Photosynth Res
DOI 10.1007/s11120-006-9122-4

REGULAR PAPER

Complete nucleotide sequence of the freshwater unicellular cyanobacterium *Synechococcus elongatus* PCC 6301 chromosome: gene content and organization

Chieko Sugita · Koretsugu Ogata · Masamitsu Shikata · Hiroyuki Jikuya · Jun Takano · Miho Furumichi · Minoru Kanehisa · Tatsuo Omata · Masahiro Sugiura · Mamoru Sugita

Received: 30 August 2006 / Accepted: 6 December 2006

Abstract The entire genome of the unicellular cyanobacterium *Synechococcus elongatus* PCC 6301 (formerly *Anacystis nidulans* Berkeley strain 6301) was sequenced. The genome consisted of a circular chromosome 2,696,255 bp long. A total of 2,525 potential protein-coding genes, two sets of rRNA genes, 45 tRNA genes representing 42 tRNA species, and several genes for small stable RNAs were assigned to the chromosome by similarity searches and computer predictions. The translated products of 56% of the potential protein-coding genes showed sequence similarities to experimentally identified and predicted proteins of known function, and the products of 35% of the genes showed sequence similarities to the translated products of hypothetical genes. The remaining 9% of genes lacked significant similarities to genes for predicted proteins in the public DNA databases. Some 139 genes coding for photosynthesis-related components were identified. Thirty-seven genes for two-component signal transduction systems were also identified. This is the smallest number of such genes identified in cyanobacteria, except for marine cyanobacteria, suggesting that only simple signal transduction systems are found in this strain. The gene arrangement and nucleotide sequence of *Synechococcus elongatus* PCC 6301 were nearly identical to those of a closely related strain *Synechococcus elongatus* PCC 7942, except for the presence of a 188.6 kb inversion. The sequences as well as the gene information shown in this paper are available in the Web database, CYORF (http://www.cyano.genome.jp/).

Keywords *Anacystis nidulans* 6301 · Cyanobacterial genome · Large inversion · Photosynthesis-related gene · *Synechococcus elongatus* PCC 6301 · Two-component signal transduction system · Transcription factor

Abbreviations

PSI	Photosystem I
PSII	Photosystem II
HIP1	Highly iterated palindrome
IS	Insertion sequence
ORF	Open reading frame

C. Sugita · M. Sugita (✉)
Center for Gene Research, Nagoya University, Chikusa, Nagoya 464-8602, Japan
e-mail: sugita@gene.nagoya-u.ac.jp

K. Ogata · M. Shikata · J. Takano
Life Science Business Unit, Shimadzu Co., Nakagyo-ku, Kyoto 604-8511, Japan

H. Jikuya
Bio-Architecture Center, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

M. Furumichi · M. Kanehisa
Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

T. Omata
Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8601, Japan

M. Sugiura
Graduate School of Natural Sciences, Nagoya City University, Nagoya 467-8501, Japan

Introduction

Cyanobacteria are bacteria capable of oxygenic photosynthesis and comprise over 1,600 species with various morphologies and species-specific characteristics such as cell movement, cell differentiation, and nitrogen fixation (Bryant 1994). The entire genome sequence of a unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803 was first described in 1996 (Kaneko et al. 1996b). To date (9 November 2006), 19 cyanobacterial genomes have been sequenced. These genomes are those of the filamentous nitrogen-fixing cyanobacterium *Anabaena* (*Nostoc*) sp. strain PCC 7120 (Kaneko et al. 2001), the thermophilic strain *Thermosynechococcus elongatus* BP-1 (Nakamura et al. 2002), the thylakoid-free strain *Gloeobacter violaceus* PCC 7421 (Nakamura et al. 2003), the marine cyanobacterium *Synechococcus* sp. strain WH8102 (Palenik et al. 2003), the *Prochlorococcus marinus* strains SS120 (Dufresne et al. 2003), MED4 (Rocap et al. 2003), MIT 9313 (Rocap et al. 2003), *Synechococcus* sp. CC9311 (Palenik et al. 2006), and others (NCBI web site; http://www.ncbi.nlm.nih.gov/genomes/static/eub_g.html). *Synechococcus elongatus* PCC 6301 is a unicellular rod-shaped cyanobacterium and inhabits freshwater. *S. elongatus* PCC 6301 is an obligate photoautotrophic organism, pigmented with chlorophylls, carotenoids and phycocyanobilin, and has long been used as a model organism for the study of photosynthesis (Bryant 1994). *S. elongatus* PCC 6301 was first isolated as a unicellular strain and made axenic by Allen (1952). Strain PCC 6301 was the first accession of the Pasteur Culture Collection. In 1975, a temporal genetic map for 19 markers was constructed (Delaney and Carr 1975) and gene analysis using modern technologies started in 1981 (Tomioka et al. 1981). We previously constructed a physical map of the *S. elongatus* PCC 6301 genome using two restriction enzymes, *Pme*I and *Swa*I, and the I-*Ceu*I endonuclease and estimated the genome size to be 2.7 Mb (Kaneko et al. 1996a).

To further acquire our knowledge of *S. elongatus* PCC 6301 and to understand the complex association between the genetics, physiology and biochemistry of photosynthesis, we determined the nucleotide sequence of the entire genome and analyzed the structure of all gene components of this organism. In this communication, we describe the overall structure of the genome and compare this genome with that of a closely related strain, *S. elongatus* PCC 7942.

Materials and methods

Cyanobacterial strain

The *S. elongatus* PCC 6301 used in this study was formerly designated *Anacystis nidulans* Berkeley strain no. 6301. This strain was originally sourced in 1972 from Dr. Riyo Kunisawa of the University of California, Berkeley, who sent the strain to Prof. Dr. Seiichi Hino of Hiroshima University, Japan (Tomioka 1983).

Cloning of DNA fragments and DNA sequencing

A *S. elongatus* PCC 6301 genomic library was constructed using 10–20 kb DNA fragments partially digested with *Sau*3AI and λDash-*Bam*HI digested arms (Stratagene, La Jolla). The λ clones were screened by plaque hybridization using *Swa*I or *Pme*I restriction fragments separated by pulsed-field gel electrophoresis (Kaneko et al. 1996a). The λDNAs were digested with *Eco*RI, subjected to Southern blot analysis, and overlap with various clones confirmed. A total number of 519 λ clones was ordered and a minimum set of 198 clones was selected for sequencing analysis. The sum of the lengths of insert DNAs was 3,523 kb (Sugita et al. 2001). The insert DNA of each λ clone was amplified by long and accurate polymerase chain reaction (LA PCR) using LA *Taq* DNA polymerase (TaKaRa, Japan) and inserts were sheared to short DNA fragments. The resulting DNA fragments were subcloned into the *Sma*I site of pUC18 and shotgun sequenced using the Dye-Terminator Cycle Sequencing kit (Amersham Biosciences) using a Shimadzu multi-capillary DNA sequencer (RISA-384). A total of 54,618 sequence files corresponding to about 10× genome-equivalents were accumulated and assembled using the Phrap program (Phil Green, University of Washington, Seattle, USA). The final gaps in the sequences were filled by primer walking. A lower threshold of acceptability for the generation of consensus sequences was set at a Phred score of 20 for each base.

