Microevolution in Cyanobacteria: Re-sequencing a Motile Substrain of *Synechocystis* sp. PCC 6803

DANIKA Trautmann, BJÖRN Voß, ANNEGRET Wilde, SALIM Al-Babili, and WOLFGANG R. Hess*

Faculty of Biology, University of Freiburg, Schänzlestr. 1, D-79104 Freiburg, Germany

*To whom correspondence should be addressed. Tel. +49 761-203-2796. Fax. +49 761-203-2445. E-mail: wolfgang.hess@biologie.uni-freiburg.de

Edited by Naotake Ogasawara
(Received 23 July 2012; accepted 5 September 2012)

Abstract

*Synechocystis* sp. PCC 6803 is a widely used model cyanobacterium for studying photosynthesis, phototaxis, the production of biofuels and many other aspects. Here we present a re-sequencing study of the genome and seven plasmids of one of the most widely used *Synechocystis* sp. PCC 6803 substrains, the glucose tolerant and motile Moscow or ‘PCC-M’ strain, revealing considerable evidence for recent microevolution. Seven single nucleotide polymorphisms (SNPs) specifically shared between ‘PCC-M’ and the ‘PCC-N and PCC-P’ substrains indicate that ‘PCC-M’ belongs to the ‘PCC’ group of motile strains. The identified indels and SNPs in ‘PCC-M’ are likely to affect glucose tolerance, motility, phage resistance, certain stress responses as well as functions in the primary metabolism, potentially relevant for the synthesis of alkanes. Three SNPs in intergenic regions could affect the promoter activities of two protein-coding genes and one *cis*-antisense RNA. Two deletions in ‘PCC-M’ affect parts of clustered regularly interspaced short palindromic repeats-associated spacer-repeat regions on plasmid pSYSA, in one case by an unusual recombination between spacer sequences.

Key words: CRISPR; genome sequence; plasmid; substrain; *Synechocystis* sp. PCC 6803

1. Introduction

With currently >4000 publications available from PubMedCentral alone, ‘Synechocystis’ is the most widely used photoautotrophic prokaryotic model organism. *Synechocystis* sp. PCC 6803 is a unicellular cyanobacterium that was isolated from a freshwater pond in Oakland, California.\(^1\) The high popularity of *Synechocystis* sp. PCC 6803 stems from the two facts that it was the first phototrophic and the third organism overall, for which a complete genome sequence was determined,\(^2\) and that it easily takes up exogenous DNA and integrates it into its chromosome by homologous recombination.\(^3-5\)

*Synechocystis* sp. PCC6803 is known to occur in several distinct substrains, all going back to the same isolate deposited in the Pasteur Culture Collection.\(^6\) Indeed, several studies reported the differences between the genome sequence of *Synechocystis* sp. PCC 6803 published in 1996 (called here the ‘GT-Kazusa’ substrain) and the actual sequence found in different laboratories.\(^7-10\) A strain history has been proposed by Ikeuchi and Tabata\(^8\) with an early branching into the motile PCC strain and the non-motile ATCC 27184 strain. The latter lost motility due to a 1-bp insertion in the *spkA* gene coding for a eukaryotic-type Ser/Thr protein kinase\(^11\) and represents the origin of the glucose-tolerant (GT) strains\(^5\) to which also the ‘GT-Kazusa’ substrain belongs.

For decades, *Synechocystis* sp. PCC 6803 has served as a simple model in photosynthesis research and to solve fundamental questions in microbial and plant physiology. More recently, cyanobacteria are increasingly being recognized as a promising resource for the production of biofuels such as hydrogen,\(^12\) ethanol,\(^13\) isobutyraldehyde and isobutanol,\(^14\) ethylene\(^15\) and alkanes.\(^16\) *Synechocystis* sp. PCC 6803 is being developed further as a model in these biotechnology- and systems biology-oriented studies. These facts as well as the search for motility-associated
genes prompted several re-sequencing studies of *Synechocystis* sp. PCC 6803 substrains, namely of the substrains GT-S, PCC-P, PCC-N, GT-I and YF. However, these studies have not included the widely used GT and motile ‘Moscow’ substrain, which we here suggest to call ‘PCC-M’. Furthermore, thus far no attention has been paid to the possible sequence variations in the seven plasmids, which constitute a total sequence length of 383 486 bp almost 10% of the total coding capacity of *Synechocystis* sp. PCC 6803. This analysis provides new and reliable sequence data for the *Synechocystis* sp. PCC 6803 substrain ‘PCC-M’, revealing several differences from the published sequence that can be interpreted as the traces of microevolution during cultivation in the laboratory.

2. Materials and methods

2.1. Origin of strain, isolation of DNA and PCR analysis

*Synechocystis* sp. PCC 6803 substrains ‘Moscow’ here called ‘PCC-M’, Kazusa (GT-Kazusa) and Vermaas’ (GT-V) were cultivated by Prof. Annegret Wilde (University of Freiburg, Germany) and maintained as frozen stocks. The ‘PCC-M’ substrate was originally obtained from the laboratory of S. Shestakov (Moscow State University) in 1993 and over the years carefully propagated for motile colonies. The ‘GT-V’ strain originates from the laboratory of W. Vermaas (Arizona State University). Genomic DNA for deep sequencing analysis was isolated from 80 ml cultures harvested on a glass microfiber filter (GF/C, 47 mm i.d. Whatman) by vacuum filtration. The frozen filter was ground in a mixer mill (Dismembrator MM301, Retsch, Germany) and the powder transferred into 1 ml SET buffer on ice (25% (w/v) sucrose, 1 mM EDTA, 50 mM Tris pH 7.5). One-fourth volume of 0.5 M EDTA, 2% SDS and 1.5 mg proteinase K (Sigma) were added for cell lysis at 50°C overnight. Following phenol/chloroform extraction, one volume of 2-propanol (Roth, Germany) was added for precipitating the DNA at room temperature for 30 min. The precipitate was washed once in H2O/2-propanol 1:1 and once in 2-propanol, followed by 10 min centrifugation at 10 000 g, 4°C. The pellet was washed with 70% EtOH, dried for 10 min and re-suspended in 50 μl H2O. One microlitre of RNase A (Sigma) was added and the tube incubated at 37°C and 260 rpm overnight. RNase was removed by another round of phenolic extraction and precipitation as described above. The DNA was re-suspended in 75 μl H2O, concentration was measured photometrically and DNA quality checked on a gel (0.8% agarose).

