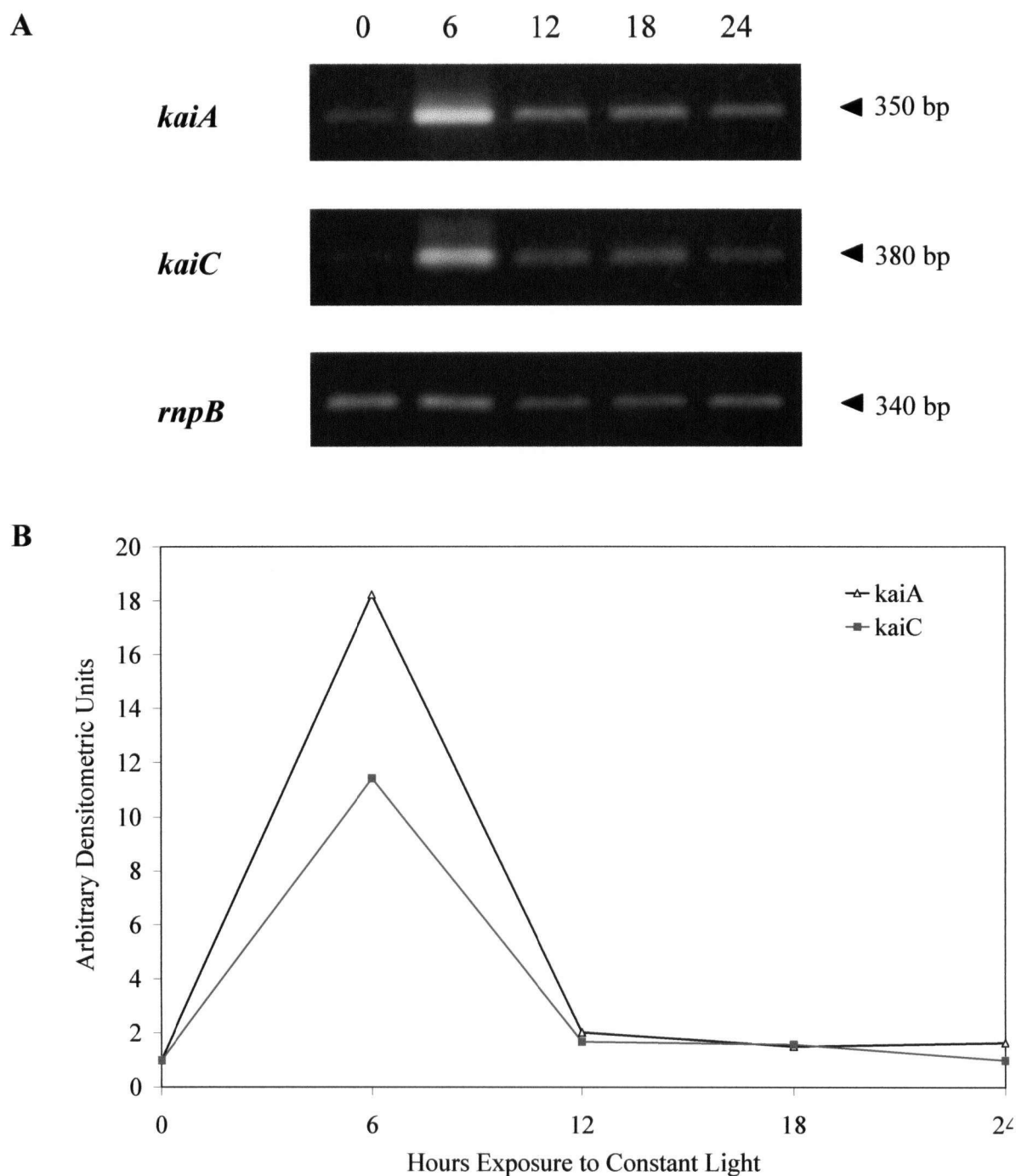
Figure 5.13. Analysis of *P. membranacea* AL#21 *kai* gene expression by relative RT-PCR.
A) RT-PCR products after electrophoresis on 1 % agarose gels. Numbers above each lane indicate the length of time (h) each cell sample had been exposed to constant light. Individual bands were quantified by densitometric analysis. B) Each *kaiA* and *kaiC* value was normalized against the corresponding *rnpB* value before being compared against the internal zero time point (arbitrarily set as one unit).

Figure 5.14. Analysis of *P. membranacea* AL#24 *kai* gene expression by relative RT-PCR. A) RT-PCR products after electrophoresis on 1 % agarose gels. Numbers above each lane indicate the length of time (h) each cell sample had been exposed to constant light. Individual bands were quantified by densitometric analysis. B) Each *kaiA* and *kaiC* value was normalized against the corresponding *rnpB* value before being compared against the internal zero time point (arbitrarily set as one unit).

temporal gene expression pattern was consistent with circadian oscillation over the 24 hour length of the time course study.

It must be noted that the actual peak of *kaiA* and *kaiC* gene expression differed between the two lichen specimens, with the greatest expression detected at CT 12-18 for AL#21 but earlier at CT 6 for AL#24. This discrepancy was consistent between the Northern hybridization and relative RT-PCR datasets. Given that these two field specimens were collected at the same time from the same geographic location, and were treated in an identical fashion in the laboratory, this result is somewhat puzzling. It would be useful to analyze additional lichen specimens, and more frequent time points, to investigate the repeatability of these data.

5.9. Discussion

In this chapter, some aspects of *Nostoc* circadian clock gene expression were characterized. Evidence supporting circadian cycling of *kai* gene expression in axenic and lichenized *Nostoc* was collected using two independent methods.

