Published in final edited form as:

Plasmid. 2008 May ; 59(3): 176–192. doi:10.1016/j.plasmid.2008.01.005.

The complete sequence and functional analysis of pANL, the large plasmid of the unicellular freshwater cyanobacterium *Synechococcus elongatus* PCC 7942

You Chen^a, C. Kay Holtman^a, Roy D. Magnuson^b, Philip A. Youderian^a, and Susan S. Golden^{a,*}

^a Department of Biology, Texas A&M University, College Station, TX 77843

^b Department of Biological Sciences, University of Alabama in Huntsville, Huntsville, AL 35899

Abstract

Two endogenous plasmids are present in *Synechococcus elongatus* PCC 7942, a model organism for studying photosynthesis and circadian rhythms in cyanobacteria. The large plasmid, pANL, was shown previously to be involved in adaptation of *S. elongatus* cells to sulfur starvation, which provided the first evidence of cellular function of a cyanobacterial plasmid. Here, we report the complete sequence of pANL, which is 46,366 bp in length with 53% GC content and encodes 58 putative ORFs. The pANL plasmid can be divided into four structural and functional regions: the replication origin region, a signal transduction region, a plasmid maintenance region, and a sulfur-regulated region. Cosmid-based deletion analysis suggested that the plasmid maintenance and replication origin regions are required for persistence of pANL in the cells. Transposon-mediated mutagenesis and complementation-based pANL segregation assays confirmed that two predicted toxin-antitoxin cassettes encoded in the plasmid maintenance region, belonging to PemK and VapC families, respectively, are necessary for plasmid exclusion. The compact and efficient organization of sulfur-related genes on pANL may provide selective advantages in environments with limited sulfur.

Keywords

Synechococcus elongatus PCC 7942; cyanobacterial plasmid; toxin-antitoxin cassette; plasmid maintenance

1. Introduction

As a group of photoautotrophic eubacteria, cyanobacteria are characterized by oxygenic photosynthesis, morphological diversity, and their blue-green color. Like most other prokaryotes, cyanobacteria generally have a single circular chromosome. In addition, many cyanobacterial strains also contain one or more endogenous plasmids (Bose and Carmichael, 1990; Felkner and Barnum, 1988; Rebière et al., 1986). The complete sequences of about 30

* Corresponding Author: 3258 TAMU, College Station, TX 77843-3258. Tel: 979-845-9824, Fax: 979-845-2891, Email: sgolden@tamu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

cyanobacterial plasmids have been deposited into the GenBank database as of December 2007. They belong to strains of six cyanobacterial genera: *Anabaena*, *Leptolyngbya*, *Microcystis*, *Nostoc*, *Synechococcus*, and *Synechocystis*, with a size range from around 1.5 kb (pPBS1, GenBank accession no. **AF176225**) to over 400 kb (pCC7120alpha, GenBank accession no. **BA000020**). There are 7 plasmids in *Synechocystis* sp. PCC 6803 (Kaneko et al., 2003; Xu and McFadden, 1997; Yang and McFadden, 1993; Yang and McFadden, 1994) and 6 in *Nostoc* (also known as *Anabaena*) sp. PCC 7120 (Kaneko et al., 2001). One of the species with no plasmid detected is the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1 (Nakamura et al., 2002). Many cyanobacterial cellular functions have long been speculated to be determined by plasmids, such as resistance to antibiotics and heavy-metals, toxin production, and gas vacuolation (Lau and Doolittle, 1979; Lau et al., 1980). However, since most of the cyanobacterial plasmids are still cryptic, none of these biological roles has been experimentally verified.

The freshwater unicellular cyanobacterium *Synechococcus elongatus* PCC 7942, formerly *Anacystis nidulans* R2, carries two endogenous plasmids, pANS (also called pUH24) and pANL (also called pUH25) (van den Hondel et al., 1980; Van der Plas et al., 1992). These two plasmids have been identified not only in the closely related strain *S. elongatus* PCC 6301, but also in *Cyanobium* (*Synechococcus*) sp. PCC 6707, which is different from *S. elongatus* PCC 7942 in both genome size and base composition (Stanier et al., 1971; van den Hondel et al., 1979). Although the spontaneous loss (curing) of pANS from *S. elongatus* PCC 7942 has been reported (Lau and Doolittle, 1979), attempts to cure pANL were not successful (Laudenbach et al., 1985).

The smaller plasmid, pANS (GenBank accession no. **S89470**), is 7,835 bp long and encodes 8 putative open reading frames (ORFs) (Van der Plas et al., 1992). Because pANS can be easily cured and has a relatively small size, it has been used to construct vectors to shuttle genes between *E. coli* and *Synechococcus* strains (Golden and Sherman, 1983; Koksharova and Wolk, 2002; Sherman and van de Putte, 1982). The large plasmid, pANL, has also been the subject of investigation and it has never been cured from the cell. Laudenbach and colleagues constructed a physical map of restriction sites for pANL and identified a likely origin of replication (Laudenbach et al., 1983; Laudenbach et al., 1985). Three *Bam*HI fragments of pANL were then sequenced (GenBank accession nos. **U20224**, **U23436**, and **AF176824**; 3.4 kb, 3.8 kb, and 11.2 kb, respectively). A number of genes on these fragments are functionally related to sulfur metabolism in *S. elongatus* PCC 7942, and thus were designated as sulfur-regulated plasmid-encoded (srp) genes (Nicholson et al., 1995; Nicholson and Laudenbach, 1995). These reports provided the first evidence of physiological function of a cyanobacterial plasmid. We completed the pANL sequence (GenBank accession no. **AF441790**) based on a combination of cosmid end sequences and transposon-mediated sequences ((Holtman et al., 2005) and this report).

Here we report sequence and functional analysis of pANL: manual annotation of the putative ORFs by protein-coding region prediction and gene function assignment, identification *via* cosmid-based deletion mapping of regions that are important for plasmid maintenance, and verification of the function of two toxin-antitoxin cassettes using transposon-mediated mutagenesis and complementation-based segregation assay. Among our findings, a previously published ~5 kb sequence fragment (GenBank accession no. **U70379**) that carries a putative *narL* gene and genes encoding a two-component system, which was originally assumed to be on the chromosome, is in fact part of pANL. Four modules can be assigned in pANL based on gene organization and functional prediction: the replication origin region, a signal transduction region, a plasmid maintenance region, and a sulfur-regulated region. Our results show that the difficulty to cure pANL is mainly due to two toxin-antitoxin cassettes in the plasmid maintenance region.