Genomic DNA for PCR was isolated from the cell pellet of 1 ml *Synechocystis* liquid culture. The pellet was washed once with a 1:10 dilution of TE buffer (10 mM Tris HCl pH 8.1 mM EDTA) and re-suspended in 70 μl of the same buffer. Cells were broken by incubation at 98°C for 10 min. After centrifugation at 14 000 g and 4°C for 5 min, the supernatant was collected and kept on ice. Two microlitres of it were used for PCR. For PCR reactions, Phusion® DNA polymerase (Finnzymes, New England Biolabs) was used according to the manufacturer’s instructions. To verify single nucleotide polymorphisms (SNPs) between the different substrains, ~500 bp fragments containing the SNP position were amplified. PCR products were excised from an agarose gel, purified (illustra GFX PCR DNA and Gel Band Purification Kit, GE Healthcare) and sent for Sanger sequencing to GATC Biotech (Konstanz, Germany). For sequencing of the small plasmids, several PCR reactions were performed to get overlapping sequences and contigs were assembled using the software ContigExpress (Vector NTI Advance 11, Invitrogen). Alignments of the sequences were performed using AlignX (Vector NTI Advance 11, Invitrogen).

2.2. Sequencing methods and mapping

Sequencing of genomic DNA was carried out on an Illumina Genome Analyzer IIx system. Prior to sequencing, the DNA was sheared by ultrasonication (Covaris, Woburn, MA, USA), resulting in fragments of 300 bp length on average. For these fragments paired-end sequencing according to the manufacturer’s protocol was carried out, resulting in 42 143 495 million 101 nt long reads. These reads were analysed with two methods in order to identify SNPs, deletions and insertions. For the first approach, we used the DNA sequence data assembler algorithm MIRA (Mimicking Intelligent Read Assembly) to perform an assembly of the reads using the ‘GT-Kazusa’ genome as the reference. In the assembly process, MIRA generates tables of candidate SNPs, insertions and deletions. We verified these results independently by mapping all sequencing reads to the assembled chromosome and plasmid sequences. This was done using segemehl, requiring at least 85% accuracy and reporting only the best hit. It should be noted that segemehl reports co-optimal best hits.

3. Results

3.1. Overview

Sequencing of the *Synechocystis* sp. PCC 6803 ‘Moscow’ substrain ‘PCC-M’ by Illumina (Solexa) yielded an average 1100-fold coverage of the chromosome and five of the seven plasmids. The
existence of the two remaining plasmids was verified individually by PCR. Following assembly of sequences, mapping to the reference strain sequences and annotation, the obtained genome and plasmid sequences were deposited in the GenBank database with the accession numbers CP003265–CP003272.

Altogether, we found 45 differences (36 SNPs and 9 indels >1 bp) between the investigated substrate 'PCC-M' and the published sequences of the 'GT-Kazusa' chromosome and plasmids used here as references (Table 1). From these differences, 41 are located in the chromosome and four in the plasmids pSYSA, pSYSM and pCA2.4. For verification, about one-third of these differences were randomly chosen and confirmed independently by PCR and Sanger sequencing of the respective regions in substrate 'PCC-M', but no misidentified mutations were found. These DNA regions were, in addition, amplified and compared with the sequences from substrains 'GT-Kazusa' and 'GT-V' for control and comparison, respectively. The GT 'GT-V' was chosen for comparison as is widely used for the dissection and analysis of photosynthetic mutants. Fully segregated PSI, PSII and Chl biosynthesis mutants were successfully generated photosynthetic mutants. Fully segregated PSI, PSII and Chl biosynthesis mutants were successfully generated.

The number of differences between 'PCC-M' and 'GT-Kazusa' are almost twice as many as reported by Tajima et al.10 for the GT (GT-S) 'Kazusa' strain, where a total of 22 differences from the published sequence were found.10 All but 3 of those 22 differences were also detected in the 'PCC-M' strain studied here. The three unique differences in the 'GT-S' and 26 differences between 'PCC-M' and 'GT-Kazusa' underline the existence of lineage splitting in the Synechocystis substrains. Moreover, we found seven SNPs (#5, 13, 15, 16, 27, 32 and 33 in Tables 1 and 2) and one larger indel (#6 in Tables 1 and 2) specifically shared between the 'PCC-M' and the 'PCC-N and PCC-P' substrains, indicating that 'PCC-M' belongs to the 'PCC' group of motile substrains.9 'PCC-M and PCC-P' are strains that both exhibit the native positive phototaxis, whereas 'PCC-N' strain shows negative phototaxis.24

3.2. SNPs in protein-coding genes

Of the total of 36 SNPs in 'PCC-M' compared with 'GT-Kazusa', all except 1 are located in the chromosome. The single base substitution that was found on the plasmid pCA2.4 within the repA gene (#42 in Table 1) seems to be no mutation but an error in the published sequence of 'GT-Kazusa', since in our PCR-control experiments, the sequence was identical in the three strains 'GT-Kazusa', 'PCC-M' and 'GT-V'. Of the 35 chromosomal SNPs compared with 'GT-Kazusa', 5 are silent base substitutions, 14 substitutions lead to amino acid substitutions, in 6 cases a single basepair is deleted and in 2 cases (#23 and #28) one basepair was inserted within an ORF, causing a frameshift mutation. Furthermore, five substitutions, two single basepair insertions and one single basepair deletion were observed in intergenic regions (IGR) of 'PCC-M' compared with the reference (Table 1).

Seven SNPs are specifically shared between the 'PCC-M', 'PCC-N and PCC-P' substrains. These are in slr1865 (#13), encoding a hypothetical protein, in slr1951 (#15), encoding a haemolysin-like protein, in slr1983 (#16), encoding a two-component hybrid sensor and regulator protein, in slr0222 (#27), encoding the histidine kinase Hik25, a silent mutation in slr0302 (#32), encoding a PAS/PAC and GAF sensors-containing diguanylate cyclase, one missing basepair, leaving the splA gene intact (#5) and, finally, in ssr1176 (#33), encoding a transposase (Tables 1 and 2).

The gene for a cell surface-localized haemolysin-like protein, HlyA (slr1951), reported to function as a barrier against the adsorption of toxic compounds,25,26 is lacking one nucleotide in 'PCC-M' compared with the reference (difference #15). In the 'GT-Kazusa', 'GT-V' as well as the 'GT-I' and 'GT-S' strains,9 the presence of the additional A leads to the fusion of two ORFs that are separate in 'PCC-M', 'PCC-N' and 'PCC-P' substrains.9 As a result, Slr1951 is 1741 amino acids in the former and only 1437 residues in the latter.

In our data, some other previously published mutations8,10 are confirmed. For instance, splA (slr1574; #5), a regulator of cellular motility via phosphorylation of membrane proteins,11,27 is disrupted by a 1-bp insertion in the non-motile 'GT-Kazusa' and 'GT-V' strains, whereas it is intact in the motile 'PCC-M' strain (Table 1). Similarly, the pilC gene (slr0162/3) required for pili assembly has been reported to carry a frameshift mutation in the 'GT-Kazusa' and 'GT-S' sequences.8,10,28 We found an intact pilC gene in 'PCC-M' (#20), as well as in the 'GT-V' substrate.