Gene assignment and annotation

Coding regions were assigned by a combination of computer prediction and similarity search. Briefly, prediction of protein-coding regions was effected with the Glimmer 2.02 program. Prior to prediction, a matrix for the *S. elongatus* PCC 6301 genome was generated by training with a dataset of 665 open

reading frames (ORFs) that showed a high degree of sequence similarity at the amino acid level to experimentally identified and predicted proteins of known function. All of the predicted protein-encoding regions equal to or longer than 90 bp were translated into amino acid sequences and these were then subjected to similarity search against the non-redundant protein database (nr-database) with the BLASTP program (Altschul et al. 1977). In parallel, the entire genomic sequence was compared with those in the nr-protein database using the BLASTX program to identify genes that had escaped from prediction and/or those smaller than 90 bp, especially in the predicted intergenic regions. For predicted genes that did not show sequence similarities to known genes, only those equal to or longer than 300 bp were considered as authentic gene candidates.

Presumed functions of predicted genes were assigned on the basis of sequence similarities of their deduced products to those of genes of known function. For genes that encoded proteins of 100 amino acid residues or more, a BLAST *E*-value of e^{-30} was considered significant. A higher *E*–value was considered significant for genes encoding smaller proteins. The tRNA-coding regions were predicted by use of the tRNA scan-SE 1.23 program (Lowe and Eddy 1997) in combination with similarity searches.

Results and discussion

Sequence determination of the entire genome

The genome of *S. elongatus* PCC 6301 is a circular chromosome of 2,696,255 bp and the average GC content is 55.5%. Nucleotides are numbered from the first A nucleotide of ATTTAAAT, at one of twenty recognition sites of *Swa*I (Fig. 1). The inner circle of Fig. 1 shows the average GC percentage of every 2 kb segment of the entire genome. No obvious uneven GC distribution was observed except for three deep spikes, L1, L2 and L3, at approximate coordinates of 875 kb, 1,480 kb, and 1,566 kb, respectively. The L1 region had 33.6% GC content and contained the *syc0782_c* gene encoding a putative bacteriophage protein (950 amino acids) and the *syc0783_c* gene encoding a putative phage tail tube protein FII (167 amino acids). The L2 region (35.3% GC content) contained the *syc1369_d* gene encoding a phage integrase-like protein (435 amino acids). In the L3 region, the *syc1448_d* encodes a tetratricopeptide repeat motifs-containing polypeptide, similar to the plant gibberellin signal transduction protein SPINDLY (At3g11540). This ORF was 1,920 bp in size and contained 34.7% GC. These sequences may have originated by lateral transfer from other organisms. GC skew analysis was performed to locate a probable origin and terminator of DNA replication but no apparent shift of skew was detected.

The highly iterated palindrome (HIP1) is an 8 bp palindromic sequence, GCGATCGC, first reported in the genomes of *S. elongatus* PCC 6301 and other cyanobacterial species (Gupta et al. 1993). A survey of the entire genome of *S. elongatus* PCC 6301 showed that 7358 copies of the HIP1 sequence were present. This sequence may be characteristic of cyanobacterial genomes (Table 1). In contrast, only 222 copies of the HIP1 sequence are found in the *Escherichia coli* genome and 51 copies occur in the *Bacillus subtilis* genome.

Assignment of protein-encoding and RNA-encoding genes

Potential protein-coding regions were assigned using a combination of computer prediction by the Glimmer program and similarity search. The total number of potential protein-coding genes finally assigned to the genome was 2,525, and the name of each gene is listed in the Cyanobacteria Gene Annotation Database, CYORF (http://www.cyano.genome.jp/). The putative protein-coding genes, starting with either ATG, GTG, or TTG codons, are denoted by a number with three letters representing the species name (syc). The numbers run from 1 to 2,525 from position 1 (the *Swa*I site) in a clockwise direction, and the direction of transcription (_c or _d) is indicated. Use of the term _c indicates that transcription is counterclockwise. For example, *syc1093_c*, encoding the photosystem (PS) II reaction center D1 protein, is transcribed in a counterclockwise direction on the genome (Fig. 1), while the *syc2044_d* gene encoding the PSI reaction center subunit Ia is oriented in the opposite direction to *syc1093_c*. Among 2,525 potential protein-coding genes identified in the genome, 1,422 (56%) were homologous to genes of known function, 871 (35%) showed similarities to conserved hypothetical genes, and the remaining 232 (9%) showed no significant similarities to any reported genes.

Two copies of an rRNA gene cluster (*rrnA* and *rrnB*) (Tomioka and Sugiura 1983; Kumano et al. 1983) were assigned on the genome, ordered as 16*S*–23*S*–5*S*, at 1,050 kb and 1,650 kb (Fig. 1). The spacing between them (measured from the 5′ ends of the clusters) is 605,653 bp. A total of 45 tRNA genes representing 42 tRNA species was identified on the genome (Fig. 1). These genes are dispersed and are likely to be transcribed as single units, except for the

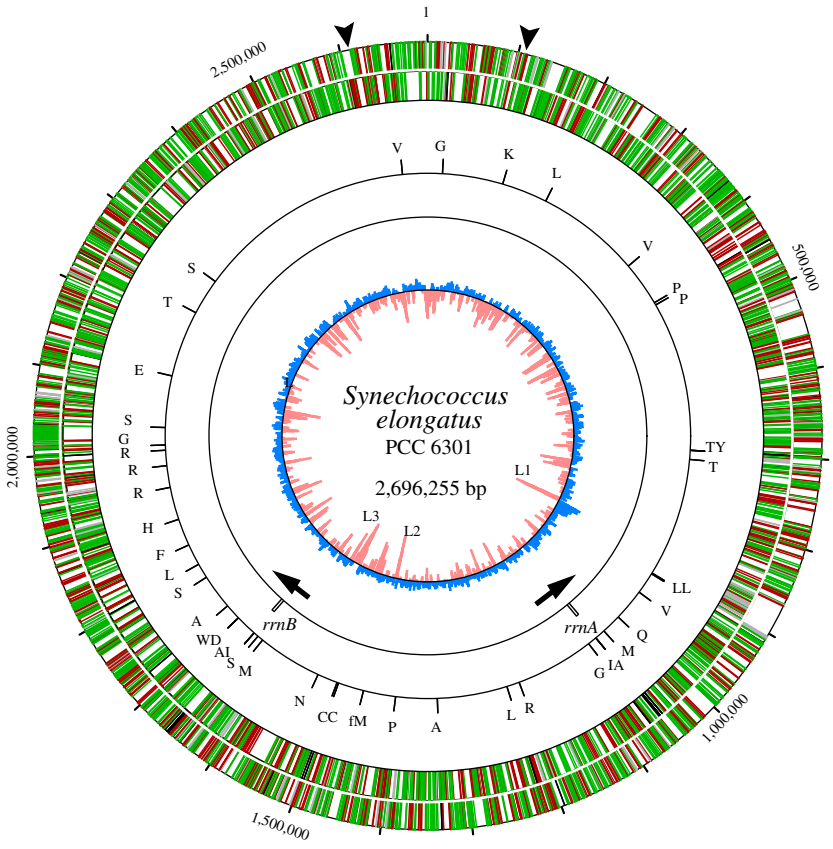

Fig. 1 Circular representation of the chromosome of *Synechococcus elongatus* PCC 6301. The scale indicates the location in bp starting from the Swa*I* recognition site (Kaneko et al. 1996a). The bars in the outermost and the second circles show positions of putative protein-coding genes in clockwise and counterclockwise directions, respectively. The genes (56% of total) whose functions could be deduced by sequence similarity to genes of known function are depicted in green and those (35%) that showed sequence similarity to conserved hypothetical genes registered in public databases are in red. The remaining genes (9%) that showed no significant similarities to any registered genes are in gray. The bars in the third circle indicate the positions of predicted tRNA genes and those in the fourth circle the positions of genes for rRNAs (*rrnA* and *rrnB*). The tRNA genes are indicated by single amino acid codes identifying the specificities of the tRNAs. The innermost circle with a scale shows the average G + C percentage calculated using window size of 2 kb. Pink spikes towards the inside of the innermost circle indicate lower GC percentages and blue spikes outside of the innermost circle indicate higher GC percentages than the average value of 55.5%. The arrowheads indicate the endpoints of the large inversion in the *S. elongatus* PCC 6301 genome, relative to that of *S. elongatus* PCC 7942