2. Materials and methods

2.1. Cyanobacterial strains, media, and culture conditions

All cyanobacterial wild-type and mutant strains were created in *Synechococcus elongatus* PCC 7942. Cyanobacterial strains were grown in BG-11 medium (Rippka et al., 1979) under continuous light conditions ($\sim 50\text{--}70 \mu\text{E}/\text{m}^2\text{s}$) at 30°C with appropriate antibiotics. The transformation of *S. elongatus* cells was conducted as previously described (Holtman et al., 2005). The cyanobacterial strains used in this study are summarized in Table 1.

2.2. Plasmid construction

Unless otherwise stated, plasmids were constructed in *Escherichia coli* strain DH10B (Invitrogen, Carlsbad, CA). All plasmids used in this study are described in Table 1 and primers in Table 2. pAM3558 is a gentamycin-resistance version of pAM1303, an NS1 cloning vector (Andersson et al., 2000). Neutral site vectors mediate homologous recombination at defined loci in the *S. elongatus* chromosome and cause no apparent phenotypes. To make pAM3558, the *aadA* gene (Sp^RSm^R) in the Ω cassette of pAM1303 was replaced by the *accC1* gene (Gm^R) from pAM3511. The *sepA1T1* operon or *sepA1* alone was amplified using primer pairs pem5 $\times$ pemk3 or pem5 $\times$ pemi3, respectively. The PCR products were then inserted into pAM1303 to form pAM3716 and pAM3715, respectively. PCR products of the *sepA2T2* operon or *sepA2* alone, which were amplified with primer pairs add2–5 $\times$ add2–34 or add2–5 $\times$ add2–33, were inserted into pAM3558 to construct pAM3617 and pAM3618, respectively.

2.3. Sequence determination and annotation

A 960-cosmid genomic library provided the templates for transposon-mediated mutagenesis and sequencing of the genome of *S. elongatus* PCC 7942 (Holtman et al., 2005). Alignment of end sequences of these cosmids with published *S. elongatus* PCC 7942 sequences in GenBank showed that many of them were located on pANL, the large plasmid, instead of the chromosome, which means that pANL fragments were cloned into the cosmid vector at a high frequency during library construction. A DNA hybridization assay (Materials and Methods 2.5) revealed that around one third of the cosmids in the library ($\sim 32\%$, 307 cosmids) carry pANL sequence. Based on the analysis of these cosmid end sequences, one cosmid of pANL-origin, 2A8, was predicted to cover most of the gaps in the assembly of cosmid-end sequences, and was chosen as a template for *in vitro* mutagenesis and sequencing to complete the pANL sequence (Holtman et al., 2005). The remaining gaps in the pANL assembly were then filled by primer walking. Primers used in primer walking assays are underlined in Table 2.

The individual sequences were arranged and assembled using the Vector NTI software package (Invitrogen, Carlsbad, CA). Protein-coding regions were predicted with FramePlot 2.3.2 (Ishikawa and Hotta, 1999). Functional assignment of putative ORFs was conducted by similarity search (BLAST) in protein and conserved domain databases of GenBank (Altschul et al., 1990; Marchler-Bauer et al., 2005). Later, the pANL complete sequence was refined with DOE Joint Genome Institute (JGI) shotgun sequences (Holtman et al., 2005) as a reference and subsequently updated (GenBank accession no. [AF441790](#)).

2.4. Boiling method for preparation of cyanobacterial PCR templates

Samples (1–2 ml) of *S. elongatus* PCC 7942 cells (OD 0.6–0.9 at 750 nm) were collected by centrifugation at 16,000 g for 1 min and then resuspended in 150–200 μl TE (Tris-EDTA, pH 8.0) (Maniatis et al., 1982). The resuspended cells were boiled for 5 min, pelleted at 16,000 g for 15 min. The supernatant fractions were then transferred to a new 1.5 ml tube. For a 50 μl PCR reaction, 2–5 μl boiled samples ($\sim 5\text{--}10 \text{ ng}$ total DNA/ μl) were used.

2.5. DNA Hybridization analysis

Cosmid 6G1 DNA (~200 ng/μl) was digested with *EcoRI*. Two large digestion fragments (~12 kb and ~16 kb), gel-purified using a DNA purification kit (BIO-RAD, Hercules, CA), were labeled with ³²P-dCTP as probes to hybridize 10 pieces of nitrocellulose membranes (courtesy of Dr. C. Thomas), each blotted with DNAs from one of the ten 96-well plates of the cosmid library. Membranes were incubated first in prehybridization solution (5× SSPE (Maniatis et al., 1982), 1% SDS, 0.1 mg/ml SSDNA) at 65°C for 4 h, and then in hybridization solution (5 × SSPE, 1% SDS) at 60°C overnight after addition of boiled probes. Next, membranes were rinsed (0.5 × SSPE, 0.2% SDS) 3 times for 3 min each at 60°C, and then washed in the same solution 3 times for 30 min each at 62–65°C. Finally, air-dried membranes were wrapped and exposed to x-ray film at –80°C overnight.