Another SNP (G–A) exists in psaA (slr1834; #9), encoding the photosynthetic P700 apoprotein subunit Ia; however, in accordance with Tajima et al.,10 we believe this is an annotation error in the database as we found an A in the respective position in all three strains dealt with in this work (Table 1). Similarly, ycf22 (slr0751; #26) is here suggested to be fused to the downstream reading frame slr0752. Indeed, in blastp comparisons, both proteins together match against a single, widely distributed, larger protein of 452 amino acids. This protein possesses a Ttg2C domain (COG1463), which is found in an ABC-type...
Table 1. Location and effects of SNPs and indels found in ‘PCC-M’ compared with the nucleotide sequence of ‘GT-Kazusa’ in the database

| Event | Start | End | Size | Ref → mut | Effect | Result | Locus | Gene name | Product |
|-------|-------|-----|------|-----------|--------|--------|-------|-----------|---------|
| 1     | 144 507 | 14 4507 | 1    | A → G | GTA → GTG | V → V | silent | slr0242 | bcp      | Bacterioferritin comigratory protein homolog |
| 2     | 386 411 | 102  | 1    | C → T | TTA → TCA | L → S | AA change | slr1609 | Acyl-ACP synthetase (AAS) |
| 3     | 527 395 | 600  | 1    | A → * | AAT → ATT | N → I | Frameshift | slr1753 | Hypothetical protein |
| 4     | 781 626 | 154  | 1    | C → T | Possible effect on infA promoter | IGR_slr2030_slr2031 | Serine phosphatase, regulator of sigma subunit |
| 5     | 831 647 | 1    | 1    | G → A | AGC → AAC | S → N | AA change | slr1898 | argB      | N-acetylglutamate kinase |
| 6     | 943 495 | 1    | 1    | G → A | GTC → ATC | V → I | AA change | slr1834 | psaA      | P700 apoprotein subunit Ia |
| 7     | 1 012 958 | 1 012 958 | 1    | G → T | AAT → AAA | N → K | AA change | sll1359 | Predicted cytochrome c |
| 8     | 1 070 839 | 1 070 839 | 1    | T → A | ISY203b missing; slr1862/3 | slr1780 (transposase); slr1862/3 | Hypothetical protein |
| 9     | 1 200 290 | 1 201 474 | 1185 | G → A | TGT → TAT | C → Y | AA change | slr1865 | Hypothetical protein |
| 10    | 1 364 187 | 1 364 187 | 1    | T → C | TGG → CTG | L → L | silent | slr0838 | pyrF      | Orotidine 5'-monophosphate decarboxylase |
| 11    | 1 423 340 | 1 423 340 | 1    | A → * | GAC → GCA | D → A | Frameshift, protein truncated | sll1951 | hlyA      | Haemolysin |
| 12    | 1 812 419 | 1 812 419 | 1    | C → T | GCG → GTG | A → V | AA change | slr1983 | Two-component hybrid sensor and regulator |
| 13    | 2 048 403 | 2 049 585 | 1183 | G → A | ISY203e missing | slr1635 (transposase); slr1636 | Hypothetical protein |
| 14    | 2 092 571 | 2 092 571 | 1    | T → A | TTA → TAA | L → * | AA change, new stop codon | slr0422 | Asparaginase |
| 15    | 2 198 893 | 2 198 893 | 1    | A → G | TTA → TGT | L → L | silent | slr0142 | Probable cation efflux system protein |
| 16    | 2 204 584 | 2 204 584 | 1    | G → * | GGT → GTT | G → V | Frameshift | slr0162 | pilC      | Part of PilC, pilin biogenesis protein, twitching motility |
| No. | Coordinates | Type | Description | References |
|-----|--------------|------|-------------|------------|
| 21 | S 2301721 2301721 | I | A → G | AA change<sup>a</sup> | slr0168 | Hypothetical protein, no conserved domains |
| 22 | I 2350285 2350286 | I | * → A | | IGR_sml0001_slr0363 | Hypothetical protein, no conserved domains |
| 23 | I 2360245 2360246 | I | * → C | Frameshift<sup>a</sup> | slr0364/slr0366 | Hypothetical protein, no conserved domains |
| 24 | S 2400722 2400722 | I | C → A | Possible effect on glcP promoter | IGR_sll0771_slr0774 (IGR_glcP_secD) | |
| 25 | D 2409244 2409244 | I | G → * | Frameshift<sup>a</sup> | slr0762 | Hypothetical protein, no conserved domains |
| 26 | D 2419399 2419399 | I | A → * | Frameshift<sup>a</sup> | slr0751 (ycf22); slr0752 | Hypothetical protein YCF22 |
| 27 | S 2521013 2521013 | I | T → C | AA change | slr0222 | hik25 Two-component hybrid sensor and regulator |
| 28 | I 2544045 2544046 | I | * → G | Frameshift<sup>a</sup> | ssl0787/ssl0788 | Hypothetical protein, no conserved domains |
| 29 | S 2602717 2602717 | I | C → A | AA change<sup>a</sup> | slr0468 | Hypothetical protein, no conserved domains |
| 30 | S 2602734 2602734 | I | T → A | AA change<sup>a</sup> | slr0468 | Hypothetical protein, no conserved domains |
| 31 | S 2748897 2748897 | I | C → T | | IGR_slr0210_ssr0332 | |
| 32 | S 3014665 3014665 | I | T → C | - silent - | slr0302 | pleD PleD-like protein |
| 33 | S 3098707 3098707 | I | T → C | AA change | ssr1176 (transposase) | Located in a mobile element (ISY100v3) |
| 34 | S 3110189 3110189 | I | G → A | | IGR_ssl0665_ssl0666 | Located in a mobile element (ISY523) |
| 35 | S 3142651 3142651 | I | T → C | - silent<sup>a</sup> | ssl0045 | spsA Sucrose phosphate synthase |
| 36 | I 3194022 3194023 | I | * → A | Possible effect on slr0534_as3 promoter | | |
| 37 | D 3260096 3260096 | I | C → * | Frameshift | IGR_slr0533_slr0534 (IGR_hik10_slt) | |
| 38 | D 3364288 3364288 | I | A → * | Frameshift | slr1496 | Mannose-1-phosphate guanylyltransferase |
| 39 | S 3371938 3371938 | I | T → A | AA change | slr1564 | sigF Group 3 RNA polymerase sigma factor |
| 40 | D 3400331 3401513 1183 | D | * → | ISY203g missing | slr1474 (transposase) | ccaS sll1473/s = His-Kinase/ATPase |
| 41 | S 3423372 3423372 | I | C → T | AA change | slr0753 | Probable transport protein (anion permease) |

Plasmid pCA2.4

| No. | Coordinates | Type | Description | References |
|-----|--------------|------|-------------|------------|
| 42 | S 1127 1127 | I | T → G | - silent<sup>a</sup> | repA Replication initiation factor | |

*Continued*
Transport system involved in resistance to organic solvents. The acronym ycf stands for hypothetical chloroplast reading frames, meaning proteins conserved in chloroplasts and also cyanobacteria. The 1-bp shorter version, which is splitted into sl0751/sll0752, is a database error in the case of ‘GT-Kazusa’ as well.