tandemly arrayed *trnY*-GUA–*trnT*-GGU (at 697 kb), *trnW*-CCA-*trnD*-GUC (at 1,693 kb) and the spacer tRNA genes *trnA*-UGC and *trnI*-GAU in the rRNA gene cluster. The *trnL*-UAA gene contained a group I intron of 240 bp (Sugita et al. 1995) as reported in several cyanobacterial strains (Paquin et al. 1997). No group II intron-containing gene was found in the *S. elongatus* PCC 6301 genome although group II introns are generally found in other cyanobacteria, for example *Synechocystis* PCC 6803, *Nostoc* PCC 7120, *T. elongatus* BP-1, and *G. violaceus* PCC 7421. Turning to small RNA-coding genes, potential genes showing sequence similarity to SRP (signal recognition particle) RNA (*ffs*, Kaneko et al. 1996b), tmRNA (*ssrA*, Watanabe et al. 1998), 6Sa RNA (*ssaA*, Watanabe et al. 1997), and RNase P subunit B (*rnpB*, Banta et al. 1992) were identified in the genome. Recently, we have identified a gene (*sncA*) between *syc2263_d* (*guaB*) and *syc2264_d* (*trxA*), encoding a novel small RNA (Nakamura et al. unpublished). Our preliminary northern blot analysis using intergenic spacer DNAs revealed the presence of numerous sequences for unidentified small stable RNAs in the *S. elongatus* PCC 6301 genome.

Genes involved in photosynthesis

One hundred thirty nine genes for components related to photosynthesis are listed in Table 2. Complete sets of PSI and PSII genes were identified in the genome. As in

Table 1 General features of the genomes of freshwater cyanobacterial species

	S. elongatus PCC 6301	*Thermosynechococcus elongatus* BP-1	*Synechocystis* PCC 6803	*Nostoc* PCC 7120	*Gloeobacter violaceus* PCC 7421
Genome size (bp)	2,696,255	2,593,857	3,573,470	6,413,771	4,659,019
GC content (%)	55.5	53.9	47.7	41.3	62.0
Highly iterated palindrome (HIP1)	7,358	3,681	3,160	5,916	318
Protein gene	2,525	2,475	3,264	6,129	4,430
rRNA operon	2	1	2	4	1
tRNA gene	45	44	42	48	45
Two-component systems[a]					
Sensory kinase	13	17	26	71	27
Response regulator	21	27	38	71	37
Hybrid type	3	6	17	53	12
Serine/threonine protein kinase	5	11	7	39	15
Transcription factor[a]	36	27	46	99	110
Sigma factor[a]	10	7	9	8	14
Transposase[a]	1	82	112	145	74

[a] The data for *Nostoc* PCC 7120 represent the sum of the chromosomal and plasmid-encoded genes

Synechocystis PCC 6803, there are three copies of *psbA* and two copies of *psbD*, *psaK*, and *psb28* in *S. elongatus* PCC 6301. The other PSI and PSII genes are present as single copies. Single genes for the subunits of the cytochrome b_6/f complex (*petA*, *B*, *C*, *D*, *G*, *M*, *N*) are found in *S. elongatus* PCC 6301. It is noteworthy that *S. elongatus* PCC 6301 contains but a single *petC* gene for the Rieske-type FeS center protein of the cytochrome b_6/f complex, in contrast to the situation in *Synechocystis* PCC 6803 and *Nostoc* PCC 7120, both of which have multiple *petC* genes. It may also be noted that three copies of the *petJ* gene (*syc0089_d*, *syc1274_d*, *syc1568_d*) encoding cytochrome *c*553, a mobile electron carrier, are present in *S. elongatus* PCC 6301 while *Synechocystis* PCC 6803 has but one copy. Three copies of the *petF* gene (*syc1175_c*, *syc1529_c*, *syc2484_c*) encoding ferredoxin are found in *S. elongatus* PCC 6301 while four copies of *petF* are found in *Synechocystis* PCC 6803. A complete set of genes for carbon fixation enzymes was found in *S. elongatus* PCC 6301, although a gene for RubisCO activase was absent, as is the case in *Synechocystis* PCC 6803. RubisCO activase gene is present in *Nostoc* PCC 7120 (Kaneko et al. 2001), *Nostoc punctiforme* ATCC 29133 (Meeks et al. 2001), *G. violaceus* PCC 7421 (Nakamura et al. 2003), *Anabaena variabilis* ATCC 29413 (NC_007413), *Trichodesmium erythraeum* IMS101 (NC_008312), and *Synechococcus* sp. JA-3-3Ab (NC_007775). Several copies of *ndhD* and *ndhF*, encoding subunits of NADH dehydrogenase, are present in both *S. elongatus* PCC 6301 and *Synechocystis* PCC 6803. Turning to phycobilisome components, *S. elongatus* PCC 6301 contains complete sets of *cpcA–cpcG* genes (for phycocyanin synthesis) and *apcA–apcF* genes (for allophycocyanin production). The *cpcA*, *cpcB*, and *cpcC* genes are duplicated in *S. elongatus* PCC 6301, but only one *cpcG* gene is present. Several copies of the *cpc* genes are present in *Synechocystis* PCC 6803, *Nostoc* PCC 7120, and *T. elongatus* BP-1. Genes required for phycoerythrocyanin are absent in *S. elongatus* PCC 6301 as is the case in other cyanobacteria, except for *Nostoc* PCC 7120 and *A. variabilis* ATCC 29413.

Genes for signal transduction

A large number of genes encoding elements of two-component signal transduction systems were first assigned in *Synechocystis* PCC 6803 (Mizuno et al. 1996), and thereafter in *Nostoc* PCC 7120, *T. elongatus* BP-1 and *G. violaceus* PCC 7421 (Table 1). This indicates that cyanobacteria have devised sophisticated signaling systems facilitating adaptive responses to their environments. In *S. elongatus* PCC 6301, 13 and 21 potential genes for sensory histidine kinases and response regulators, respectively, were identified in the genome (Table 3). Three hybrid sensory kinase genes were also assigned. The gene total of 37 is the smallest number assigned in freshwater cyanobacteria. The marine cyanobacterial *Prochlorococcus marinus* SS120 has 11 such genes (Dufresne et al. 2003). Of the 37 genes, a half of genes were described previously in *S. elongatus* PCC 6301 and related species. They include highly conserved genes such as those for phosphate