3. Results and discussion

3.1. General features of pANL

The complete sequence of pANL is 46,366 bp in length with an ~53% GC content, lower than that of the chromosome (~55%, GenBank accession no. **CP000100**) and pANS (~59%). There are 58 putative ORFs predicted on pANL (Table 3). On average, there are ~1.25 coding regions per kb of pANL sequence. The coding regions, ~54% in GC content, cover ~83% of the total sequence. The GC content in the intergenic regions is ~47%. As shown in Table 3, 27 ORFs are assigned with functions based on their significant hits in BLAST searches for conserved domains and close homologues; 26 ORFs encode conserved hypothetical proteins with homologues found; 2 ORFs encode hypothetical small proteins (approximately 100 amino acid residues) without any significant hits in the protein database; additionally, there are three other small ORFs, *anL09*, *anL26*, and *anL27*, that share similarity only with each other in both nucleotide and protein sequences, but with nothing else in the database. Because the GC content of these three ORFs is close to that of the other coding sequences, they are probably pANL-specific loci. Most of the proteins encoded by pANL have cyanobacterial homologues. There are 23 ORFs that have significant paralogues on the chromosome of *S. elongatus* PCC 7942 (Table 3). Among those, both homologues of *anL03*, a conserved hypothetical protein, are *S. elongatus* PCC 7942 chromosomal genes. On the contrary, *anL02*, *anL25*, *anL51*, and *anL58* did not match any cyanobacterial proteins in a BLAST search. The GC content of two of them, *anL02* and *anL58* (36% and 38%, respectively), is much lower than that of the other ORFs or of the chromosome, suggesting that they may have arisen by lateral transfer from an exogenous source.

3.2. Repetitive sequences

An octameric highly iterated palindrome (HIP1), GCGATCGC, is over-represented in DNA sequences of many cyanobacterial strains, including *Synechococcus* species (Gupta et al., 1993). The chromosome of *S. elongatus* PCC 7942 has 7,402 HIP1 sites, around 1 site per 364 bp. There are 95 HIP1 sites found in pANL, averaging 1 site every 488 bp, a lower frequency than that in both the chromosome and pANS (1 site/~412 bp). The distribution of HIP1 sites around pANL is roughly random, with a few hot spots and blank regions. One interesting finding is that the longest HIP1 gap (coordinates 43193–274, ~3.5 kb) corresponds to the putative replication origin. It has been suggested that HIP1 sites participate in homologous recombination and other sequence rearrangement events in cyanobacterial genomes (Gupta et al., 1993; Robinson et al., 1997). The absence of HIP1 sites in this region supports its importance in maintenance of pANL.

Self-alignment of pANL sequence was performed with the ‘BLAST 2 SEQUENCES’ tool (bl2seq) to identify repetitive sequences (Tatusova and Madden, 1999). Five pairs of relatively large direct or inverse repeat sequences were identified and all of them are located in putative

coding regions (Table 4). Based on location and structure, all five pairs of repetitive sequences may be related to mobile elements and can serve as structural separators for different regions of pANL. The details of these repeat sequence pairs will be discussed along with the putative ORFs where they reside.

3.3. Four functional regions

Plasmid pANL can be divided into four modules by separation of repetitive sequences and the functional assignments of putative genes: the replication origin region; a signal transduction region; a plasmid maintenance region; and a sulfur-regulated region.

3.3.1. The replication origin region—This region starts from *anL55.5* (Fig. 1A) and ends at *anL12* (Fig. 2), including 18 ORFs and several pairs of repeat sequences. The replication origin of pANL has been narrowed down to an ~11.7 kb *Bam*HI fragment in this region (Laudenbach et al., 1983). Even though two replication origins in two adjacent *Kpn*I fragments in this part of pANL were proposed (Laudenbach et al., 1985), our experimental data (Table 5) suggest that there is only one replication origin (coordinates 42913–44307, ~1.4 kb) located in the *Kpn*I-*Bam*HI fragment that is closest to the first nucleotide of the pANL sequence accession (Fig. 1A).

At the borders of this region, there are two pairs of overlapping genes. One pair, *anL56* and *anL56.5* (Figs. 1A and 2), is very similar to the other pair, *anL11* and *anL12* (Fig. 2), in their size and structure, as well as encoded protein sequences. The putative products of ORF *anL56.5* and *anL12* are ~45% identical and both carry a copy of the COG3514 conserved domain (no function characterized). A pair of 26 bp direct repeat sequences (RSP-4, Table-4) lies in their coding regions. The similarity between *anL56* and *anL11* is too low to be detected at the protein sequence level. However, they both contain a COG2929 domain conserved in bacteria. These two pairs of overlapping ORFs are probably the consequence of either a gene duplication or transposition event. On either side of *anL56* and *anL56.5* are two small paralogs, *anL55.5* and *anL57*, sharing about 80% protein sequence identity, and a pair of 149 bp direct repeats (RSP-2, Table-4). Among the other ORFs, *anL01* predicts a conserved hypothetical protein product (999 aa) that carries a likely MobB (COG1763) domain, which is involved in molybdopterin-guanine dinucleotide biosynthesis. ANL01 was speculated to be a RepA (replication initiator protein) homologue. However, no significant sequence similarity between ANL01 and RepA of other bacteria can be detected. As mentioned above, both *anL58* and *anL02* have low GC content (38% and 36%, respectively), and have no cyanobacterial homologues, but many significant hits in other bacteria, and might have been transferred from other bacteria not very long ago. A small protein encoded by *anL09* (105 aa) shows high similarity only to *anL26* and *anL27* (48% and 60%, respectively), and to no other proteins in the database. These three ORFs possibly resulted from successive duplication and inversion events as they carry both direct and inverted repeats (RSP-3&5, Table 4). The *anL10* locus encodes a protein almost identical (95% identity) to the C-terminal domain of the *anL28* protein, predicted to be a site-specific recombinase. The largest pair of direct repeats in pANL is located in these two ORFs (RSP-1, Table 4). These repetitive sequences are likely recombination sites recognized by site-specific recombinases.

3.3.2. The signal transduction region—This region starts with *anL14*, which encodes a hypothetical protein (Fig. 2). ORFs *anL26* and *anL27*, which are closely related to *anL09*, are at the other end. The most striking feature in this region is the presence of genes that encode a two-component signal transduction system (Stock et al., 2000). ORF *anL22* and its upstream overlapping *anL23* encode a histidine protein kinase (HPK) and its cognate response regulator (RR), respectively. The HPK encoded by *anL22*, 439 aa in length, contains a C-terminal histidine protein kinase domain with the conserved phosphoacceptor histidine residue. No

identifiable domain is present at its N-terminal part except two transmembrane regions. The RR protein (225 aa) AnL23 carries a CheY-like receiver (REC) domain with the conserved aspartic acid residue and a C-terminal winged-helix DNA-binding domain. Three other two-component systems have been identified on the plasmids of *Synechocystis* sp. PCC 6803 (Kaneko et al., 2003), and they were shown not to be involved in the perception of hyperosmotic stress (Paithoonrangsarid et al., 2004). The real function of these plasmid-born two-component systems is yet to be explored.