3.2.1. SNPs unique to ‘PCC-M’ Six of the 10 SNPs unique to ‘PCC-M’ are located within coding regions and cause amino acid substitutions or alter the length of the respective reading frame.

A single basepair transversion in the gene sigF (slr1564; #39 in Table 1) is leading to a M231K substitution within the −35 element DNA-binding region^29 of a group 3 sigma factor required for phototactic movement^30 and salt-stress response. This SNP cannot lead to impaired motility as ‘PCC-M’ is motile but it might influence the DNA–protein interaction of SigF because positively charged residues such as lysine located in this part of the σ42 region can directly interact with DNA.^29

Another transversion, in argB (slr1898; #8 in Table 1), leads to an S2N amino acid substitution in N-acetylglutamate kinase, the enzyme performing the first committed step of Arg biosynthesis. Transitions in slr1359 and slr1609 (#11 and #3 in Table 1) result in an N–K substitution at a very conserved position within a predicted cytochrome and an L608S (L548S) substitution in the long-chain acyl-CoA-synthetase Slr1609 that has been found crucial for fatty acid activation and the biosynthesis of alkanes.32 Interestingly, an unrelated SNP exists at position 488923 within the slr1609 coding sequence in a strain ‘YF’, leading to a G546L (G486L) substitution.17 It should be noted that the slr1609 reading frame has been annotated 60 codons shorter (636 instead of 696 amino acids) during recent re-sequencing analyses,9,10 compared with the original annotation of ‘GT-Kazusa’ (numbers in brackets). The shorter Slr1609 protein of 636 amino acids is also consistent with the mapped start site of transcription at position 487352,33 located 115 nt upstream of the revised start codon.

A transition in slr0753 (#41 in Table 1) leads to a P113L substitution in a putative chloride efflux transporter involved in maintaining the chloride ion concentration homoeostasis as required for a functional photosystem II.34

A single basepair deletion in slr1496 (#38 in Table 1), encoding mannose-1-phosphate guanylyltransferase, causes a frameshift and premature stop of the gene in ‘PCC-M’. The resulting protein is with 515 instead of 643 amino acids severely truncated and may be rendered function-less.

Table 1. Continued

| Event | Locus | Gene name | Effect | Size | Nucl change | Ref AA change | Result |
|-------|-------|-----------|--------|------|-------------|---------------|--------|
| 43 D 17 34 | slr0751 | Hypothetical protein | S | 399 | 2399 | | |
| 44 D 71 57 | slr0752 | Hypothetical protein | D | 159 | 71596 | | |
| 45 D 117 26 | slr1359 | Hypothetical protein | I | 1183 | 118451 | | |

The events are numbered (column #), the type of mutation (M) is indicated as S, substitution, D, deletion or I, insertion, together with the respective start and end positions in the ‘GT-Kazusa’ reference sequence. For each event the respective nucleotide change is indicated on the forward strand, together with the resulting codon modification (Ref. → Mut.) and amino acid change, if any. Highlighted in italics are four instances of missing ISY203 copies and in bold all SNPs affecting intergenic spacer regions (IGR).
Table 2. Comparison of SNPs and indels found in the chromosome of 'PCC-M' with sequences from other substrains

| Event | Comparison of strains: literature + this work | GT-Kazusa \(^{9,10}\) | GT-S\(^{1,10}\) | GT-I\(^{1}\) | PCC-P\(^{9}\) | PCC-N\(^{9}\) | PCC-M |
|-------|---------------------------------|-----------------|----------------|------------|--------------|--------------|------|
| 1     | S                              | —               | —              | —          | —            | —            | ✓    |
| 2     | I                              | —               | —              | ✓          | —            | —            | —    |
| 3     | S                              | —               | —              | —          | —            | —            | —    |
| 4     | D                              | ✓\(^{a}\)       | —              | —          | —            | —            | —    |
| 5     | D                              | —               | —              | —          | ✓            | ✓            | ✓    |
| 6     | I                              | —               | —              | ✓          | ✓            | ✓            | ✓    |
| 7     | S                              | —               | —              | —          | —            | —            | —    |
| 8     | S                              | —               | —              | —          | —            | —            | —    |
| 9     | S                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 10    | S                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 11    | S                              | —               | —              | —          | —            | —            | —    |
| 12    | D                              | —               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 13    | S                              | —               | —              | —          | ✓            | ✓            | ✓    |
| 14    | S                              | ✓               | ✓              | —          | ✓            | ✓            | ✓    |
| 15    | D                              | —               | —              | ✓          | ✓            | ✓            | ✓    |
| 16    | S                              | —               | —              | —          | ✓            | ✓            | ✓    |
| 17    | D                              | —               | —              | ✓          | ✓            | ✓            | ✓    |
| 18    | S                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 19    | S                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 20    | D                              | —               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 21    | S                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 22    | I                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 23    | I                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 24    | S                              | —               | —              | —          | —            | —            | —    |
| 25    | D                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 26    | D                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 27    | S                              | —               | —              | —          | ✓            | ✓            | ✓    |
| 28    | I                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 29    | S                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 30    | S                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 31    | S                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 32    | S                              | —               | —              | —          | ✓            | ✓            | ✓    |
| 33    | S                              | —               | —              | —          | ✓            | ✓            | ✓    |
| 34    | S                              | —               | —              | —          | ✓            | ✓            | ✓    |
| 35    | S                              | ✓               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 36    | I                              | —               | —              | —          | —            | —            | ✓    |
| 37    | D                              | —               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 38    | D                              | —               | —              | —          | —            | —            | ✓    |
| 39    | S                              | —               | —              | —          | —            | —            | ✓    |
| 40    | D                              | —               | ✓              | ✓          | ✓            | ✓            | ✓    |
| 41    | S                              | —               | —              | —          | —            | —            | ✓    |

All events are numbered (column #) as in Table 1. The presence of the respective 'PCC-M' mutation in the different substrains is indicated by the check marks.