5.9.1. Northern blot and relative RT-PCR assays are complementary

S. elongatus cell samples were used to validate the performance of two different methodological approaches for the quantification of mRNA. Northern blot hybridization is a time-honored and well-understood technique that requires large expenditures of sample and hands-on work. Relative RT-PCR is an assay based upon complex reaction kinetics, combining greater sensitivity and selectivity with the capacity for high-throughput analyses. The data sets collected using these two methods are complementary. Northern hybridization detected mRNA length and offered a glimpse at stability and degradation characteristics of target transcripts. The sensitivity of relative RT-PCR allowed improved detection of oscillation patterns with low amplitude, and the form of signal output was accommodated densitometric analysis well. In addition, the relative RT-PCR assay detected *kaiA* mRNA in lichen samples whereas Northern blots failed in this regard. Both methods achieved the primary objective of detecting temporal changes in target mRNA abundance, and the consistency between datasets increases confidence in the conclusions.

5.9.2. *Nostoc* and *S. elongatus* - similarities and differences in clock gene expression

It was found that *Nostoc* shared certain *kai* gene characteristics with the cyanobacterial circadian model organism *S. elongatus*: 1) *kaiABC* is expressed as two transcriptional units, 2) *kai* transcripts are unstable, 3) kaiC mRNA has rhythmic abundance, peaking at CT 12 and

bottoming out at CT 0. This supports the hypothesis that the *Nostoc kaiABC* system is a circadian oscillation mechanism as discovered in *S. elongatus*. However, unlike what is known for the *S. elongatus* model, rhythmic abundance of *Nostoc kaiA* is strongly detected with amplitude similar to that of *Nostoc kaiC*. This implies that the differential promoter regulation between *kaiA* and *kaiC* in *S. elongatus* does not exist in *Nostoc*. In addition, *Nostoc kaiB* appear to be transcribed on both the *kaiA* and the *kaiC* mRNAs. The functional relevance of these differences between the *S. elongatus* model and the *Nostoc* system is unknown, but given that current theoretical models of the molecular mechanism underlying clock rhythm describe complicated and changing ratios of Kai proteins within multimeric complexes (37, 53, 78), these disparities are likely to be significant.

5.9.3. Why are rhythms in *kai* expression more pronounced in lichenized *Nostoc*?

For both the cultured *Nostoc* strain PCC 9709 and lichen specimens, temporal patterns of *kaiA* and *kaiC* gene expression were generally consistent with circadian oscillation. However, Northern blot analysis of *kaiC* expression showed that rhythms were far stronger (greater amplitude) in the lichenized *Nostoc* than in the isolated cultured cyanobacterial strain. One possible explanation could involve the relative metabolic states of the two types of cells. It has been reported that circadian rhythms dampen with cultures age into senescence and decreased metabolic activity (36). The *Nostoc* PCC 9709 was observed to be very slow-growing, and it seems reasonable to associate this with low metabolic activity in axenic culture. In contrast, cyanobionts in a lichenized state are observed to have greater nitrogen-fixation capabilities and also fix carbon for both themselves and their fungal symbionts (1, 49). Perhaps the lichenized *Nostoc* had a greater metabolic load that necessitated the physiological optimization provided by a *kai*-encoded circadian regulatory program. The reasons for this difference in *kai* expression level are unknown, and intriguing.

CHAPTER SIX

Conclusions

The results of this study demonstrate that the prokaryotic circadian clock gene cluster *kaiABC* is conserved in *Nostoc* PCC 9709. I have identified a single homologous copy of the *kaiABC* gene cluster (encoding the *S. elongatus* central clock oscillator) in the *Nostoc* genome. Deduced amino acid sequence similarity (46 % identity for KaiA, 81 % for KaiB, 80 % for KaiC) is consistent with conservation of function in accordance with available structure-function data for the *S. elongatus* model. Functional studies conducted to assess *kai* transcriptional activity provide evidence that temporal patterns in *Nostoc* clock gene expression are consistent with circadian oscillation, particularly for lichenized cyanobionts. *Nostoc kaiABC* differed from the model system in a few aspects: 1) the deduced KaiA polypeptide is two-thirds the length of its *S. elongatus* counterpart and shows sequence conservation only within the C-terminal third of the protein, 2) *kaiB* appears to be transcribed on both *kai* mRNAs and not just with *kaiC*, and 3) transcriptional activity of *kaiA* and *kaiB* are differentially regulated in *S. elongatus* but appear to be similarly regulated in *Nostoc*. In my opinion, these differences do not weaken the overall conclusion that the *Nostoc kaiABC* gene cluster encodes a circadian oscillator, but they may provoke adjustment of any 'universal' model for the molecular mechanism underlying the cyanobacterial circadian clock.

CHAPTER SEVEN

Future Research

Nostoc sp. PCC 9709 is a less-than-ideal laboratory subject for many common types of molecular studies due to the simple fact of its poor growth characteristics in axenic culture. A more prudent choice of filamentous cyanobacterial strain would be *Nostoc punctiforme* ATCC 29133 (also known as PCC 73102), which grows rapidly even under dark heterotrophic culture conditions (60), is amenable to genetic manipulations (12), has a known genome sequence (50) and can reconstitute symbiotic associations with the hornwort *Anthoceros* and the angiosperm *Gunnera* (49). Nevertheless, there are some additional studies of *Nostoc kaiABC* that may be performed on PCC 9709 with a relatively modest expenditure of research effort. The experiments proposed in the three sections below would complement and add to the work already described in this thesis. The last section briefly describes potential studies on the circadian clocks of cyanobacteria in symbiosis and utilizes *Nostoc punctiforme* ATCC 29133.