Other proteins encoded in this region may also participate in signal transduction and regulation. The *anL15* locus encodes a protein carrying a predicted transcriptional regulator domain. The protein encoded by ORF *anL16* looks like an NADPH-specific quinone reductase. There is a Type 1 glutamine amidotransferase homologue encoded by *anL18*. Putative ORF *anL19* encodes a likely one-component system protein (Ulrich et al., 2005), carrying an N-terminal GAF domain (no conserved cysteine for chromophore binding) and a C-terminal HTH_LUXR domain, which probably binds DNA directly. The ORF *anL25* encodes a protein that may also bind DNA through its C-terminal HTH_DEOR domain.

3.3.3. The plasmid maintenance region—As with many other low-copy number plasmids, pANL carries a full set of genes (Fig. 1B) necessary for maintaining its stable inheritance in cells. These genes encode: 1) a site specific recombinase for resolution of multimers, ANL28 (XerC homologue); 2) a partitioning system for equal distribution of plasmids into daughter cells (ANL31 and ANL32, ParB and ParA homologues, respectively); 3) two toxin-antitoxin cassettes for post-segregational killing of pANL-cured cells (ANL29 and ANL29.5, a PemK family toxin and its cognate antitoxin; ANL33 and ANL34, a VapC-like toxin-antitoxin system). These cassettes are designated *sepA1/T1* (*anL29.5/anL29*, PemK-like system) and *sepA2/T2* (*anL33/anL34*, VapC-like system) for *Synechococcus elongatus* plasmid-born antitoxin or toxin genes (Table 3 & Fig. 1B).

No significant conserved domains were identified for putative antitoxins SepA1 and SepA2. Their functional assignment is mainly based on their gene context and conservation among other cyanobacteria. Both of these ORFs not only are located upstream of, but also overlap with, their cognate toxin genes, as seen in other toxin-antitoxin cassettes. The protein sequence of SepA1 (ANL29.5, 71 aa) shows ~48% identity to N9414_18895, a 79 aa small protein in *Nodularia spumigena* CCY9414. The gene downstream of N9414_18895, N9414_18900, also encodes a PemK-like protein. Thus SepA1 and N9414_18895 are very likely antitoxins, even though they are both featureless hypothetical proteins. The first 28 aa of SepA1 also shares ~64% identity with that of Ssl2420 in *Synechocystis* sp. PCC 6803, an 84 aa small hypothetical protein. Sll1225, encoded downstream of *ssl2420* and carrying a PIN domain (COG4113), is possibly a toxin protein. It is interesting to see that antitoxin proteins from two different toxin families are very similar in their N-terminal parts. Thus, a composite and modular mechanism, which has been described from a toxin-antitoxin system on *Bacillus thuringiensis* plasmid pG11 (Fico and Mahillon, 2006), is also adopted for cyanobacterial toxin-antitoxin systems, in which similar antitoxins can be assembled with different toxins. There is no significant match in the GenBank database for SepA2 (ANL33), the putative antitoxin for the SepT2 (ANL34) protein. The closest homologues of SepT2, the putative VapC family toxin, are RS9917_04395 from *Synechococcus* sp. RS9917 (37% identity) and SYNW0214 from *Synechococcus* sp. WH 8102 (36% identity). Upstream of each of these two toxins is encoded a small ribbon-helix-helix protein. However, they are not homologous of SepA2 despite their similar size and gene context. The function of SepA1 or SepA2 as antitoxin was then confirmed by functional analysis (section 3.4).

The last gene in this region, pANL30, maps between the two putative operons of *pemIK* and *parAB*. The encoded protein contains a nucleotidyl transferase domain (pfam01909) and may

add a nucleotidyl group onto an unknown substrate, perhaps functioning in antibiotic resistance.

3.3.4. The sulfur-regulated region—Adjacent to the plasmid maintenance region is the sulfur metabolism related region, including 13 previously designated *srp* genes (*srpA*→*srpM*) (Nicholson et al., 1995; Nicholson and Laudenbach, 1995)(Fig. 2). This region can be divided into two gene clusters that are in opposite orientations. Within the clusters, all genes are expressed in the same direction. The counter-clockwise cluster comprises ORFs from *anL35* to *anL40*. The clockwise cluster includes ORFs from *anL44* to *anL55*. The two clusters are separated by two head-to-head genes: *anL42* and *anL43*, both of which encode conserved hypothetical proteins (Table 3).

In the counter-clockwise cluster, a rhodanese-related sulfurtransferase domain (COG0607) is present in the ANL35 protein sequence. Its downstream gene *anL36* encodes a cysteine desulfurase domain (COG1104). Proteins encoded by *anL40* and *anL39*, SrpG and SrpH, correspond to enzymes for catalyzing two consecutive steps in the cysteine biosynthesis pathway, O-acetylserine (thiol) lyase and serine acetyltransferase, respectively (Nicholson et al., 1995).

Among the proteins encoded in the clockwise cluster are a putative transcriptional regulator ANL44 (COG1959), a putative carbonic anhydrase ANL45 (cd03379), a 36 kD putative catalase (ANL46/SrpA) that is induced in *S. elongatus* PCC 7942 cells under sulfur stress, and a likely chromate transporter (ANL48/SrpC) (Nicholson and Laudenbach, 1995). ORF *anL47* (*srpB*) carries an MgtC domain with no known function (pfam02308). There is another O-acetylserine (thiol) lyase homologue encoded by *anL49* (*srpD*). SrpE (ANL50) is a predicted gamma-Glutamyltranspeptidase (pfam01019), a key enzyme in glutathione metabolism. SrpJ (ANL52) is a homologue of acyl-CoA dehydrogenase (COG1960). The downstream three ORFs that are likely in an operon, *srpKLM*, encode subunits of putative TauABC complex, an ABC-type nitrate/sulfonate/bicarbonate transport system.