\(^{a}\)The deletion of 0.6 kb in the gene slr1753 compared with the reference was also verified here in 'GT-Kazusa'.
3.3. Point mutations in IGRs

Compared with the reference, eight SNPs are located in IGRs, three of these (#7, 24 and 36) are ‘PCC-M’ specific. One of these (#36 in Table 1) SNPs is predicted to affect one of the recently reported cis-antisense RNAs. The additional A between positions 3194022 and 3194023 is located in the IGR between genes slr0533 and slr0534, encoding histidine kinase 10 (Hik10) and the soluble lytic transglycosylase Slt. On the reverse strand, the additional T falls within the predicted −10 element of the slr0534_as3 promoter. Instead of the high-scoring CATAAT the motif is changed to ATTAAT. Hence, a modulation of slr0534_as3 expression compared with the reference is possible. In contrast to its designation, this cis-antisense RNA overlaps the 3’ end of genes slr0533 and hik10 (due to an error in the annotation used as the reference). In microarray analyses, slr0534_as3 of strain ‘PCC-M’ was found to be moderately to highly expressed under four tested conditions. Compared with the accumulation of the hik10 mRNA, it appeared even stronger. A function of transcription was mapped to position 2400 722; #7 and #24 in Table 1) could have an a physiological effect in the salt stress response. The slr0534_as3 transcript may play a silencing role with regard to hik10 under non-inducing conditions. Mutation of its promoter element may hence cause a physiological effect in the salt stress response.

Two other SNPs (at positions 831 647 and 2400 722; #7 and #24 in Table 1) could have an impact on the promoter strength or the regulation of the genes infA and glcP. For glcP, the initiation site of transcription was mapped to position 2400 666 and for infA to position 831 635 (unpublished). Thus, these two SNPs are located 12 and 56 nt upstream of the respective initiation site of transcription. In the case of the infA promoter, the transition replaces a nucleotide within the putative −10 element, changing it from TGTGAT to TATGAT, a much more typical motif for a −10 element in Synechocystis. The mutation 56 nt upstream of the initiation site of transcription of glcP might be functionally relevant as well. The gene product, a glucose transporter, is directly relevant for the physiological ability to use glucose; its gene expression is affected by mutation of the gene for the AbrB-type transcription factor Slr0822. The region at position −56 might well be part of the recognized sequence.

3.4. Larger indels and plasmids

In addition to this relatively large number of SNPs, only seven larger deletions were found on the chromosome and two plasmids. Compared with the reference, a deletion of 0.6 kb exists in the gene slr1753 (#4 in Table 1), which encodes, according to our data, a giant protein comprising 1549 amino acids that probably is transported to the cell surface. However, we found this deletion in our verification also in ‘GT-Kazusa’ and ‘GT-V’. Moreover, the deleted/inserted region consists of long series of DNA repeats (Fig. 1), an evidence for a possible assembly or annotation error in the original sequence analysis.

Given the very scarce available information concerning biological functions of the plasmids in Synechocystis sp. PCC 6803, it was interesting that all seven plasmids were detected during our analysis. Two, pCC5.2 and pCB2.4, were initially not found. However, as they were amplified easily by PCR, we re-inspected the unmapped sequencing reads, but still could not detect a single read matching these plasmids. This observation may relate to a lower copy number of these compared with the other plasmids, but this was not tested in the current study. Analysing the plasmid sequences, we observed a remarkable genetic stability. In addition to a single-base substitution in the plasmid pCA2.4 that might rather constitute an error in the reference sequence (see above) and a missing mobile element on the plasmid pSYSM, two mutations were observed, both in the plasmid pSYSA.

Two major mutations affect the clustered regularly interspaced short palindromic repeats-CRISPR-associated proteins (CRISPR-Cas) system, located on the plasmid pSYSA. CRISPR-Cas systems provide in many archaea and bacteria an adaptive immunity against invading DNA. The plasmid pSYSA encodes the three independent systems CRISPR1, CRISPR2 and CRISPR3. A 2399-bp deletion encompassing the spacer-repeat regions 15–47 of CRISPR1 was detected in ‘PCC-M’ (#43), which also eliminated the relatively short genes ssr7018, ssl7019, ssl7020 and ssl7021, annotated within the spacer-repeat array of CRISPR1. However, the theoretical protein sequences of these gene products show no conservation at all and might not constitute real genes. Nevertheless, the deletion of spacer-repeat regions 15–47 of CRISPR1 is severe, since compared with the reference, it has eliminated two-thirds, 33 of its 49 spacer-repeat units. The sequence analysis suggests that the recombination events leading to the deletion of spacer-repeat regions 15–47 must have occurred within the direct repeats. Thus, this recombination is in agreement with previous observations that the downstream ends of the repeat clusters are conserved such that deletions and recombination events occur internally.

A very different type of deletion was noticed for the CRISPR2 system located on the same plasmid. In this case, 159 bp were deleted (event #44 in Table 1). These 159 deleted bases correspond to positions
71 499–71 657 in the reference. The deletion encompasses two repeats including the spacer 41 in between. It is very surprising that the recombination did not occur within the repeat sections but in the adjacent spacers 40 and 42, thus generating a new ‘hybrid’ spacer 40 at positions 69 082–69 111 in the pSYSA plasmid of ‘PCC-M’ (Fig. 2). As a result, spacers 40, 41 and 42 of the original sequence are missing and became replaced by this hybrid sequence. The vast majority of described deletions in the CRISPR system occur between the direct repeats.\textsuperscript{45} Non-homologous recombination between two different spacers is rare, the deletion observed here in CRISPR2 of the plasmid pSYSA is generating additional sequence diversity in the CRISPR system. Due to the two deletions in the plasmid pSYSA, we determined its total length as 100 749 bp, compared with 103 307 bp for the reference.

3.5. Mobile elements

As can be seen in Tables 1 and 2 (differences #12, 17, 40 and 45), the ‘PCC-M’ substrain lacks four insertion elements of the ISY203 type present in ‘GT-Kazusa’.\textsuperscript{7} These elements are ISY203b, e and g on the chromosome and ISY203j on the plasmid pSYSM. These four indels have the exact same size of 1183 bp, only one is 1185 bp.

In the ‘GT-S’ strain re-sequenced by Tajima \textit{et al}.\textsuperscript{10} one of these four elements, ISY203e, is already present, placing this strain (in accordance with Ikeuchi and Tabata)\textsuperscript{8} before ‘GT-Kazusa’ in the strain history. The absence of ISY203b, e and g in ‘PCC-M’ is further shared with the strains ‘GT-I’, ‘PCC-N’ and ‘PCC-P’,\textsuperscript{9} whereas no statement is possible with regard to the possible presence of ISY203j on the plasmid pSYSM in the latter.

With respect to the described mobile elements, ‘PCC-M’ appears as one of the least-derived substrains.