Table 2 List of photosynthesis-related genes

Gene name	ORF ID	Products or definition
Photosystem I (16)		
psaA	*syc2044_d*	Reaction center subunit Ia
psaB	*syc2045_d*	Reaction center subunit Ib
psaC	*syc0986_d*	Reaction center subunit VII
psaD	*syc0543_c*	Reaction center subunit II
psaE	*syc0231_c*	Reaction center subunit IV
psaF	*syc0300_d*	Reaction center subunit III
psaI	*syc1760_d*	Reaction center subunit VIII
psaJ	*syc0301_d*	Reaction center subunit IX
psaK1	*syc0622_d*	Reaction center subunit X
psaK2	*syc1109_d*	Reaction center subunit X
psaL	*syc1761_d*	Reaction center subunit XI
psaM	*syc2182_c*	Reaction center subunit XII
btpA	*syc1593_c*	Photosystem I biogenesis protein btpa
ycf3	*syc1781_c*	Photosystem I assembly related protein
ycf4	*syc0874_d*	Photosystem I assembly related protein
ycf37	*syc0278_d*	Photosystem I assembly related protein
Photosystem II (32)		
psbAI	*syc1093_c*	Reaction center D1 protein
psbAII	*syc0166_d*	Reaction center D1 protein
psbAIII	*syc0647_c*	Reaction center D1 protein
psbB	*syc0833_d*	Core light harvesting protein CP47
psbC	*syc0872_c*	Reaction center CP43 protein
psbDI	*syc0873_c*	Reaction center D2 protein
psbDII	*syc2448_d*	Reaction center D2 protein
psbE	*syc0373_d*	Cytochrome b559 alpha chain
psbF	*syc0374_d*	Cytochrome b559 beta chain
psbH	*syc1288_c*	PsbH protein
psbI	*syc2386_c*	PsbI protein
psbJ	*syc0376_d*	PsbJ protein
psbK	*syc1062_c*	PsbK protein
psbL	*syc0375_d*	PsbL protein
psbM	*syc0831_c*	PsbM protein
psbN	*syc1289_d*	PsbN protein
psbO	*syc1218_c*	Manganese-stabilizing polypeptide
psbP	*syc0509_c*	Oxygen-evolving complex 23 kD protein PsbP
psbQ	*syc2412_d*	Extrinsic protein PsbQ
psbTc	*syc0834_d*	PsbTc protein
psbU	*syc2213_d*	12 kD extrinsic protein
psbV	*syc2085_d*	Cytochrome c550
psb27	*syc1170_c*	11 kD extrinsic protein
psb28	*syc2411_c*	13 kD extrinsic protein
psb28	*syc1626_d*	13 kD extrinsic protein
psb29	*syc0013_d*	PSII component
psbX	*syc2079_d*	PsbX protein
psbY	*syc2133_c*	PsbY protein
psbZ	*syc1853_d*	Plant Ycf9 homolog
isiA	*syc0002_c*	Iron-stress inducing chlorophyll-binding protein (CP43')
ycf39	*syc1681_d*	Chaperon-like protein for quinone binding
ycf48	*syc0372_d*	Photosystem II stability/assembly factor, Ycf48-like protein
Cytochrome b_6/f complex (7)		
petA	*syc0319_d*	Apocytochrome f component of cytochrome b_6/f complex
petB	*syc1771_c*	Cytochrome b_6
petC	*syc0318_d*	Cytochrome b_6/f complex iron-sulfur subunit
petD	*syc1770_c*	Cytochrome b_6/f complex subunit 4
petG	*syc2465_d*	Cytochrome b_6/f complex subunit V
petM	*syc1680_d*	Cytochrome b_6/f complex subunit PetM
petN	*syc1044_d*	Cytochrome b_6/f complex subunit VIII
Soluble electron carriers (12)		
petE	*syc0461_c*	Plastocyanin
petF	*syc1175_c*	Ferredoxin, petf-like protein

Table 2 continued

Gene name	ORF ID	Products or definition
petF	*syc1529_c*	Ferredoxin, petf-like protein
petF	*syc2484_c*	Ferredoxin, petf protein
petH	*syc0566_c*	Ferredoxin-NADP oxidoreductase
petJ	*syc1274_d*	Cytochrome c553
petJ	*syc1568_d*	Cytochrome c553
petJ	*syc0089_d*	Cytochrome c553
petL	*syc1391_c*	PetL-like
isiB	*syc0001_c*	Flavodoxin
ftrC	*syc2357_c*	Ferredoxin-thioredoxin reductase catalytic chain
ftrV	*syc0418_c*	Ferredoxin-thioredoxin reductase variable subunit
ATP synthase (10)		
atpA	*syc1177_c*	ATP synthase alpha chain of F_1
atpB	*syc1787_c*	ATP synthase beta chain of F_1
atpC	*syc1176_c*	ATP synthase gamma chain of F_1
atpD	*syc1178_c*	ATP synthase delta chain of F_1
atpE	*syc1786_c*	ATP synthase epsilon chain of F_1
atpF	*syc1179_c*	ATP synthase b chain of Fo
atpG	*syc1180_c*	ATP synthase b′ chain of Fo
atpH	*syc1181_c*	ATP synthase lipid-binding c chain of Fo
atpI	*syc1182_c*	ATP synthase a chain of Fo
atp1	*syc1183_c*	ATP synthase protein 1 subunit
NADH dehydrogenase (21)		
ndhA	*syc0210_c*	Subunit 1
ndhB	*syc0140_d*	Subunit 2
ndhC	*syc0370_c*	Subunit 3
ndhD1	*syc2120_d*	Subunit 4
ndhD2	*syc0117_d*	Subunit 4
ndhD3	*syc2001_c*	Subunit 4
ndhD4	*syc0915_c*	Subunit 4
ndhD5	*syc2459_c*	Subunit 4
ndhE	*syc0207_c*	Subunit 4 L
ndhF1	*syc2119_d*	Subunit 5
ndhF3	*syc2002_c*	Subunit 5
ndhF4	*syc1204_d*	Subunit 5
ndhG	*syc0208_c*	Subunit 6
ndhH	*syc2348_c*	Subunit 7
ndhI	*syc0209_c*	Subunit NdhI
ndhJ	*syc0368_c*	Subunit I
ndhK	*syc0369_c*	Subunit NdhK
ndhL	*syc1104_d*	Subunit NdhL
ndbA	*syc1403_c*	Type 2 NADH dehydrogenase NdbA
ndbB	*syc1598_c*	Type 2 NADH dehydrogenase NdbB
ndbC	*syc1313_d*	Type 2 NADH dehydrogenase NdbC
Carbon fixation (23)		
ccmA	*syc1811_d*	Probable carboxysome formation protein
ccmK1	*syc0135_c*	Carboxysome shell protein
ccmK3	*syc1228_c*	Carboxysome shell protein
ccmK4	*syc1227_c*	Carboxysome shell protein
ccmL	*syc0134_c*	Putative carboxysome assembly protein
ccmM	*syc0133_c*	Putative carboxysome structural protein (related to rbcs)
ccmN	*syc0132_c*	Putative carboxysome assembly protein
ccmO	*syc0131_c*	Putative carboxysome assembly protein (related to ccmk)
icfA	*syc0110_c*	Carboxysome-localized carbonic anhydrase
rbcL	*syc0130_c*	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit
rbcS	*syc0129_c*	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit
rbcX	*syc2521_d*	Rubisco chaperonin
pgk, cbbK	*syc0433_c*	Phosphoglycerate kinase
gap2, cbbG	*syc2349_c*	NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase
tpiA, cbbJ	*syc0290_d*	Triosephosphate isomerase
fbaA, cbbA	*syc0113_c*	Fructose-1,6-bisphosphate/sedoheptulose-1,7-bisphosphate aldolase
fbpI, glpX, cbbF	*syc1015_d*	Fructose-1,6-bisphosphatase/sedoheptulose-1,7-bisphosphatase