Most of the genes in this region either have been shown or are predicted to be involved in sulfur metabolism or transportation. Many of them have close paralogs in the chromosome (Table 3), where they probably originated.

3.4. Functional analysis of pANL maintenance genes

To experimentally define the pANL regions required for plasmid maintenance, cosmid-based deletion analysis was performed. The identified regions were then scrutinized *via* transposon-mediated mutagenesis of individual genes.

3.4.1. Cosmid-based deletion analysis of pANL—Deletion mapping was conducted using pANL-bearing cosmids (Table 5 & Fig. 2). Many cosmids in our collection carry a segment of pANL sequence that was randomly produced during *Sau3AI* partial digestion in the process of library construction (Holtman et al., 2005). If the pANL sequence on a cosmid includes all essential parts for pANL maintenance, the cosmid should be able to completely replace pANL in the cyanobacterial cell under constant selection for antibiotic-resistance markers on the cosmid vector. If some components required for plasmid maintenance are present, both pANL and the cosmid would co-exist in the cells even under strong selection, and would probably form recombinant variants of them (Laudenbach et al., 1985). The composition of the pANL/cosmid pool in the cyanobacterial cells can be detected by PCR amplification using primer pairs specific for the region of pANL that is lost in the cosmid or for the cosmid vector.

With both ends checked by sequencing, 27 cosmids missing different parts of pANL were chosen for deletion analysis (Table 5 & Fig. 2). Transformation and selection of these cosmids in *S. elongatus* wild-type strain AMC06 were done in 24-well microplates. Total DNA was then extracted using a boiling method (Materials and methods 2.4) directly from *S. elongatus* transformants (not plated for single colonies) after ~100 generations of selection with kanamycin (> 3 months). PCR amplification using primer pairs specific for the cosmid vector SuperCos I confirmed the presence of cosmids in all transformed strains (Table 5). No transformants could be obtained when *S. elongatus* cells were transformed with the SuperCos I vector only that carries no pANL sequence. Then, for each cosmid-transformed strain, primer pair(s) specific for the region shared by pANL and the transformed cosmid, as well as primer pair(s) specific for the region of pANL absent from the cosmid, were used to detect the segregation of pANL in the cells.

As summarized in Table 5 & Fig. 2, strains transformed with some cosmids were completely missing the region of pANL that is absent from the introduced cosmid, which indicates that no endogenous pANL is present in these cells and the region is not required for plasmid maintenance. Strains transformed with other cosmids, however, still maintained the region that is missing from the cosmid; that is, pANL could not be fully segregated out and the region is necessary for plasmid maintenance. When all of the regions not required for plasmid maintenance were lined up in the sequence order, four gaps that are candidates for regions required for plasmid maintenance appeared (Fig. 2 & Table 5). Gap-2 (coordinates 33455–33890) is located in *anL46* (putative catalase), while gap-3 (35060–35811) lies in *anL48* (putative chromate transporter). Both cosmids 4H2 and 2B4, which lack the N-terminal coding region of *anL46* and *anL48*, respectively, completely replaced endogenous pANL when introduced into wild-type *S. elongatus* cells (Table 5 & Fig. 2). Thus, function of the protein products of the ORFs in which the two gaps reside is not likely to be involved in plasmid maintenance. Furthermore, cosmids that lack gaps 2 or 3 also lack gaps 1 or 4. Although we cannot strictly exclude a role for gaps 2 and 3 in plasmid maintenance, the defects in these cosmids are more parsimoniously assigned to gaps 1 and 4. Gap-1 (coordinates 22171–23818) includes ORFs *anL29.5*, *anL30*, *anL31*, and *anL32*; gap-4 (coordinates 42913–44307) contains ORFs *anL55.5*, *anL56*, and *anL56.5* (Fig. 2 & Table 5). Gap-1 corresponds to the plasmid maintenance region, while gap-4 is located in the replication origin region. The deletion analysis results obtained from cyanobacterial strains that were transformed with cosmids 2B10 or 4H2, both missing the C-terminal part of *sepA2*, all of *sepT2*, and some other downstream ORFs, suggest the putative antitoxin gene *sepA2* (*anL33*) was not required for plasmid maintenance (Table 5). Conversely, the endogenous pANL could not be segregated in the strain transformed with cosmid 4F10, which is missing the *sepA2T2* operon and part of the *para* (*anL32*) ORF. Transposon-mediated site-specific mutagenesis assay (3.4.2, below) clearly showed that it is *sepA2*, not *para*, that is required for plasmid maintenance. The segregation of pANL in strains transformed by the cosmids 2B10 and 4H2 suggests that a recombination event not detected by PCR amplification assays in the cosmid-based deletion assay may have occurred in those cases, yielding a false segregation result.

Other sequences, including the signal transduction and the sulfur-regulated regions, do not contribute to pANL maintenance. Functional redundancy of many of these genes with paralogues on the chromosome may render them dispensable (Table 3). However, the results also suggest that these genes provide significant advantages for cells in adaptation to sulfur starvation, because all known plasmid maintenance mechanisms are present to avoid the loss of pANL.

3.4.2. Transposon-mediated mutagenesis of plasmid maintenance genes—To further investigate the requirement of individual genes in the plasmid maintenance region, *Mu*Cm transposons that have inserted into pANL cosmids were used for mutagenesis of these

genes *in vivo*. As *S. elongatus* PCC 7942 favors double homologous recombination, once transformed and selected by chloramphenicol (Cm), the transposons will be integrated into their corresponding loci on pANL and disrupt the function of the putative ORFs, if applicable. The cosmid itself will be lost because no selection pressure for the cosmid is exerted. If the disrupted pANL gene is not necessary for exclusion, then native pANL with no insertion will be lost eventually. PCR amplification of the insertion sites was used to detect the presence of intact pANL in transposon mutant strains. *Mu*Cm transposons located on pANL cosmids 2A8, 6C4, and 5H4 were introduced into wild-type *S. elongatus* strain AMC06 (Fig. 1B). The coordinates of these insertions are also listed in the GenBank entry of pANL (GenBank accession No. [AF441790](#)). Transformation of *S. elongatus* cells with *Mu* insertions (2A8E9 and 5H4D7 in Fig. 1B) located in *sepA2* (*anL33*), encoding antitoxin for the VapC-like toxin SepT2, resulted in either no colony or tiny colonies with very slow growth, consistent with an antitoxin function. Transformants with *Mus* in other ORFs were normal in growth. Primer pairs flanking the insertion loci were used in PCR amplification from boiled cyanobacterial samples. In the wild-type strain the PCR product is usually smaller than 1 kb as designed. In insertional mutant strains, the expected PCR product is > 2 kb due to the insertion of mini*Mu*Cm (~1.4 kb). Probably due to the low quality of DNA, PCR amplification with boiled samples could not always produce products bigger than 1.5 kb. Thus, sometimes no signal can be seen on the agarose gel if all pANL molecules in the cells carry the transposon insertion. Otherwise, a wild-type band should be detected.