4. Discussion

4.1. Strain history

‘PCC-M’ shows sequence differences in several genes compared with the reference sequence of ‘GT-Kazusa’ and also to the recently sequenced ‘GT-S’ strain. Kanesaki \textit{et al}.\textsuperscript{9} concluded that 15 differences
between the resequenced strains and the published GT-Kazusa sequence were annotation errors in the latter due to sequencing artefacts, a list to which we add two more putative errors in the database, differences #4 and #42 in Table 1. According to the proposed strain history in Ikeuchi and Tabata, the early division of Synechocystis sp. PCC 6803 into two branches occurred due to an insertion in spkA. Thus, our data suggest that the motile ‘PCC-M’ strain belongs to the motile PCC 6803 branch, whereas the non-motile ‘GT-Kazusa’, ‘GT-S’ and ‘GT-V’ strains are more closely related to each other and belong to the ATCC 27 184 branch. However, the 1-bp insertion in the pilC leading to ‘GT-Kazusa’ as described in the proposed strain history is not present in either ‘GT-S’ or ‘GT-V’, characterizing ‘GT-Kazusa’ as a more derived strain.

That ‘PCC-M’ belongs to the motile PCC 6803 branch is further reinforced by our finding of six SNPs specifically shared between the ‘PCC-M’ and the ‘PCC-N and PCC-P’ substrains (Tables 1 and 2). These six SNPs are in slr1084, slr1983, slr2031, slr0222 and slr0302 (Tables 1 and 2) and must already have been present in the progenitor strain to ‘PCC-M’, ‘PCC-N’ and ‘PCC-P’. Additional support comes from the analysis of two larger indels (#2 and #6 in Table 1). The preceding paper, Kanesaki et al. described difficulties in finding indels between direct repeat sequences such as slr1084 and slr2031 by short read type re-sequencing data. Therefore, these two regions were analysed by PCR and Sanger sequencing in addition to the re-sequencing analysis. Indeed, the finding of indels between direct repeat sequences in genes slr1084 and slr2031 turned out as not been straightforward in our analysis as well. Compared with the reference, we found in both cases the additional sequences of 102 and 154 bp to be present in ‘PCC-M’. This result is relevant for lineage relationships among substrains. The additional 102 bp in gene slr1084 are shared between ‘PCC-M’ and the other substrains ‘PCC-P’, ‘PCC-N’ and ‘GT-I’. Therefore, this must be a deletion in the lineage leading to GT-Kazusa and GT-S. In contrast, the additional 154 bp within and upstream of gene slr2031 are shared between ‘PCC-M’, ‘PCC-P’ and ‘PCC-N’ and are absent from all studied GT substrains. These 154 bp comprise the conserved start codon of slr2031 and extend the gene by 29 codons compared with ‘GT-Kazusa’. Hence, the lack of these 154 bp in GT strains indicate a functionally adverse deletion there. In fact, the 154-bp deletion in GT substrains was noticed before, as well as the activity of slr2031 in the original Synechocystis sp. PCC 6803 substrains. From these considerations, the tree shown in Fig. 3 can be derived. In this tree, ‘GT-Kazusa’ is displayed as the strain with the longest evolutionary distance from the original isolate, whereas the ‘PCC-M’ sub-branch belongs to the ‘PCC’ group of substrains and is probably close to the original characteristics. All strains belonging to the ‘PCC’ group of substrains exhibit twitching motility as was shown also for the original PCC strain deposited in the Pasteur Culture Collection with variations in the motility behaviour. Since ‘PCC-M’ shows motility and is tolerant to glucose, it appears physiologically as a sort of intermediate between the two major branches: the motile and GT branches, consistent with its characterization as being close to the original characteristics.

4.2. Re-sequencing studies of Synechocystis sp. PCC 6803

The analysis of genome sequences of cyanobacteria has had a large impact on photosynthesis, ecology and biotechnology research. The present re-sequencing project delivers the new and complete sequence of the Synechocystis sp. PCC 6803 ‘PCC-M’, a substrain used in many laboratories and in several aspects close to the original isolate. Altogether, there are now chromosomal sequences for seven substrains of Synechocystis sp. PCC 6803 available: ‘PCC-M’ (this study); ‘PCC-P’ (positive phototaxis) and ‘PCC-N’ (negative phototaxis), both based on single colonies isolated from the PCC strain and designated according to their direction of phototactic movement; ‘GT-I’, the standard strain in Dr Ikeuchi’s group; ‘GF’, and ‘GT-S’, a current derivative of the original stock of Synechocystis sp. PCC 6803 from which the chromosomal reference sequence for ‘GT-Kazusa’ was determined in 1996 and for the large plasmids in

Figure 2. Non-homologous recombination in the plasmid pSYSA affecting spacers 40, 41 and 42 of CRISPR2. As a result of the 159-bp deletion in ‘PCC-M’ compared with ‘GT-Kazusa’, a novel hybrid spacer 40 was generated. The direct repeats are presented as squares and the nucleotide positions in the ‘GT-Kazusa’ are given according to the GenBank file NC_005230.
2003, whereas the three small plasmids had been sequenced already before.\cite{37,51,52}

4.3. Mutations potentially linked to phenotype

It is likely that most of the identified differences between the sequenced substrains result from distinct differences in the cultivation conditions in the different laboratories that have selected for fixing one or the other mutation. That also implies that the majority of identified mutations are not silent but linked to a certain effect. Indeed, most mutations in coding regions are not silent as might be expected but lead to frameshifts, amino acid substitutions or the...

Figure 3. Visualization of phylogenetic relationships between various strains of *Synechocystis* sp. PCC 6803. The occurrence of the identified SNPs and other known events are indicated along the branches. The eight events separating the ‘GT’ and ‘PCC’ strains from each other are given at the branch point where these two lineages split or on the respective branches where they occurred. Putative insertions and deletions are labelled ‘Ins’, and ‘Del’, respectively.
truncation of reading frames. Similarly, SNPs in non-coding regions are probably biologically meaningful, too. This idea received support here by linking three ‘PCC-M’-specific SNPs in IGRs to the promoter regions controlling the expression of two protein-coding and one antisense RNA.