7.1. Advances in techniques for quantifying gene expression

Northern blot analysis of gene expression is a time-honoured technique, which can be accomplished with the tools available in most genetic research laboratories. Its use for the quantification of gene expression has, in recent years, become less favoured due the advent of real-time PCR. Quantitative real-time RT-PCR, which monitors the emissions of fluorescent dyes during the course of PCR amplification, offers certain advantages: it requires far less sample material, far less hands-on effort, and – once a specific assay has been developed – far less time to obtain results. Balancing these advantages are the relatively high cost of an instrument, expensive reagents and consumables and also a different cost in terms of potentially greater time and effort spent on assay development, as it is not yet a technique that is as well-understood as Northern blot analysis.

Quantitative real-time RT-PCR analysis of the *Nostoc kaiABC* system offers the opportunity to verify Northern blot and relative PCR data described in Chapter Five. As this method requires very little sample material per reaction and has high throughput capacity in 96-well and 384-well systems, many more experimental conditions and replicate measurements could be analyzed thereby offering improvements in the accuracy and precision of the data gathered.

An ideal future experiment would utilize a multiplex assay for *kai* gene expression in cyanolichens, using 16s rRNA as the control or baseline by which to measure changes in target mRNA levels. In the event that a multiplex assay was found to be difficult to optimize, the experiment could be performed using separate quantitative RT-PCR reactions for each target or control mRNA; this approach is somewhat less elegant in conception, but it is also less fraught with technical challenges. Lichens are an intriguingly cryptic class of organisms, and it would be worthwhile to improve upon the promising *kai* expression data already obtained for *P. membranacea*. In particular, the current dataset would be enhanced by the expansion of the time period beyond 48 or even 72 hours, to prove that the 24 hour peak-and-trough pattern is repeated through consecutive cycles.

7.2. Defining the transcript map of the *Nostoc kaiABC* gene cluster

In Section 5.4, RT-PCR mapping of transcriptional units indicated that the *Nostoc kaiABC* gene cluster is expressed as two separate *kaiAB* and *kaiBC* mRNAs. This suggests that the *Nostoc* system is somewhat different than the *S. elongatus* model, in which *kaiA* is transcribed separately from *kaiB* and *kaiC*. This possibility could be conclusively addressed by performing S1 nuclease protection assays to define the 5' and 3' ends of the two Nostoc *kai* mRNAs, as the positions of the *kaiB* primers (Figure 5.4a) do not prove than the *kaiB* sequence is completely intact on either of the two *Nostoc* transcripts. This study has functional significance in that the separation of *kaiA* expression from *kaiB* and *kaiC* expression, as well as relative abundances of KaiA, KaiB and KaiC proteins, are thought to be integral components of the oscillation mechanism (32, 37, 78). A change in the promoter for *kaiB* or having *kaiB* expression from more than one promoter may affect current theoretical models for the mechanism of rhythm generation in cyanobacterial circadian programs.

7.3. Protein function of *Nostoc* KaiA, KaiB and KaiC

Based on numerous published reports (29, 37, 53, 76), it now seems clear that the function of the *S. elongatus* circadian oscillation mechanism depends on multiple complex interactions between the three *kai* gene products as well as other protein components. It follows that research on other circadian systems in related organisms should address the protein function of homologous genes.

In an experiment performed at the laboratory of Dr. Susan Golden (Department of Biology, Texas A&M University), wildtype, *kaiA⁻* and *kaiC⁻* versions of the *Nostoc kaiABC*

gene cluster were inserted into NSI (neutral site I) of the *S. elongatus* PCC 7942 genome using suicide vector pAM2314. The *S. elongatus* host strains were either wildtype, *ΔkaiA*, *ΔkaiB*, or *ΔkaiC* at the *kai* locus and all carried a P$_{kaiB}$:luciferase reporter cassette inserted into NSII, which enabled detection of the rhythm of the circadian clock. Reporter luminescence for *kaiB* promoter activity in up to 15 individual transformants from each complementation was monitored for assessment of temporal patterns. No transformants of *kai* mutant strains exhibited a reversion of reporter gene expression to wildtype patterns. However, luminescence graphs were altered when compared to untransformed controls (data not shown). Additionally, insertion of *Nostoc kaiABC* into wildtype *S. elongatus* destroyed the rhythmic circadian pattern of the host strain (Appendix D). Together, these results suggest that the *Nostoc kai* elements did interact with the *Synechococcus kai* system in some manner but stronger conclusions cannot be drawn from the data.

Given that mutational analysis of *S. elongatus kaiABC* has demonstrated that physical interactions between Kai proteins can be easily perturbed or destroyed by many different single site mutations (54, 67), it is perhaps naive to expect *Nostoc* KaiA, KaiB or KaiC to be capable of functional complementation in *S. elongatus* mutants. Alternately, physical interactions amongst the *Nostoc* Kai homologues could be studied using yeast two-hybrid systems, as previously described for *S. elongatus* clock proteins in (34). Two-hybrid experiments are a suitable choice for investigation of *Nostoc* Kai protein function because these required neither genetic manipulation of the *Nostoc* PCC 9709 nor use of surrogate (and possibly incompatible) strains.

7.4. Adaptive significance of cyanobacterial circadian clocks in symbiotic associations

Circadian programming is a global regulatory system that brings about optimization of an organism's biochemical processes. A functional *kai* gene cluster is not a requirement for cyanobacterial growth, as *S. elongatus kai* mutants grow in well in pure culture at the same rate in constant light and in alternating light/dark cycles (36). However, the *kaiABC*-encoded circadian clock has been clearly demonstrated to have a functional role in competitive environments, such that a strain with a circadian period most-closely matching the environmental light/dark cycle has greater reproductive fitness.