As shown in Fig. 3, no native intact pANL could be detected in *Mu* insertional strains carrying: 5H4D2 (*anL28*, *xerC*, encoding a putative site-specific recombinase), 5H4H11 and 5H4B2 (both in *anL29*, a *pemK* homologue), 5H4A2 (*anL30*, encoding a putative nucleotidyl-transferase), 6C4Q4 (*anL31*, a *parB* homologue), 5H4C12 and 5H4E8 (both in *anL32*, a *parA* homologue), as well as 5H4A10 (*anL34*, encoding a VapC-like toxin). The results suggest that these genes are not required for maintaining pANL in cyanobacteria cells. The *sepA1* (*anL29.5*) gene, encoding the antidote for PemK, is necessary for plasmid exclusion as indicated by the presence of a PCR product from un-segregated pANL in strains transformed with the respective mini*Mu*Cm 2A8D4. The cosmid-based deletion analysis and the specific gene inactivation results support two toxin-antitoxin cassettes on pANL as previously predicted by bioinformatics analysis.

3.4.3. Both toxin-antitoxin cassettes are necessary for exclusion—Toxin-antitoxin cassettes prevent curing of plasmids by a post-segregational killing mechanism which is based on the fact that toxins are much more stable than antitoxins (Hayes, 2003). These genes are widely distributed in plasmids and chromosomes of many free-living bacteria with functions including plasmid maintenance, programmed cell death, or stress response (Gerdes et al., 2005; Hayes, 2003). Our mutagenesis data suggest that both antitoxins (SepA1 and SepA2) for either PemK-like or VapC-like toxins are necessary for plasmid exclusion. To confirm their functions, complementation-mediated pANL segregation assays were performed in strains that carry cosmids missing important maintenance regions (Fig. 4). Cosmid 2C10 lacks *anL18–30*, including the entire *sepA1T1* cassette (Fig. 1B & Table 5). Thus the endogenous pANL could not be completely segregated out by introducing this cosmid (Table 5 & Fig. 4A). When a copy of the *sepA1T1* operon was inserted into NS1 (see Materials and methods 2.2) on the chromosome and sufficient selection was performed for antibiotic-resistance markers associated with *sepA1T1* interruption allele, no pANL-specific PCR products were detected from two independent clones, indicating loss of pANL in the strain that had the cassette encoded on the chromosome (Fig. 4A). Complementation by the antitoxin gene *sepA1* alone did not succeed for unknown reasons (data not shown). Another cosmid, 4F10, is missing the *sepA2T2* operon and hence could not compete out pANL (Fig. 1B & Table 5). However, endogenous pANL was completely lost in the strain carrying 4F10 when either the *sepA2T2* cassette or just *sepA2* was inserted into NS1 (Fig. 4B). These results confirmed that both toxin-

antitoxin cassettes are functional and that nothing else in this region is required for maintenance of pANL.

With the knowledge of pANL regions necessary for exclusion, attempts to cure pANL were conducted with a *sacB*-based counter-selection strategy. Selection with *sacB*, which encodes the secretory levansucrase from *Bacillus subtilis* (Ried and Collmer, 1987), has been successfully used in cyanobacteria (Andersson et al., 2000; Cai and Wolk, 1990). First, a copy of *sacB* was inserted onto cosmid 2F10, in which the toxin gene in the *sepA1T1* operon is partially deleted and the antitoxin gene is intact (Fig. 1B), and the segregation of pANL was complete in strains transformed with 2F10 (Table 5). Then the 2F10-*sacB* construct was introduced into a strain that carries a copy of the *sepA2T2* operon in NS1. In this strain the requirement for both pANL-born toxin-antitoxin cassettes was abolished, as was demonstrated individually for each, and 2F10-*sacB* was expected to be dispensable to the cells. However, spontaneous loss of the construct did not occur after culturing the cells in media with no selective antibiotics for many generations, probably due to the presence of other plasmid maintenance components, such as the partitioning system (data not shown). Next, cells were transferred to medium containing 5% sucrose. Only cells that lost the 2F10-*sacB* construct should be able to survive. However, due to technical problems, no sucrose-resistant colonies have been recovered on sucrose-containing medium (data not shown). Trouble-shooting analysis is underway and other counter-selection markers are being tested.

4. Conclusions

It was previously shown that genes from all but one *Bam*HI fragment of pANL display increased transcription under sulfur-deficient conditions (Nicholson and Laudenbach, 1995). The only exception, the Ba3 fragment, covers part of the replication region. Thus, the majority of pANL is involved in response to sulfur starvation. In the BG-11 medium that is widely used for culturing cyanobacteria (Rippka et al., 1979), the main sulfur source, $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, is in a concentration of 0.3 mM, or about 30 mg/L sulfate. Even though many rivers and lakes today are enriched by nitrate and sulfate due to human activities, natural fresh water systems usually contain much lower levels of sulfate (< 10 mg/L). In contrast, sulfate concentration in average seawater is much higher (~185mg/L) (Nelson, 2007). Deficiency of sulfur or other major essential nutrients can result in degradation of the cyanobacterial light-harvesting complex, the phycobilisome (known as photobleaching), as well as modification of other metabolic pathways and physiological processes (Collier et al., 1994; Schwarz and Forchhammer, 2005). Thus, the incurable plasmid pANL is likely to provide important advantages for the fresh water cyanobacterium *S. elongatus* PCC 7942 because it encodes almost all of the components required for adaptation to sulfur deficiency. These include genes functioning in signal transduction, gene expression regulation, and sulfate metabolism. They might enhance the ability of cyanobacterial cells to sense environmental changes in sulfate, elevate the expression levels from genes to increase the import of sulfate, make more cysteine, and modulate other steps in the sulfur metabolism to overcome the sulfur stress. Many of these genes have paralogs on the chromosome. However, the compact and efficient organization of these genes on pANL presumably facilitates cells' reaction in response to sulfur limitation. The potential advantages of retention of genes on pANL under stress conditions may explain the presence of several strategies for plasmid maintenance and the persistence of pANL in *S. elongatus*.