For all these reasons, it appears likely that several of the mutations specific to ‘PCC-M’ or shared with ‘PCC-P’ and ‘PCC-N’ may be related to the known phenotypes of these strains. For example, the truncation of *sl1951* (haemolysin) and possible truncation of *slr1753* (surface protein) may contribute to a stress-induced clumping phenotype. Several other mutations might cause alterations in glucose tolerance or phototactic behaviour of these substrains. Differences at other loci may affect the phage resistance, stress response or functions in the primary metabolism, potentially relevant for the synthesis of alkanes or the N and C metabolism. The absence of *isy203g* in the *slr1473-5* regions in PCC substrains leads to an intact photoreceptor that regulates the expression of an alternative phycobilisome linker gene.53 Regarding phenotypic differences among motile PCC substrains, it might be noteworthy that ‘PCC-M’, despite its general ability to be motile, is not phototactic towards blue light (see direct comparison of strains in Fig. 1 of Fiedler et al.48). Here, the SNP #39 in the *sigF* gene, known to be involved in the control of phototactic movement30 might be considered, as the resulting M231K substitution could influence the DNA–protein interaction of this group 3 sigma factor in a very subtle way. For sure, the subtle differences in genome sequences have to be considered when choosing a particular substrain for certain experiments and when comparing phenotypes of mutant lines from different laboratories with the wild-type strain. Information on the re-sequenced genome and plasmid sequences including precisely annotated SNPs can be found in the eight sequence files available from GenBank under the accession numbers CP003265–CP003272.

### Funding

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7-ENERGY-2010-1) under grant agreement no. [256808] and from the German Research Foundation (DFG) project FOR1680 ‘Unravelling the Prokaryotic Immune System’ (grant HE 2544/8-1) to WRH and from grant AL 892/1-4 to SAB.

### References

1. Stanier, R.Y., Kunisawa, R., Mandel, M. and Cohen-Bazire, G. 1971, Purification and properties of unicellular blue-green algae (order Chroococcales), *Bacteriol. Rev.*, **35**, 171–205.
2. Kaneko, T., Sato, S., Kotani, H., et al. 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.
3. Haselkorn, R. 1991, Genetic systems in cyanobacteria, *Methods Enzymol.*, **204**, 418–30.
4. Porter, R.D. 1988, DNA transformation, *Methods Enzymol.*, **167**, 703–12.
5. Williams, J.G.K. 1988, Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803, *Methods Enzymol.*, **167**, 766–78.
6. Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M. and Stanier, R.Y. 1979, Generic assignments, strain histories and properties of pure cultures of cyanobacteria, *J. Gen. Microbiol.*, **111**, 1–61.
7. Okamoto, S., Ikeuchi, M. and Ohmori, M. 1999, Experimental analysis of recently transposed insertion sequences in the cyanobacterium *Synechocystis* sp. PCC 6803, *DNA Res.*, **6**, 265–73.
8. Ikeuchi, M. and Tabata, S. 2001, *Synechocystis* sp. PCC 6803—a useful tool in the study of the genetics of cyanobacteria, *Photosynth. Res.*, **70**, 73–83.
9. Kanesaki, Y., Shioya, Y., Tajima, N., et al. 2012, Identification of substrate-specific mutations by massively parallel whole-genome resequencing of *Synechocystis* sp. PCC 6803, *DNA Res.*, **19**, 67–79.
10. Tajima, N., Sato, S., Maruyama, F., et al. 2011, Genomic structure of the cyanobacterium *Synechocystis* sp. PCC 6803 strain GT-S, *DNA Res.*, **18**, 393–9.
11. Kamei, A., Yuasa, T., Orikawa, K., Geng, X.X. and Ikeuchi, M. 2001, A eukaryotic-type protein kinase, SpkA, is required for normal motility of the unicellular Cyanobacterium *Synechocystis* sp. strain PCC 6803, *J. Bacteriol.*, **183**, 1505–10.
12. McKinlay, J.B. and Harwood, C.S. 2010, Photobiological production of hydrogen gas as a biofuel, *Curr. Opin. Biotechnol.*, **21**, 244–51.
13. Deng, M.D. and Coleman, J.R. 1999, Ethanol synthesis by genetic engineering in cyanobacteria, *Appl. Environ. Microbiol.*, **65**, 523–8.
14. Atsumi, S., Higashide, W. and Liao, J.C. 2009, Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde, *Nat. Biotechnol.*, **27**, 1177–80.
15. Takahama, K., Matsuoka, M., Nagahama, K. and Ogawa, T. 2003, Construction and analysis of a recombinant cyanobacterium expressing a chromosomally inserted gene for an ethylene-forming enzyme at the *psbA1* locus, *J. Biosci. Bioeng.*, **95**, 302–5.
16. Schirmer, A., Rude, M.A., Li, X., Popova, E. and del Cardayre, S.B. 2010, Microbial biosynthesis of alkanes, *Science*, **329**, 559–62.
17. Aoki, R., Takeda, T., Omata, T., Ihara, K. and Fujita, Y. 2012, MarR-type transcriptional regulator ChlR activates expression of tetrapyrrole biosynthesis genes in response to low-oxygen conditions in cyanobacteria, J. Biol. Chem., 287, 13500–7.
18. Chevreux, B., Wetter, T. and Suhai, S. 1999, In: Proceedings of the German Conference on Bioinformatics (GCB). Hannover, Germany, pp. 45–56.
19. Hoffmann, S., Otto, C., Kurtz, S., et al. 2009, Fast mapping of short sequences with mismatches, insertions and deletions using index structures, PLoS Comput. Biol., 5, e1000502.
20. Kaneko, T., Nakamura, Y., Sasamoto, S., et al. 2003, Structural analysis of four large plasmids harboring in a unicellular cyanobacterium, Synechocystis sp. PCC 6803, DNA Res., 10, 221–8.
21. He, Q., Brune, D., Nieman, R. and Vermaas, W. 1998, Chlorophyll alpha synthesis upon interruption and deletion of por coding for the light-dependent NADPH: protochlorophyllide oxidoreductase in a photosystem-I-less/chl^_\_ strain of Synechocystis sp. PCC 6803, Eur. J. Biochem., 253, 161–72.
22. Shen, G. and Vermaas, W.F. 1994, Chlorophyll in a Synechocystis sp. PCC 6803 mutant without photosystem I and photosystem II core complexes. Evidence for peripheral antenna chlorophylls in cyanobacteria, J. Biol. Chem., 269, 13904–10.
23. Sobotka, R., Duhring, U., Komenda, J., et al. 2008, Importance of the cyanobacterial Gun4 protein for chlorophyll metabolism and assembly of photosynthetic complexes, J. Biol. Chem., 283, 25794–802.
24. Yoshihara, S., Geng, X., Okamoto, S., et al. 2001, Mutational analysis of genes involved in pilus structure, motility and transformation competency in the unicellular motile cyanobacterium Synechocystis sp. PCC 6803, Plant Cell Physiol., 42, 63–73.
25. Sakiyama, T., Araie, H., Suzuki, I. and Shiraiwa, Y. 2011, Functions of a hemolysin-like protein in the cyanobacterium Synechocystis sp. PCC 6803, Arch. Microbiol., 193, 565–71.
26. Sakiyama, T., Ueno, H., Homma, H., Numata, O. and Kuwabara, T. 2006, Purification and characterization of a hemolysin-like protein, Sll1951, a nontoxic member of the RTX protein family from the Cyanobacterium Synechocystis sp. strain PCC 6803, J. Bacteriol., 188, 3535–42.
27. Panichkin, V.B., Arakawa-Kobayashi, S., Kanaseki, T., et al. 2006, Serine/threonine protein kinase SpkA in Synechocystis sp. strain PCC 6803 is a regulator of expression of three putative pilA operons, formation of thick pili, and cell motility, J. Bacteriol., 188, 7696–9.
28. Bhaya, D., Vaulot, D., Amin, P., Takahashi, A.W. and Grossman, A.R. 2000, Isolation of regulated genes of the cyanobacterium Synechocystis sp. strain PCC 6803 by differential display, J. Bacteriol., 182, 5692–9.
29. Lane, W.J. and Darst, S.A. 2006, The structural basis for promoter-35 element recognition by the group IV sigma factors, PLoS Biol., 4, e269.
30. Bhaya, D., Watanabe, N., Ogawa, T. and Grossman, A.R. 1999, The role of an alternative sigma factor in motility and pilus formation in the cyanobacterium Synechocystis sp. strain PCC6803, Proc. Natl Acad. Sci. US A, 96, 3188–93.
31. Huckauf, J., Nomura, C., Forchhammer, K. and Hagemann, M. 2000, Stress responses of Synechocystis sp. strain PCC 6803 mutants impaired in genes encoding putative alternative sigma factors, Microbiology, 146, 2877–89.
32. Gao, Q., Wang, W., Zhao, H. and Lu, X. 2012, Effects of fatty acid activation on photosynthetic production of fatty acid-based biofuels in Synechocystis sp. PCC6803, Biotechnol. Biofuels, 5, 17.
33. Mitschke, J., Georg, J., Scholz, I., et al. 2011, An experimentally anchored map of transcriptional start sites in the model cyanobacterium Synechocystis sp. PCC6803, Proc. Natl Acad. Sci. USA, 108, 2124–9.
34. Kobayashi, M., Katoh, H. and Ikeuchi, M. 2006, Mutations in a putative chloride efflux transporter gene suppress the chloride requirement of photosystem II in the cytochrome c550-deficient mutant, Plant Cell Physiol., 47, 799–804.
35. Shoumskaya, M.A., Paithoonrangsarid, K., Kanesaki, Y., et al. 2005, Identical Hik-Rre systems are involved in perception and transduction of salt signals and hyperosmotic signals but regulate the expression of individual genes to different extents in Synechocystis, J. Biol. Chem., 280, 21531–8.
36. Ishii, A. and Hihara, Y. 2008, An AbrB-like transcriptional regulator, Sll0822, is essential for the activation of nitrogen-regulated genes in Synechocystis sp. PCC 6803, Plant Physiol., 148, 660–70.
37. Yang, X. and McFadden, B.A. 1993, A small plasmid, pCA2.4, from the cyanobacterium Synechocystis sp. strain PCC 6803 encodes a rep protein and replicates by a rolling circle mechanism, J. Bacteriol., 175, 3981–91.
38. Al-Attar, S., Westra, E.R., van der Oost, J. and Brouns, S.J. 2011, Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes, Biol. Chem., 392, 277–89.
39. Deveau, H., Garneau, J.E. and Moineau, S. 2010, CRISPR/Cas system and its role in phage-bacteria interactions, Annu. Rev. Microbiol., 64, 475–93.
40. Horvath, P. and Barrangou, R. 2010, CRISPR/Cas, the immune system of bacteria and archaea, Science, 327, 167–70.
41. Karginov, F.V. and Hannon, G.J. 2010, The CRISPR system: small RNA-guided defense in bacteria and archaea, Mol. Cell., 37, 7–19.
42. Marraffini, L.A. and Sontheimer, E.J. 2010, Self versus non-self discrimination during CRISPR RNA-directed immunity, Nature, 463, 568–71.
43. Terns, M.P. and Terns, R.M. 2011, CRISPR-based adaptive immune systems, Curr. Opin. Microbiol., 14, 321–27.
44. Wiedenheft, B., Sternberg, S.H. and Doudna, J.A. 2012, RNA-guided genetic silencing systems in bacteria and archaea, Nature, 482, 331–8.
45. Liljestol, R.K., Shah, S.A., Brugger, K., et al. 2009, CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties, Mol. Microbiol., 72, 259–72.
46. Katoh, A., Sonoda, M. and Ogawa, T. 1995, A possible role of 154-base pair nucleotides located upstream of ORF440 on CO2 transport of *Synechocystis* sp. PCC 6803. In: Mathis, P. (ed.), *Photosynthesis: From Light to Biosphere*. Kluwer Academic Publishers: Dordrecht, The Netherlands, pp. 481–4.