Table 2 continued

Gene name	ORF ID	Products or definition
tktA, cbbT	*syc0983_c*	Transketolase
ppe, cbbE	*syc0920_c*	Pentose-5-phosphate-3-epimerase
rpiA, cbbI	*syc0939_d*	Ribose-5-phosphate isomerase
prk, cbbP	*syc0567_d*	Phosphoribulokinase
cp12	*syc1152_c*	CP12 protein
cp12	*syc1261_d*	Putative CP12 protein
Phycobilisome (18)		
apcA	*syc1186_d*	Allophycocyanin alpha chain
apcB	*syc1187_d*	Allophycocyanin beta subunit
apcC	*syc1188_d*	Lc 7.8 apoprotein (core components of the phycobilisomes)
apcD	*syc1273_d*	Allophycocyanin alpha-B chain
apcE	*syc1185_d*	Phycobilisome core-membrane linker polypeptide
apcF	*syc1936_c*	Phycobilisome core component
cpcA	*syc0495_c*	Phycocyanin alpha subunit
cpcA	*syc0500_c*	Phycocyanin alpha subunit
cpcB	*syc0496_c*	Phycocyanin beta subunit
cpcB	*syc0501_c*	Phycocyanin beta subunit
cpcC	*syc0498_c*	Phycobilisome rod linker polypeptide
cpcC	*syc0499_c*	Phycobilisome rod linker polypeptide
cpcD	*syc0497_c*	Phycocyanin linker protein 9 K
cpcE	*syc0494_c*	Phycobilisome maturation protein cpce
cpcF	*syc0493_c*	Phycocyanin alpha-subunit phycocyanobilin lyase
cpcG	*syc2065_d*	Phycobilisome rod-core linker polypeptide
nblA	*syc1965_c*	Phycobilisome degradation protein
nblB	*syc2271_d*	Phycocyanin alpha phycocyanobilin lyase related protein

sensing, *sphS* (*syc0424_c*) and *sphR* (*syc0533_d*) (Aiba et al. 1993), for a protein similar to prokaryotic histidine kinases (*syc2502_c*) (Miller et al. 2000), and for *Synechococcus* response regulators A and B, *srrA* (*syc1690_d*) and *srrB* (*syc0533_d*) (Anandan et al. 1996). The genes *cikA* (*syc0882_c*) and *sasA* (*syc1978_c*) were first identified as genes for a circadian input histidine kinase (Schmitz et al. 2000) and for a KaiC-interacting sensory histidine kinase (Iwasaki et al. 2000, Nagaya et al. 1993), respectively. The *nblS* gene (*syc0534_d*) was identified as a gene for a membrane-bound histidine kinase serving as a non-bleaching sensor (van Waasbergen et al. 2002). The *nblR* (syc1796_d) gene is the OmpR transcriptional regulator of *nblA* (Schwarz and Grossman 1998). The *psfR* gene (*syc1354_c*) is a *psbAI*-stimulating response regulator (Thomas et al. 2004). Recently, Takai et al. (2006) have identified the gene product of *syc1409_d* (*rpaA*) as a response regulator of SasA. Maeda et al. (2006a, b) have recently demonstrated that *syc2277_d*, *syc2278_d*, and *syc2279_d* were involved in latent nitrate transport activity, and designated the genes as *ltnA*, *ltnB*, and *ltnC* respectively. A phosphorelay from LtnB to LtnA and that from LtnA to LtnC were shown.

Among the 37 genes, 30 orthologs (*hik* and *rre*) of *Synechocystis* PCC 6803 were assigned, but the remaining seven genes were not found in *Synechocystis* PCC 6803. Instead, orthologs of *syc0965_c*, *syc1148_c*, *syc1354_c*, *syc1790_c*, *syc1856_d*, and *syc2222_c* were found in other freshwater cyanobacteria. In contrast, orthologs of *syc1690_d* (*srrA*) were found only in marine cyanobacteria, suggesting that a particular signal transduction system occurs only in *S. elongatus* PCC 6301/*S. elongatus* PCC 7942 and marine cyanobacteria but not in other freshwater cyanobacteria including *Synechocystis* PCC 6803. Thus, there are two types of sensory histidine kinases and response regulators. One type includes genes well conserved and widely distributed in various cyanobacterial species. The other type includes genes distributed among specific cyanobacterial strains.

Takai et al. (2006) recently constructed knockout mutants of 21 response regulator genes. Knockout of most response regulator genes was successful, but two genes (*syc0140_c* and *syc2234_d*) could not be disrupted. This suggests that these two genes are indispensable for *S. elongatus* PCC 7942 viability in the laboratory conditions that were tested. Thus, both *S. elongatus* PCC 6301 and *S. elongatus* PCC 7942 are useful for comprehensive study of signal transduction in cyanobacteria.

A total of five genes for serine/threonine protein kinases is found in *S. elongatus* PCC 6301. These include *syc0259_c*, *syc0428_d*, *syc0438_d*, *syc0757_d*, and *syc0923_d*. As is the case with genes for two-component signal transduction, the products of these genes show higher amino acid sequence identities with the corresponding proteins from *Nostoc* PCC 7120 and

Table 3 List of the genes encoding two-component regulatory systems

ORF-ID	aa length	Gene name	Ortholog in *Synechocystis* PCC 6803	Function or Definition in *S. elongatus* PCC 6301/7942 and related species, or ortholog in other cyanobacteria	Ref.
Sensory histidine kinase					
syc0043_d	436		*hik27* (*manS*)		
syc0198_d	1165		*hil35* (*cph1*)	Phytochrome-like	
syc0424_c	311		*hik36* (*pilL-N*, *cheA*)		
syc0534_d	413	*sphS*	*hik7*	Phosphate sensing	Aiba et al. (1993)
syc0618_d	628	*nblS*	*hik33* (*dfr*, *nblS*)	Non-bleaching sensor for phycobilisome degradation	van Waasbergen et al. (2002)
syc0882_c	754	*cikA*	*hik24*	circadian input kinase, phytochrome-like protein	Schmitz et al. (2000)
syc1039_c	577		*hik5*		
syc1065_c	408		*hik2*	Similarity to *E. coli phoM* and *envZ*	Fukuta et al. (1994)
syc1818_d	637		*hik26*		
syc1856_d	330			*alr0117*, ortholog in *Nostoc* PCC 7120	
syc1978_c	399	*sasA*	*hik8*	Synechococcus adaptive-response sensory kinase A	Nagaya et al. (1993)
				KaiC-interacting sensory histidine kinase	Iwasaki et al. (2000)
syc2222_c	482			*alr1192*, ortholog in *Nostoc* PCC 7120	
syc2502_c	457		*hik34*	Significant similarity to prokaryotic histidine kinases	Miller et al. (2000)
Hybrid sensor and regulator					
syc0532_d	928		*hik43* (*pilL-C*, *cheA*)		
syc0681_c	951		*hik18* (*pixL*, *cheA*)		
syc2278_d	937	*ltnB*	*hik6*	Latent transport activity for nitrate, phosphorelay to LtnA	Maeda et al. (2006a)
Response regulator					
syc0104_c	218	*rpaB*	*rre26* (*rpaB*, *ycf27*)	OmpR family	
syc0151_d	229		*rre16* (*manR*)	OmpR family	
syc0199_d	151		*rre27* (*rcp1*)	CheY family	
syc0200_d	929		*rre41*		
syc0329_d	125	*cheY*	*rre7*-like	CheY family	Ishiura et al. (1998)
syc0439_c	208		*rre13*	CheY family	
syc0533_d	257	*sphR*	*rre29* (*phoB*, *sphR*)	OmpR family, phosphate-sensing	Aiba et al. (1993)
syc0684_c	120		*rre7* (*pilH*)	CheY family	
syc0685_c	237		*rre36* (*pixG*)	PatA family	
syc0965_d	237	*srrB*		OmpR family, *Synechococcus* response regulator B	Anandan et al. (1996)
syc1148_c	209			*CYB_2377*, ortholog in *Synechococcus* sp. JA-2–3 B′a	
syc1354_c	796	*psfR*		*psbA1*-stimulating factor	Thomas et al. (2004)
syc1409_d	249	*rpaA*	*rre31* (*rpaA*)	OmpR subfamily, homologous to *Synechocystis rpaA* response regulator of SasA	Takai et al. (2006)
syc1638_c	206		*rre37*	OmpR family	
syc1690_d	255	*srrA*		OmpR family, *Synechococcus* response regulator A	Anandan et al. (1996)
syc1796_c	228	*nblR*		OmpR family, non-bleaching regulatory	Schwarz and Grossman (1998)
syc2221_c	226		*rre28*	OmpR family	
syc2234_d	235		*rre1* (*ycf29*)	NarL family	
syc2277_d	126	*ltnA*	*rre21*	CheY family, latent transport activity for nitrate	Maeda et al. (2006a)
syc2279_d	411	*ltnC*	*rre22*	Latent transport activity for nitrate, phosphorelay from LtnA	Maeda et al. (2006b)
syc2397_d	133		*rre42* (*divK*)		