One intriguing question is whether a functional circadian program is as necessary for symbiosis-adapted cyanobacteria as it is for free-living strains. Depending on the nature of the

symbiotic association, cyanobionts are isolated from the environmental conditions experienced by free-living varieties. In the case of *Gunnera*, the cyanobiont is located intracellularly within the interior of the stem glands, has a increased nitrogen fixation rate, a highly decreased photosynthesis rate, and receives fixed carbon from the plant host (5). With regards to lichen *Peltigera membranacea*, cyanobacteria are entwined with fungal hyphae in a layer just below the surface of the thallus and supplies fixed carbon and likely fixed nitrogen to the mycobiont (1). Particularly with *Gunnera*, cyanobionts may not need to be closely attuned to the normal day/night cycle. This idea could be investigated using genetically engineered *kai* mutants of *N. punctiforme* ATCC 29133. Under controlled light/dark conditions (no available fixed nitrogen), would *Gunnera* specimens associating with wildtype, long-period mutants, short-period mutant or null *kai* mutants grow equally well? Expression patterns of clock genes and nitrogen fixation genes could be monitored on the mRNA and the protein levels to determine whether the cyanobiont's inherent circadian program was the dominating regulator of nitrogen fixation or whether the angiosperm partner could exert control over this metabolic pathway. If it is possible that the *kai* clock mechanism may not be a universally critical function for cyanobacteria, it seems likely that evidence for this may be found amongst the strains that engage in the most intricate symbiotic relationships.

Along a similar line of thought, Dvornyk et al. (18) suggest that *kai* gene polymorphism in free-living *Nostoc linkia* increases with higher selective pressure in harsh environments, resulting in high sequence diversity and even multiple duplications of the *kai* gene cluster. It follows then, in the converse fashion, that cyanobacterial strains which specialize in sheltered, long term symbiotic associations may exhibit evidence of lowered selective pressure on the *kaiABC* gene cluster. This possibility could be addressed by performing a genetic survey to compare *kai* sequences to common phylogenetic markers such as 16S rRNA or *rbcL* (ribulose bisphosphate carboxylase-oxygenase large subunit) (62, 71).

REFERENCES

1. **Ahmadjian, V.** 1993. The Lichen Symbiosis. John Wiley & Sons.

2. **Allen, M. M.** 1968. Simple conditions for growth of unicellular blue-green algae on plates. J. Phycol. **4:**1-4.

3. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403-410.

4. **Ausubel, F. M., R. Brent, R. E. Kingston, D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** Current Protocols in Molecular Biology, 2001 ed. John Wiley & Sons, New York.

5. **Bergman, B., A. N. Rai, C. Johansson, and E. Soderback.** 1992. Cyanobacterial - plant symbioses. Symbiosis **14:**61-81.

6. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513-23.

7. **Brown, J. W.** 1999. The ribonuclease P database. Nucleic Acids Res. **27:**314.

8. **Bustos, S. A., and S. S. Golden.** 1991. Expression of the *psbDII* gene in *Synechococcus* sp. strain PCC 7942 requires sequences downstream of the transcription start site. J. Bacteriol. **173:**7525-7533.

9. **Chamot, D., W. C. Magee, E. Yu, and G. W. Owttrim.** 1999. A cold shock-induced cyanobacterial RNA helicase. J. Bacteriol. **181:**1728-32.

10. **Chamot, D., and G. W. Owttrim.** 2000. Regulation of cold shock-induced RNA helicase gene expression in the cyanobacterium *Anabaena* sp. strain PCC 7120. J. Bacteriol. **182:**1251-6.

11. **Chevillard, S.** 1993. A method for sequential extraction of RNA and DNA from the same sample, specially designed for a limited supply of biological material. Biotechniques **15:**22-4.

12. **Cohen, M. F., J. C. Meeks, Y. A. Cai, and C. P. Wolk.** 1998. Transposon mutagenesis of heterocyst-forming filamentous cyanobacteria. Methods Enzymol. **297:**3-17.

13. **Curtis, S. E., and J. A. Martin.** 1994. The transcriptional apparatus and the regulation of transcription initiation, p. 613-639. *In* D. A. Bryant (ed.), The Molecular Biology of Cyanobacteria, vol. 1. Kluwer Academic Publishers, The Netherlands.

14. **Doolittle, W. F.** 1979. The cyanobacterial genome, its expression, and the control of that expression, p. 1-102. *In* A. H. Rose and J. F. Wilkinson (ed.), Advances in Microbial Physiology, vol. 20. Academic Press, New York.

15. **Doolittle, W. F.** 1973. Postmaturational cleavage of 23*s* ribosomal ribonucleic acid and its metabolic control in the blue-green alga *Anacystis nidulans*. J. Bacteriol. **113:**1256-63.

16. **Dunlap, J. C.** 1999. Molecular bases for circadian clocks. Cell **96:**271-90.

17. **Duplaa, C., T. Couffinhal, L. Labat, C. Moreau, J. M. Lamaziere, and J. Bonnet.** 1993. Quantitative analysis of polymerase chain reaction products using biotinylated dUTP incorporation. Anal. Biochem. **212:**229-236.

18. **Dvornyk, V., and E. Nevo.** 2003. Genetic polymorphism of cyanobacteria under permanent natural stress: A lesson from the "Evolution Canyons". Res. Microbiol. **154:**79-84.

19. **Dvornyk, V., O. Vinogradova, and E. Nevo.** 2003. Origin and evolution of circadian clock genes in prokaryotes. Proc. Natl. Acad. Sci. U. S. A. **100:**2495-500.

20. **Falquet, L., M. Pagni, P. Bucher, N. Hulo, C. J. Sigrist, K. Hofmann, and A. Bairoch.** 2002. The PROSITE database, its status in 2002. Nucleic Acids Res. **30:**235-8.

21. **Frank, D. N., and N. R. Pace.** 1998. Ribonuclease P: unity and diversity in a tRNA processing ribozyme. Annu. Rev. Biochem. **67:**153-80.