Acknowledgements

We acknowledge the DOE Joint Genome Institute for providing reference sequence of pANL as part of the *Synechococcus elongatus* PCC 7942 genome shotgun sequencing project. We thank Larry Harris-Haller and the Gene Technologies Laboratory staff for technical advice and support. This work was supported by grants from the National

Science Foundation (MCB-9818031), the US Department of Energy (DE-FG03-00ER15055), and the National Institute of Neurological Disorders and Stroke (P01 NS39546).

References

- Altschul SF, et al. Basic local alignment search tool. *J Mol Biol* 1990;215:403–10. [PubMed: 2231712]
- Andersson CR, et al. Application of bioluminescence to the study of circadian rhythms in cyanobacteria. *Methods Enzymol* 2000;305:527–42. [PubMed: 10812624]
- Bose SG, Carmichael WW. Plasmid distribution among unicellular and filamentous toxic cyanobacteria. *Journal of Applied Phycology* 1990;V2:131–136.
- Cai YP, Wolk CP. Use of a conditionally lethal gene in *Anabaena* sp strain PCC 7120 to select for double recombinants and to entrap insertion sequences. *J Bacteriol* 1990;172:3138–45. [PubMed: 2160938]
- Collier JK, et al. Changes in the cyanobacterial photosynthetic apparatus during acclimation to macronutrient deprivation. *Photosynthesis Research* 1994;V42:173–183.
- Felkner RH, Barnum SR. Plasmid content and homology of 16 strains of filamentous, nonheterocystous cyanobacteria. *Curr Microbiol* 1988;V17:37–41.
- Fico S, Mahillon J. *TasA-tasB*, a new putative toxin-antitoxin (TA) system from *Bacillus thuringiensis* pGI1 plasmid is a widely distributed composite *mazE*-doc TA system. *BMC Genomics* 2006;7:259. [PubMed: 17038198]
- Gerdes K, et al. Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol* 2005;3:371–82. [PubMed: 15864262]
- Golden SS, Sherman LA. A hybrid plasmid is a stable cloning vector for the cyanobacterium *Anacystis nidulans* R2. *J Bacteriol* 1983;155:966–72. [PubMed: 6309751]
- Gupta A, et al. Deletion within the metallothionein locus of cadmium-tolerant *Synechococcus* PCC 6301 involving a highly iterated palindrome (HIP1). *Mol Microbiol* 1993;7:189–95. [PubMed: 8446026]
- Hayes F. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science* 2003;301:1496–9. [PubMed: 12970556]
- Holtman CK, et al. High-Throughput Functional Analysis of the *Synechococcus elongatus* PCC 7942 Genome. *DNA Res* 2005;12:103–15. [PubMed: 16303742]
- Ishikawa J, Hotta K. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. *FEMS Microbiol Lett* 1999;174:251–3. [PubMed: 10339816]
- Kaneko T, et al. Structural analysis of four large plasmids harboring in a unicellular cyanobacterium, *Synechocystis* sp PCC 6803. *DNA Res* 2003;10:221–8. [PubMed: 14686584]
- Kaneko T, et al. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res* 2001;8:205–13. 227–53. [PubMed: 11759840]
- Koksharova OA, Wolk CP. Genetic tools for cyanobacteria. *Appl Microbiol Biotechnol* 2002;58:123–37. [PubMed: 11876404]
- Lau RH, Doolittle WF. Covalently closed circular DNAs in closely related unicellular cyanobacteria. *J Bacteriol* 1979;137:648–652. [PubMed: 104976]
- Lau RH, et al. Cyanobacterial plasmids: Their widespread occurrence, and the existence of regions of homology between plasmids in the same and different species. *Molecular and General Genetics MGG* 1980;V178:203–211.
- Laudenbach DE, et al. The large endogenous plasmid of *Anacystis nidulans*: Mapping, cloning and localization of the origin of replication. *Molecular and General Genetics MGG* 1983;V192:402–407.
- Laudenbach DE, et al. Evidence for two distinct origins of replication in the large endogenous plasmid of *Anacystis nidulans* R2. *Molecular and General Genetics MGG* 1985;V199:300–305.
- Maniatis, T., et al. *Molecular cloning: A Laboratory Manual*. Cold Spring Harbor Lab Press; Cold Spring Harbor, New York: 1982. p. 545
- Marchler-Bauer A, et al. CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* 2005;33:D192–6. [PubMed: 15608175]
- Nakamura Y, et al. Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. *DNA Res* 2002;9:123–30. [PubMed: 12240834]