T. elongatus BP-1 than with the corresponding proteins of *Synechocystis* PCC 6803.

Genes for transcription and sigma factors

Thirty-six genes were assigned as encoding transcription factors in *S. elongatus* PCC 6301. Of 36 genes, nine encoded transcription factors belonging to the response regulator OmpR family (*syc0104_c*, *syc0151_d*, *syc 0533_d*, *syc0965_d*, *syc1409_d*, *syc1638_c*, *syc1690_d*, *syc1796_c*, and *syc2221_c*). The others fall into the LysR family (four genes: *syc0243_c*/*cmpR*/*ndhR*, *syc0308_c*/*ntcB*, *syc1688_d*/*lrrA*, and *syc2116_c*/*rbcR*), the CRP/Fnr family (three genes: *syc1377_c*/*ntcA*, *syc1747_d*, and *syc2406_c*/*cysR*), the ArsR family (three genes: *syc 0262_c*/*ziaR*, *syc0604_c*, and *syc0841_c*) and the FUR family (five genes: *syc0557_d*/*fur*, *syc0722_d*, *syc1925_c*, *syc2290_d*, and *syc2438_d*) while the others are single genes. These single genes include a representative of the LuxR/NarL type-1 family (*syc1834_d*), a chromosomal replication initiator protein DnaA-encoding gene (*syc0449_d*/*dnaA*), genes encoding homologs of the transcriptional regulator ExsB (*syc0221_c*), the heat-inducible transcription repressor HrcA (*syc0908_c*), and the transcription factor MarR family (*syc0391_d*). Other such genes include genes encoding homologs of the circadian period extender PadR family (*syc0852_d*/*pex*), of the glucokinase xylose repressor ROK family (*syc 1981_c*), of the phosphate transport system regulatory protein PhoU family (*syc2183_d*), of the sugar fermentation stimulation protein SfsA (*syc2054_d*), of the TetR/AcrR family (*syc0924_d*), of the CRO/Cl type-3 family (*syc2056_d*), and of the Gnt type-1 repressor (*syc1414_c*). No genes belonging to the LexA or AraC families were found.

Ten genes encoding RNA polymerase sigma factors (*rpoD1–D6*, *sigF1, F2*, *sigG*, and *sigI*) were identified in the genome. There are one gene (*syc0879_d*) encoding the principal sigma factor *rpoD1*, five genes (*syc0953_d*, *syc2345_d, syc2245_c*, *syc0857_c, syc0015_c*) encoding the group 2 sigma factor *rpoD2* to *rpoD6* (Goto-Seki et al. 1999), two genes (*syc2309_c*, *syc2495_c*) for the group 3 sigma factor SigF, and two genes (*syc2171_d*, *syc2091_c*) for group 3 ECF (extracytoplasmic function)-type sigma factors SigG and SigI.

Transposase genes and insertion sequence

A notable feature of gene content in cyanobacteria is that many bacterial transposase genes are spread throughout the genomes. There are 82 such genes in *T. elongatus* BP-1, 112 in *Synechocystis* PCC 6803, 145 in *Nostoc* PCC 7120, and 74 in *G. violaceus* PCC 7421. In contrast, only a single gene encoding a putative bacterial transposase of 422 amino acids (*syc2252_c*) was present in the *S. elongatus* PCC 6301 genome. Two genes (*syc0688_c*, *syc0708_d*) encode pseudotransposases of 158 and 74 amino acids, respectively.

Many insertion sequences (ISs) were found in *T. elongatus* BP-1 (70 ISs), *Synechocystis* PCC 6803 (more than a hundred ISs), and *G. violaceus* PCC 7241 (47 ISs). In contrast, no IS was identified in the *S. elongatus* PCC 6301 genome after BLAST searching using IS families of *Synechocystis* PCC 6803 as query sequences. Large numbers of intact and disrupted IS-like elements in cyanobacteria strongly suggest that rearrangement of the genome occurred frequently during and after establishment of these ancient microorganisms. In general, gene organization and gene arrangement varies greatly among cyanobacterial species. The IS elements may permit genome divergence, offering plasticity permitting organisms to adapt to varying intracellular or extracellular environments.

Comparison of *S. elongatus* PCC 6301 and 7942 genomes

Synechococcus elongatus PCC 7942 (formerly designated *Anacystis nidulans* R2) was the first cyanobacterium reliably transformable by exogenous DNA (Shestakov and Khyen 1970). This allowed advances in genome manipulation, and permitted study of gene expression and function in this strain (Holtman et al. 2005). *S. elongatus* PCC 6301 is closely related to *S. elongatus* PCC 7942 (Golden et al. 1989; Kaneko et al. 1996a). The entire genome sequence of *S. elongatus* PCC 7942 was recently completed by the US Department of Energy Joint Genome Institute (JGI). The genome size is 2,695,903 bp and 2,662 protein-coding genes were assigned (NC_007604). The gene organization and the nucleotide sequence of the *S. elongatus* PCC 6301 genome are nearly identical to those of the *S. elongatus* PCC 7942 genome except for a large inversion of 188.6 kb. As shown (Fig. 2), the positions of the endpoint of the large inversion lie within *syc0094_c* and *syc2450_d* (*somB*) in *S. elongatus* PCC 6301 and within *synpcc7942_1463* and *synpcc 7942_1635* in *S. elongatus* PCC 7942. These genes encode probable major outer membrane porin proteins and the deduced amino acid sequences are highly conserved. The *synpcc7942_1463* sequence is fused by the 5′ portion (245 bp) of *syc0094_c* to the 3′ terminal region (1393 bp) of *syc2450_d*. Similarly, *synppc 7942_1635* is a fused sequence of *syc2450_d* and *syc0094_c*. It is noteworthy that a 20 bp sequence, 5′-ATCGTTGGTTATCCCGATCG-3′ occurs at both