22. **Freeman, W. M., S. J. Walker, and K. E. Vrana.** 1999. Quantitative RT-PCR: pitfalls and potential. Biotechniques **26:**112-22, 124-5.

23. **Gish, W., and D. J. States.** 1993. Identification of protein coding regions by database similarity search. Nature Genet. **3:**266-272.

24. **Golden, S. S., M. Ishiura, C. H. Johnson, and T. Kondo.** 1997. Cyanobacterial circadian rhythms. Annu. Rev. Plant Physiol. Plant Mol. Biol. **48:**327-354.

25. **Goward, T., B. McCune, and D. Meidinger.** 1994. The lichens of British Columbia, illustrated keys: Part I, Foliose and squamulose species. British Columbia Ministry of Forests, Victoria, British Columbia.

26. **Grierson, D., and H. Smith.** 1973. The synthesis and stability of ribosomal RNA in blue-green algae. Eur. J. Biochem. **36:**280-285.

27. **Haas, E. S., and J. W. Brown.** 1998. Evolutionary variation in bacterial RNase P RNAs. Nucleic Acids Res. **26:**4093-4099.

28. **Harmer, S. L., S. Panda, and S. A. Kay.** 2001. Molecular bases of circadian rhythms. Annu. Rev. Cell Dev. Biol. **17:**215-253.

29. **Hayashi, F., H. Suzuki, R. Iwase, T. Uzumaki, A. Miyake, J. R. Shen, K. Imada, Y. Furukawa, K. Yonekura, K. Namba, and M. Ishiura.** 2003. ATP-induced hexameric ring structure of the cyanobacterial circadian clock protein KaiC. Genes Cells **8:**287-96.

30. **Herdman, M., M. Janvier, J. B. Waterbury, R. Rippka, and R. Y. Stanier.** 1979. Deoxyribonucleic acid base composition of cyanobacteria. J. Gen. Microbiol. **111:**63-71.

31. **Honda, D., A. Yokota, and J. Sugiyama.** 1999. Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine *Synechococcus* strains. J. Mol. Evol. **48:**723-39.

32. **Ishiura, M., S. Kutsuna, S. Aoki, H. Iwasaki, C. R. Andersson, A. Tanabe, S. S. Golden, C. H. Johnson, and T. Kondo.** 1998. Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobacteria. Science **281:**1519-1523.

33. **Iwasaki, H., T. Nishiwaki, Y. Kitayama, M. Nakajima, and T. Kondo.** 2002. KaiA-stimulated KaiC phosphorylation in circadian timing loops in cyanobacteria. Proc. Natl. Acad. Sci. U. S. A. **99:**15788-93.

34. **Iwasaki, H., Y. Taniguchi, M. Ishiura, and T. Kondo.** 1999. Physical interactions among circadian clock proteins KaiA, KaiB and KaiC in cyanobacteria. Embo J. **18:**1137-1145.

35. **Iwasaki, H., S. B. Williams, Y. Kitayama, M. Ishiura, S. S. Golden, and T. Kondo.** 2000. A KaiC-interacting sensory histidine kinase, SasA, necessary to sustain robust circadian oscillation in cyanobacteria. Cell **101:**223-233.

36. **Johnson, C. H., S. S. Golden, and T. Kondo.** 1998. Adaptive significance of circadian programs in cyanobacteria. Trends Microbiol. **6:**407-10.

37. **Kageyama, H., T. Kondo, and H. Iwasaki.** 2003. Circadian formation of clock protein complexes by KaiA, KaiB, KaiC, and SasA in cyanobacteria. J. Biol. Chem. **278:**2388-95.

38. **Kaneko, T., Y. Nakamura, C. P. Wolk, T. Kuritz, S. Sasamoto, A. Watanabe, M. Iriguchi, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, M. Kohara, M. Matsumoto, A. Matsuno, A. Muraki, N. Nakazaki, S. Shimpo, M. Sugimoto, M. Takazawa, M. Yamada, M. Yasuda, and S. Tabata.** 2001. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. DNA Res. **8:**205-13; 227-53.

39. **Kaneko, T., S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda, and S. Tabata.** 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions (supplement). DNA Res. **3:**185-209.

40. **Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y.**

References 81

Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, and H. Kikuchi. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. DNA Res. **5**:55-76.

41.   **Kimball, G. R., and M. Goldrich.** Simultaneous extraction of RNA and DNA, Technical Bulletin 161. Ambion.

42.   **Kitayama, Y., H. Iwasaki, T. Nishiwaki, and T. Kondo.** 2003. KaiB functions as an attenuator of KaiC phosphorylation in the cyanobacterial circadian clock system. Embo J. **22**:2127-34.

43.   **Klenk, H. P., R. A. Clayton, J. F. Tomb, O. White, K. E. Nelson, K. A. Ketchum, R. J. Dodson, M. Gwinn, E. K. Hickey, J. D. Peterson, D. L. Richardson, A. R. Kerlavage, D. E. Graham, N. C. Kyrpides, R. D. Fleischmann, J. Quackenbush, N. H. Lee, G. G. Sutton, S. Gill, E. F. Kirkness, B. A. Dougherty, K. McKenney, M. D. Adams, B. Loftus, J. C. Venter, and et al.** 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. Nature **390**:364-70.

44.   **Leipe, D. D., L. Aravind, N. V. Grishin, and E. V. Koonin.** 2000. The bacterial replicative helicase DnaB evolved from a RecA duplication. Genome Res. **10**:5-16.

45.   **Liu, Y., N. F. Tsinoremas, C. H. Johnson, N. V. Lebedeva, S. S. Golden, M. Ishiura, and T. Kondo.** 1995. Circadian orchestration of gene expression in cyanobacteria. Genes Dev. **9**:1469-78.