47. Kamei, A., Ogawa, T. and Ikeuchi, M. 1998, Identification of a novel gene (*slr2031*) involved in high-light resistance in the cyanobacterium *Synechocystis* sp. PCC 6803. In: Garab, G. (ed.), *Photosynthesis: Mechanism and Effects*. Kluwer Academic Publishers: Dordrecht, The Netherlands, pp. 2901–5.

48. Fiedler, B., Börner, T. and Wilde, A. 2005, Phototaxis in the cyanobacterium *Synechocystis* sp. PCC 6803: role of different photoreceptors, *Photochem. Photobiol.*, **81**, 1481–8.

49. Narikawa, R., Suzuki, F., Yoshihara, S., et al. 2011, Novel photosensory two-component system (PixA-NixB-NixC) involved in the regulation of positive and negative phototaxis of cyanobacterium *Synechocystis* sp. PCC 6803, *Plant Cell Physiol.*, **52**, 2214–24.

50. Hess, W.R. 2011, Cyanobacterial genomics for ecology and biotechnology, *Curr. Opin. Microbiol.*, **14**, 608–14.

51. Xu, W. and McFadden, B.A. 1997, Sequence analysis of plasmid pCC5.2 from cyanobacterium *Synechocystis* PCC 6803 that replicates by a rolling circle mechanism, *Plasmid*, **37**, 95–104.

52. Yang, X. and McFadden, B.A. 1994, The complete DNA sequence and replication analysis of the plasmid pCB2.4 from the cyanobacterium *Synechocystis* PCC 6803, *Plasmid*, **31**, 131–7.

53. Hirose, Y., Shimada, T., Narikawa, R., Katayama, M. and Ikeuchi, M. 2008, Cyanobacteriochrome CcaS is the green light receptor that induces the expression of phycobilisome linker protein, *Proc. Natl Acad. Sci. USA*, **105**, 9528–33.