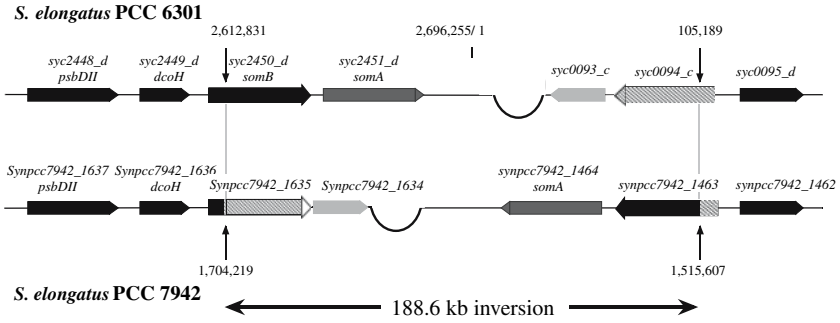

Fig. 2 Gene arrangements near the endpoints of the large inversion in the *Synechococcus elongatus* PCC 6301 genome, relative to that of *S. elongatus* PCC 7942. Nucleotide positions of the endpoints are indicated

endpoints of the large inversion. This is likely to be required for the generation of the large inversion. The 188.6 kb sequence of the *S. elongatus* PCC 6301 inversion showed 99.86% nucleotide identity with that of *S. elongatus* PCC 7942. Nucleotide differences can be explained as single-nucleotide insertions, deletions, or substitutions. Many of these differences may be due to single nucleotide polymorphisms. This implies that small differences in porin-like sequences located at the inversion points might render *S. elongatus* PCC 7942 naturally competent for transformation while such competence is absent in *S. elongatus* PCC 6301. In addition of the large inversion, two small regions are deleted in *S. elongatus* PCC 7942. They include a 52 bp sequence (nucleotide positions 704,012 to 704,063) and a 243 bp sequence (1,783,884 to 1,784,126) in the *S. elongatus* PCC 6301 genome.

Lau et al. (1980) reported two distinct plasmids of 5.3×10^6 daltons (8.2 kb) and 30.5×10^6 daltons (48 kb) in *Synechococcus* strain IU 625 (=*S. elongatus* PCC 6301). Rebière et al. (1986) reported three distinct plasmids of 7.9 kb, 50.8 kb, and 1000–1500 kb in *S. elongatus* PCC 7942. Two plasmids, pUH24 (8 kb) and pANL (46 kb) of *S. elongatus* PCC 7942 have been sequenced (Van der Plas et al. 1992; Holtman et al. 2005) and contain 8 and 59 potential protein-encoding genes, respectively, including a putative two-component sensor histidine kinase. A plasmid pBA1 (8 kb) of *S. elongatus* PCC 6301 (Shinozaki et al. 1982) may correspond to pUH24. A large and mega-plasmid reported by Rebière et al. (1986) has never been further characterized and may correspond to the chromosome itself or a part of the chromosome.

We hope that the *S. elongatus* PCC 6301 genome information will complement that of *S. elongatus* PCC 7942. The sequences as well as the gene information shown in this paper are available in the Web database, CYORF (http://www.cyano.genome.jp/). The sequence data analyzed in this study have been registered in DDBJ/GenBank/EMBL under accession number AP008231.

Acknowledgements This work was supported by Grant-in-Aid for Scientific Research on Priority Area "Genome Biology" (13206027 to MS) of the Ministry of Education, Sports, Science, Culture and Technology, Japan. We thank Tohru Matsubayashi, Takakazu Kaneko and Takahiko Tsudzuki, for valuable advises, and Nikolay Tzverkov, Makoto Suzuki, Tsuneo Fujishiro, Atsuko Tubai, Misato Ogura, Harumi Yokota, Yuriko Hara for technical assistance. We also thank Dr. Kan Tanaka, Prof. Masahiko Ikeuchi (The University of Tokyo), Prof. Ken Nishikawa, Dr. Yoshiaki Minesaki (National Institute of Genetics, Mishima), Dr. Shin-ichi Maeda, Prof. Takao Kondo (Nagoya University), Dr. Hideo Iwasaki (Wasada University) for annotation of protein-encoding genes, and an anonymous referee for valuable comments on the manuscript.

References

Aiba H, Nagaya M, Mizuno T (1993) Sensor and regulator proteins from the cyanobacterium *Synechococcus* species PCC 7942 that belong to the bacterial signal-transduction protein families: implication in the adaptive response to phosphate limitation. Mol Microbiol 8:81–91

Allen MB (1952) The cultivation of Myxophyceae. Arch Mikrobiol 17:34–53

Altschul SF, Madden TL, Scaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Anandan S, Nalty MS, Cogdell DE, Golden SS (1996) Identification of two classes of transcriptional regulator genes in the cyanobacterium *Synechococcus* sp. strain PCC 7942. Arch Microbiol 166:58–63

Banta AB, Haas ES, Brown JW, Pace NR (1992) Sequence of the ribonuclease P RNA gene from the cyanobacterium *Anacystis nidulans*. Nucleic Acids Res 20:911

Bryant DA (ed) (1994) The molecular biology of cyanobacteria. Kluwer Academic Publishers, Dordrecht, Boston, London

Delaney SF, Carr NG (1975) Temporal genetic mapping in the blue-green alga *Anacystis nidulans* using ethyl methanesulphonate. J Gen Microbiol 88:259–268

Dufresne A, Salanoubat M, Partensky F, Artiguenave F, Axmann IM, Barbe V, Duprat S, Galperin MY, Koonin EV, Le Gall F, Makarova KS, Ostrowski M, Oztas S, Robert C, Rogozin IB, Scanlan DJ, Tandeau de Marsac N, Weissenbach J, Wincker P, Wolf YI, Hess WR (2003) Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. Proc Natl Aca Sci USA 100:10020–10025

Fukuta M, Wakasugi T, Sugiura M (1994) Nucleotide sequence of a putative sensory kinase gene from the cyanobacterium, *Anacystis nidulans* 6301. DNA Res 1:97–98

Golden SS, Nalty MS, Cho DSC (1989) Genetic relationship of two highly studied *Synechococcus* strains designated *Anacystis nidulans*. J Bacteriol 171:24–29

Goto-Seki A, Shirokane M, Masuda S, Tanaka K, Takahashi H (1999) Specificity crosstalk among group 1 and group 2 sigma factors in the cyanobacterium *Synechococcu*s sp. PCC7942: *in vitro* specificity and a phylogenetic analysis. Mol Microbiol 34:473–484

Gupta A, Morby AP, Turner JS, Whitton BA, Robinson NJ (1993) Deletion within the metallothionein locus of cadmium-tolerant *Synechococcus* PCC 6301 involving a highly iterated palindrome (HIP1). Mol Microbiol 7:189–195

Holtman CK, Chen Y, Sandoval P, Gonzales A, Nalty MS, Thomas TL, Youderian P, Golden SS (2005) High-throughput functional analysis of the *Synechococcus elongatus* PCC 7942 genome. DNA Res 12:103–115

Ishiura M, Kutsuna S, Aoki S, Iwasaki H, Andersson CR, Tanabe A, Golden SS, Johnson CH, Kondo T (1998) Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobacteria. Science 281:1519–1523

Iwasaki H, Williams SB, Kitayama Y, Ishiura M, Golden SS, Kondo T (2000) A KaiC-interacting sensory histidine kinase, SasA, necessary to sustain robust circadian oscillation in cyanobacteria. Cell 101:223–233