46.   **Lorne, J.** 1999. Investigating the presence of the *kaiC* gene in various cyanobacteria. TerraGen Discovery Inc.

47.   **Lorne, J., J. Scheffer, A. Lee, M. Painter, and V. P. W. Miao.** 2000. Genes controlling circadian rhythm are widely distributed in cyanobacteria. FEMS Microbiol. Lett. **189**:129-133.

48.   **Ma, J., A. Campbell, and S. Karlin.** 2002. Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. J. Bacteriol. **184**:5733-45.

49.   **Meeks, J. C.** 1998. Symbiosis between nitrogen-fixing cyanobacteria and plants - The establishment of symbiosis causes dramatic morphological and physiological changes in the cyanobacterium. Bioscience **48**:266-276.

50.   **Meeks, J. C., J. Elhai, T. Thiel, M. Potts, F. Larimer, J. Lamerdin, P. Predki, and R. Atlas.** 2001. An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium. Photosynth. Res. **70**:85-106.

51.   **Miao, V. P. W., A. Rabenau, and A. Lee.** 1997. Cultural and molecular characterization of photobionts of *Peltigera membranacea*. Lichenologist **29**:571-586.

52. **Mohamed, A., and C. Jansson.** 1989. Influence of light on accumulation of photosynthesis-specific transcripts in the cyanobacterium *Synechocystis* 6803. Plant Mol.Biol. **13**:693-700.

53. **Mori, T., S. V. Saveliev, Y. Xu, W. F. Stafford, M. M. Cox, R. B. Inman, and C. H. Johnson.** 2002. Circadian clock protein KaiC forms ATP-dependent hexameric rings and binds DNA. Proc. Natl. Acad. Sci. U. S. A. **99**:17203-8.

54. **Nishimura, H., Y. Nakahira, K. Imai, A. Tsuruhara, H. Kondo, H. Hayashi, M. Hirai, H. Saito, and T. Kondo.** 2002. Mutations in KaiA, a clock protein, extend the period of circadian rhythm in the cyanobacterium *Synechococcus elongatus* PCC 7942. Microbiol. **148**:2903-9.

55. **Nishiwaki, T., H. Iwasaki, M. Ishiura, and T. Kondo.** 2000. Nucleotide binding and autophosphorylation of the clock protein KaiC as a circadian timing process of cyanobacteria. Proc. Natl. Acad. Sci. U. S. A. **97**:495-499.

56. **Payne, P. I., and T. A. Dyer.** 1972. Characterization of the ribosomal ribonucleic acids of blue-green algae. Arch. Mikrobiol. **87**:29-40.

57. **Rabenau, A.** 1996. Molecular analysis of colormorphs in the lichen *Peltigera membranacea*: a genetic test using the 16S - 23S rRNA spacer region. Diploma. Technische Universitat Braunschweig, Braunschweig, Germany.

58. **Rippka, R.** 1988. Recognition and identification of cyanobacteria. Methods Enzymol. **167**:28-67.

59. **Rippka, R., J. Deruelles, J. B. Waterbury, M. Herdman, and R. Y. Stanier.** 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J. Gen. Microbiol. **111**:1-61.

60. **Rippka, R., and M. Herdman.** 1992. Pasteur culture collection of cyanobacteria catalogue & taxonomic handbook. Volume 1. Institut Pasteur.

61. **Rozen, S., and H. J. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers, p. 365-386. *In* S. Krawetz and S. Misener (ed.), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, New Jersey. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.

62. **Rudi, K., O. M. Skulberg, and K. S. Jakobsen.** 1998. Evolution of cyanobacteria by exchange of genetic material among phyletically related strains. J. Bacteriol. **180**:3453-61.

63. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. **74**:5463-7.

64.   **Sharma, R. C., and R. T. Schimke.** 1996. Preparation of electrocompetent *E. coli* using salt-free growth medium. Biotechniques **20**:42-4.

65.   **Sigrist, C. J., L. Cerutti, N. Hulo, A. Gattiker, L. Falquet, M. Pagni, A. Bairoch, and P. Bucher.** 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief. Bioinform. **3**:265-74.

66.   **Smith, D. R., L. A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, J. N. Reeve, and et al.** 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. J. Bacteriol. **179**:7135-55.

67.   **Taniguchi, Y., A. Yamaguchi, A. Hijikata, H. Iwasaki, K. Kamagata, M. Ishiura, M. Go, and T. Kondo.** 2001. Two KaiA-binding domains of cyanobacterial circadian clock protein KaiC. FEBS Lett. **496**:86-90.

68.   **Thellin, O., W. Zorzi, B. Lakaye, B. De Borman, B. Coumans, G. Hennen, T. Grisar, A. Igout, and E. Heinen.** 1999. Housekeeping genes as internal standards: use and limits. J. Biotechnol. **75**:291-5.

69.   **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. Clustal W - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673-4680.

70.   **Tous, C., M. A. Vega-Palas, and A. Vioque.** 2001. Conditional expression of RNase P in the cyanobacterium *Synechocystis* sp. PCC6803 allows detection of precursor RNAs. Insight in the in vivo maturation pathway of transfer and other stable RNAs. J. Biol. Chem. **276**:29059-66.

71.   **Turner, S., K. M. Pryer, V. P. W. Miao, and J. D. Palmer.** 1999. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. J. Eukaryot. Microbiol. **46**:327-338.

72.   **Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman.** 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. **3**:1-11.

73.   **Vioque, A.** 1992. Analysis of the gene encoding the RNA subunit of ribonuclease P from cyanobacteria. Nucleic Acids Res. **20**:6331-6337.

74.   **Vioque, A.** 1997. The RNase P RNA from cyanobacteria: short tandemly repeated repetitive (STRR) sequences are present within the RNase P RNA gene in heterocyst-forming cyanobacteria. Nucleic Acids Res. **25**:3471-3477.

75. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. Embo J. **1:**945-51.

76. **Williams, S. B., I. Vakonakis, S. S. Golden, and A. C. LiWang.** 2002. Structure and function from the circadian clock protein KaiA of *Synechococcus elongatus*: a potential clock input mechanism. Proc. Natl. Acad. Sci. U. S. A. **99:**15357-62.

77. **Xu, Y., T. Mori, and C. H. Johnson.** 2000. Circadian clock-protein expression in cyanobacteria: rhythms and phase setting. Embo J. **19:**3349-3357.

78. **Xu, Y., T. Mori, and C. H. Johnson.** 2003. Cyanobacterial circadian clockwork: roles of KaiA, KaiB and the kaiBC promoter in regulating KaiC. Embo J. **22:**2117-26.

79. **Zheng, H., W. Yan, J. Toppari, and P. Harkonen.** 2000. Improved nonradioactive RT-PCR method for relative quantification of mRNA. Biotechniques **28:**832-834.

Appendix A:
Nucleotide alignment of the *Nostoc* PCC 9709 *kaiC* PCR product against the region of *S. elongatus kaiC* bracketed by the O-488/O-489 primer pair.
Invariant bases are shaded in grey. The PCC 7942 sequence shown here corresponds to bases 93 through 687 of the *kaiC* gene (GenBank Accession AB010691). The GC content is 41.2 % for the PCC 9709 sequence and 49.8 % for the PCC 7942 sequence.


```
PCC 9709  TAGTCATGGTGGTTTACCAATTGGTAGAACTACCTTGATCAGTGGTACCTCCGGCACAGG   60
PCC 7942  TAGTCATGGCGGTCTTCCAATCGGGCGATCGACCCTCGTTAGTGGTACTTCAGGAACCGG   60

PCC 9709  CAAAACTTTATTCTCTCTTCAGTTTCTCTATAACGGTATCACCTACTTTGATGAAGCAGG   120
PCC 7942  CAAGACCCTTTTTTTCTATTCAATTTCTCTATAACGGTATTATCGAGTTTGATGAGCCTGG   120

PCC 9709  AGTATTTGTTACCTTTGAAGAATCACCCAGTGATATTATTAAAAATGCCCATGTTTTTGG   180
PCC 7942  GGTTTTTCGTTACTTTCGAAGAAACCCCGCAAGATATCATTAAAAACGCCCGTAGTTTTGG   180

PCC 9709  TTGGAACTTGCCACGCCTAATTGAAGAAGGCAAGTTATTTATTCTTGATGCATCTCCCGA   240
PCC 7942  CTGGGATTTAGCCAAGCTGGTCGATGAGGGCAAACTATTTATTCTTGATGCTTCACCCGA   240

PCC 9709  TCCAGAAGGTCAAGATATCGTTGGTAATTTTGACCTTTCTGCACTCATTGAACGCTTGCA   300
PCC 7942  TCCAGAAGGTCAAGAGGTTGTTGGCGGCTTCGATCTCTCTGCTCTGATTGAGCGGATTAA   300

PCC 9709  ATATGCCATCCGTAAATACAAAGCTAAACGAGTTTCAATCGACTCAATAACAGCAGTATT   360
PCC 7942  TTATGCAATTCAAAAGTATCGAGCGCGGCGGGTTTCAATTGACTCGGTCACGTCCGTTTT   360

PCC 9709  TCAGCAGTATGAAGCGATGGGAGTAGTGCCGACGTGAGATTTTTCGCCTGGTAGCACGTCT   420
PCC 7942  CCAGCAATATGATGCCTCTTCTGTGGTTCGCCGCGAACTCTTTCGGTTGGTAGCTCGCCT   420

PCC 9709  GAAATTATTGAATGTCACCACTGTAATTACCACTGAACGTGGTGAAGAATATGGGCCTGT   480
PCC 7942  AAAACAAAATTGGGGCAACTACGGTCATGACCACCGAGCGTATCGAGGAATATGGCCCGAT   480

PCC 9709  TGCCTCTTTCGGAGTAGAAGAATTTGTTTCTGATAATGTAGTAATTGTTCGCAACGTTTT   540
PCC 7942  CGCTCGTTACGGTGTTGAGGAATTTGTCTCCGATAACGTCGTGATTCTCCGCAACGTTTT   540

PCC 9709  AGAAGGAGAACGTCGCCGTCGCACAATTGAAATTCTCAAGTTGCGCGGGACAACT       595
PCC 7942  GGAAGGGGAGCGCCGTCGCCGCACCCTCGAAATCCTCAAGCTACGTGGCACCAGC       595
```

Appendix B: Map analysis of clones recovered from the *Nostoc* genomic library

The probe region is denoted by the striped area. Vector regions are diagrammed as grey boxes. PCR primer orientations are indicated with arrows. The mapped positions of library clones are shown by thick black lines (the dotted line indicates chimeric sequence). The position of the Kpn I site on inserts is indicated by the short vertical lines.
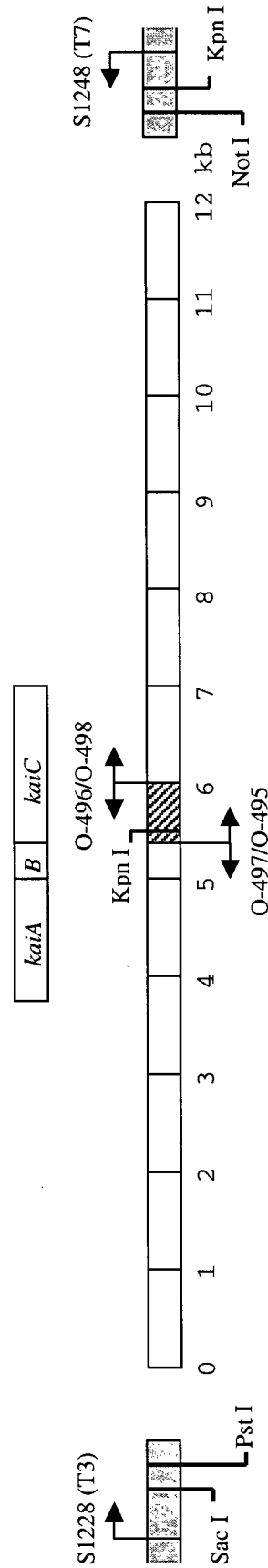