Kaneko T, Matsubayashi T, Sugita M, Sugiura M (1996a) Physical and gene maps of the unicellular cyanobacterium *Synechococcus* sp. strain PCC6301 genome. Plant Mol Biol 31:193–201

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S (1996b) 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. DNA Res 3:109–136

Kaneko T, Nakamura Y, Wolk CP, Kuritz T, Sasamoto S, Watanabe A, Iriguchi M, Ishikawa A, Kawashima K, Kimura T, Kishida Y, Kohara M, Matsumoto M, Matsuno A, Muraki A, Nakazaki N, Shimpo S, Sugimoto M, Takazawa M, Yamada M, Yasuda M, Tabata S (2001) Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. DNA Res 8:205–213

Kumano M, Tomioka N, Sugiura M (1983) The complete nucleotide sequence of a 23S rRNA gene from a blue-green alga, *Anacystis nidulans*. Gene 24:219–225

Lau RH, Sapienza C, Doolittle WE (1980) Cyanobacterial plasmids: their widespread occurrence, and existence of regions of homology between plasmids in the same and different species. Mol Gen Genet 178:203–211

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

Maeda S, Sugita C, Sugita M, Omata T (2006a) Latent nitrate transport activity of a novel sulfate permease-like protein of the cyanobacterium *Synechococcus elongatus*. J Biol Chem 281:5869–5876

Maeda S, Sugita C, Sugita M, Omata T (2006b) A new class of signal transducer in His-Asp phosphorelay systems. J Biol Chem 281:37868–37876

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

Miller B, Heuser T, Zimmer W (2000) Functional involvement of a deoxy-D-xylulose 5-phosphate reductoisomerase gene harboring locus of *Synechococcus leopoliensis* in isoprenoid biosynthesis. FEBS Lett 481:221–226

Mizuno T, Kaneko T, Tabata S (1996) Compilation of all genes encoding bacterial two-component signal transducers in the genome of the cyanobacterium, *Synechocystis* sp. strain PCC 6803. DNA Res 3:407–414

Nagaya M, Aiba H, Mizuno T (1993) Cloning of a sensory-kinase-encoding gene that belongs to the two-component regulatory family from the cyanobacterium *Synechococcus* sp. PCC7942. Gene 131:119–124

Nakamura Y, Kaneko T, Sato S, Ikeuchi M, Katoh H, Sasamoto S, Watanabe A, Iriguchi M, Kawashima K, Kimura T, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Sugimoto M, Takeuchi C, Yamada M, Tabata S (2002) Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. DNA Res 9:123–130

Nakamura Y, Kaneko T, Sato S, Mimuro M, Miyashita H, Tsuchiya T, Sasamoto S, Watanabe A, Kawashima K, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Takeuchi C, Yamada M, Tabata S (2003) Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. DNA Res 10:137–145

Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P, Lamerdin J, Regala W, Allen EE, McCarren J, Paulsen I, Dufresne A, Partensky F, Webb EA, Waterburg J (2003) The genome of a motile marine *Synechococcus*. Nature 424:1037–1042

Palenik B, Ren Q, Dupont CL, Myers GS, Heidelberg JF, Badger JH, Madupu R, Nelson WC, Brinkac LM, Dodson RJ, Durkin AS, Daugherty SC, Sullivan SA, Khouri H, Mohamoud Y, Halpin R, Paulsen IT (2006) Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. Proc Natl Acad Sci USA 103:13555–13559

Paquin B, Kathe SD, Nierzwicki-Bauer SA, Shub DA (1997) Origin and evolution of group I introns in cyanobacterial tRNA genes. J Bacteriol 179:6798–6806

Rebière MC, Castets AM, Houmard J, Tandeau de Marsac N (1986) Plasmid distribution among unicellular and filamentous cyanobacteria: occurrence of large and mega-plasmids. FEMS Microbiol Lett 37:269–275

Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgrn NA, Arellano A, Coleman M, Hauser L, Hess WR, Johnson ZI, Land M, Lindell D, Post AF, Regala W, Shah M, Shaw SL, Steglich C, Sullivan MB, Ting CS, Tolonen A, Webb EA, Zinser ER, Chisholm SW (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature 424:1042–1047

Schmitz O, Katayama M, Williams SB, Kondo T, Golden SS (2000) CikA, a bacteriophytochrome that resets the cyanobacterial circadian clock. Science 289:765–768

Schwarz R, Grossman AR (1998) A response regulator of cyanobacteria integrates diverse environmental signals and is critical for survival under extreme conditions. Proc Natl Acad Sci USA 95:11008–11013

Shestakov SV, Khyen NT (1970) Evidence for genetic transformation in blue-green alga *Anacystis nidulans*. Mol Gen Genet 107:372–375

Shinozaki K, Tomioka N, Yamada C, Sugiura M (1982) Cloning and characterization of a plasmid DNA from *Anacystis nidulans* 6301. Gene 19:221–224

Sugita M, Luo L, Ohta M, Itadani H, Matsubayashi T, Sugiura M (1995) Genes encoding the group I intron-containing $tRNA^{Leu}$ and subunit L of NADH dehydrogenase from the cyanobacterium *Synechococcus* PCC 6301. DNA Res 2:71–76

Sugita M, Tsudzuki T, Sugita C, Ishiura M, Ogata K, Jikuya H, Takano J, Sugiura M (2001) Genome analysis of *Synechococcus* sp. strain PCC6301. In: PS2001 Proceedings: 12th International Congress on Photosynthesis, S41–009, CSIRO Publishing, Melbourne (CD-ROM)

Takai N, Nakajima M, Oyama T, Kito R, Sugita C, Sugita M, Kondo T, Iwasaki H (2006) A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. Proc Natl Acad Sci USA 103:12109–12114

Thomas C, Andersson CR, Canales SR, Golden SS (2004) PsfR, a factor that stimulates *psbAI* expression in the cyanobacterium *Synechococcus elongatus* PCC 7942. Microbiol 150:1031–1040

Tomioka N (1983) Effect of culture conditions on hydrogenase activity and H_2-supported nitrogenase activity in some blue-green algae. J Sci Hiroshima Univ Ser B, Div 2 (Bot) 18:173–190

Tomioka N, Shinozaki K, Sugiura M (1981) Molecular cloning and characterization of ribosomal RNA genes from a blue-green alga, *Anacystis nidulans*. Mol Gen Genet 184:359–363

Tomioka N, Sugiura M (1983) The complete nucleotide sequence of a 16S ribosomal RNA gene from a blue-green alga, *Anacystis nidulans*. Mol Gen Genet 191:46–50

Van der Plas J, Oosterhoff-Teertstra R, Borrias M, Weisbeek P (1992) Identification of replication and stability functions in the complete nucleotide sequence of plasmid pUH24 from the cyanobacterium *Synechococcus* sp. PCC 7942. Mol Microbiol 6:653–664

van Waasbergen LG, Dolganov N, Grossman AR (2002) *nblS*, a gene involved in controlling photosynthesis-related gene expression during high light and nutrient stress in *Synechococcus elongatus* PCC 7942. J Bacteriol 184:2481–2490

Watanabe T, Sugiura M, Sugita M (1997) A novel small stable RNA, 6Sa RNA, from the cyanobacterium *Synechococcus* sp. strain PCC6301. FEBS Lett 416:302–306

Watanabe T, Sugita M, Sugiura M (1998) Identification of 10Sa RNA (tmRNA) homologues from the cyanobacterium *Synechococcus* sp. strain PCC6301 and related organisms. Biochim Biophys Acta 1396:97–104