Library Clone	Vector Orientation (5' – 3')
pkai2A-1	T7 – T3
pkai2B-3, pkai2C-1	T3 – T7
pkai2D-5	T3 – T7
pkai2E-1	T3 – T7
pkai2F-1	not determined
pkai2G-1, pkai2N-1	T7 – T3
pkai2H-1	T7 – T3
pkai2I-1, pkai2M-1	T7 – T3
pkai2J-1	T7 – T3
pkai2K-1	T3 – T7
pkai2L-2	T3 – T7

Appendix C:
Field collection site of *P. membranacea* specimens.
The map was published by the Government of British Columbia, Ministry of Water, Land and Air Protection. The collection site is denoted by star symbol.
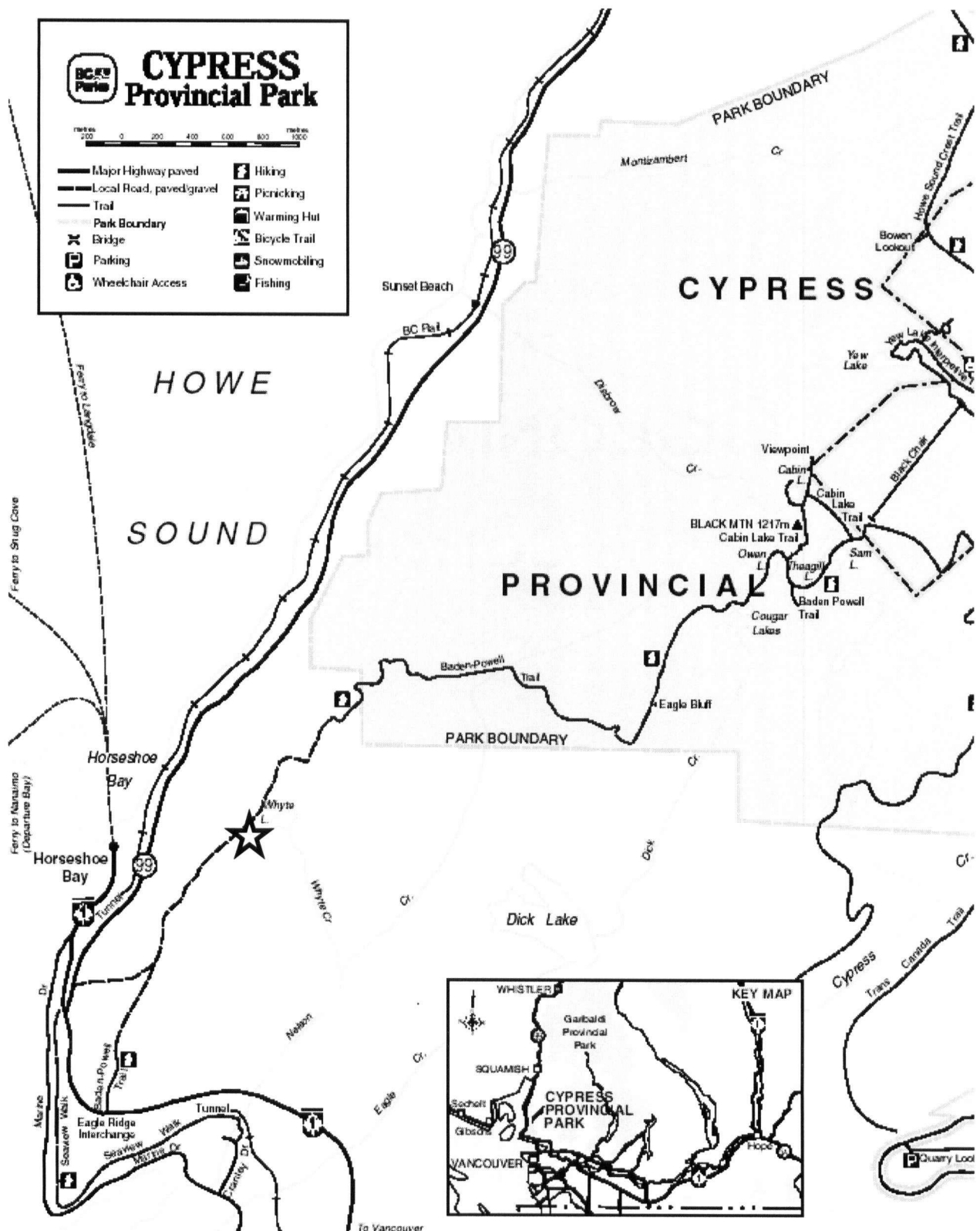
Appendix D:
Genetic complementation of *S. elongatus* with *Nostoc kaiABC* disrupts wildtype circadian rhythms.

S. elongatus strain AMC 541 carries a luciferase reporter gene under the control of the *kaiBC* promoter for monitoring circadian rhythms of expression. A) Wildtype phenotype of host strain, B-D) Individual transformants carrying *Nostoc kaiABC* on the chromosome. After three days, the luminescence trace dampens as cultures become overgrown in 96-well culture plates.