- Nelson, DO. Fresh Water, Natural Composition of Water Encyclopedia. 2007. 2007.
- Nicholson ML, et al. Two enzymes together capable of cysteine biosynthesis are encoded on a cyanobacterial plasmid. *Molecular and General Genetics MGG* 1995;V247:623–632.
- Nicholson ML, Laudenbach DE. Genes encoded on a cyanobacterial plasmid are transcriptionally regulated by sulfur availability and CysR. *J Bacteriol* 1995;177:2143–50. [PubMed: 7536734]
- Paithoonrangsarid K, et al. Five histidine kinases perceive osmotic stress and regulate distinct sets of genes in *Synechocystis*. *J Biol Chem* 2004;279:53078–86. [PubMed: 15471853]
- Rebière M-C, et al. Plasmid distribution among unicellular and filamentous cyanobacteria: Occurrence of large and mega-plasmids. *FEMS Microbiology Letters* 1986;37:269–275.
- Ried JL, Collmer A. An *nptI-sacB-sacR* cartridge for constructing directed, unmarked mutations in gram-negative bacteria by marker exchange-eviction mutagenesis. *Gene* 1987;57:239–46. [PubMed: 3319780]
- Rippka R, et al. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* 1979;111:1–61.
- Robinson PJ, et al. HIP1 propagates in cyanobacterial DNA via nucleotide substitutions but promotes excision at similar frequencies in *Escherichia coli* and *Synechococcus* PCC 7942. *Mol Microbiol* 1997;24:181–9. [PubMed: 9140975]
- Schwarz R, Forchhammer K. Acclimation of unicellular cyanobacteria to macronutrient deficiency: emergence of a complex network of cellular responses. *Microbiology* 2005;151:2503–2514. [PubMed: 16079330]
- Sherman LA, van de Putte P. Construction of a hybrid plasmid capable of replication in the bacterium *Escherichia coli* and the cyanobacterium *Anacystis nidulans*. *J Bacteriol* 1982;150:410–3. [PubMed: 6277870]
- Stanier RY, et al. Purification and properties of unicellular blue-green algae (order *Chroococcales*). *Bacteriol Rev* 1971;35:171–205. [PubMed: 4998365]
- Stock AM, et al. Two-component signal transduction. *Annu Rev Biochem* 2000;69:183–215. [PubMed: 10966457]
- Tatusova TA, Madden TL. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* 1999;174:247–50. [PubMed: 10339815]
- Ulrich LE, et al. One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol* 2005;13:52–6. [PubMed: 15680762]
- van den Hondel CA, et al. Homology of plasmids in strains of unicellular Cyanobacteria. *Plasmid* 1979;2:323–33. [PubMed: 113796]
- van den Hondel CA, et al. Introduction of transposon Tn901 into a plasmid of *Anacystis nidulans*: preparation for cloning in cyanobacteria. *Proc Natl Acad Sci USA* 1980;77:1570–4. [PubMed: 6246495]
- Van der Plas J, et al. Identification of replication and stability functions in the complete nucleotide sequence of plasmid pUH24 from the cyanobacterium *Synechococcus* sp PCC 7942. *Mol Microbiol* 1992;6:653–64. [PubMed: 1552863]
- Xu W, McFadden BA. Sequence analysis of plasmid pCC5.2 from cyanobacterium *Synechocystis* PCC 6803 that replicates by a rolling circle mechanism. *Plasmid* 1997;37:95–104. [PubMed: 9169201]
- Yang X, McFadden BA. 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* 1993;175:3981–91. [PubMed: 8320214]
- Yang X, McFadden BA. The complete DNA sequence and replication analysis of the plasmid pCB2.4 from the cyanobacterium *Synechocystis* PCC 6803. *Plasmid* 1994;31:131–7. [PubMed: 8029321]

Fig. 1.

Graphical representations of the replication origin region (**A**) and the plasmid maintenance region (**B**). Putative ORFs are shown as grey boxes with arrowheads. (**A**) *KpnI* and *BamHI* sites are shown, as is the position of the first nucleotide of pANL numbering (1); the thick horizontal black line above the ORFs indicates the putative replication origin suggested by deletion mapping analysis. (**B**) Black arrowheads are *Mu* insertions; dashed thick lines under ORFs indicate regions of pANL that are absent from the cosmid labeled below each; the coordinates of *Mu* insertions are listed in brackets: 5H4D2 (20823–20827), 5H4H11 (21297–21301), 5H4B2 (21508–21512), 2A8D4 (21759–21763), 5H4A2 (21905–21909), 6C4Q4 (22625–22630), 5H4C12 (23120–23124), 5H4E8 (23604–23608), 2A8E9 (23867–23871), 5H4D7 (23898–23902), and 5H4A10 (24319–24323).

Fig. 2.

Graphical representation of cosmid-based deletion analysis. Putative pANL ORFs are shown in black arrows. Functional regions are labeled above the ORFs: origin, the replication origin region (*anL55.5–anL12*); signal, the signal transduction region (*anL14–anL27*); maintenance, the plasmid maintenance region (*anL28–anL34*); sulfur, the sulfur-related region (*anL35–anL55*). The first and last ORFs, *anL01* and *anL59*, as well as the last ORF of each functional region are also labeled. Regions of pANL that are missed in each cosmid are displayed as white boxes in the order of their locations (see Table 5). + or – marks after the plasmid names refer to positive or negative PCR results from boiled samples of cyanobacterial cells transformed with cosmids (see Results and Discussion 3.2.1.). Gaps shared by cosmids that were unable to replace pANL are indicated by vertical lines and pairs of arrowheads below the ORFs.

Fig. 3.

The segregation of pANL in transposon-mediated mutant strains of plasmid maintenance genes. PCR amplification data from boiled samples are shown. In both upper and lower panels, bottom labels: M, 100-bp DNA ladder (Invitrogen); 1–16 (upper panel) or 1–20 (lower panel), boiled cyanobacterial samples; top labels: WT, wild-type strain AMC06; T, primer pair specific for regions surrounding the *Mu* insertion; S, primer pair specific for region with no insertion and shared by both pANL and the cosmid. Primer pairs used for detecting the insertion of mini*Mu* and wild-type copy of pANL sequence are: 5H4D2, 2ie3 × 614ie5 (T) and pan11 × pan12 (S); 5H4H11, 2ie3 × pan10-2 (T) and pan11 × pan12 (S); 5H4B2, 2ie3 × pan10-2 (T) and pan11 × pan12 (S); 2A8D4, pem5 × pemi3 (T) and 2ie3 × pan10-2 (S); 5H4A2, pan11 × pan12 (T) and 2ie3 × pan10-2 (S); 6C4Q4, pan11 × pan12 (T) and pan13 × pan14 (S); 5H4C12, pan13 × pan14 (T) and pan11 × pan12 (S); 5H4E8, pan13 × pan14 (T) and pan11 × pan12 (S); and 5H4A10, add2–5 × add2–34 (T) and pan11 × pan12 (S).

Fig. 4.

Complementation analysis of toxin-antitoxin cassettes. **(A)** Complementation with the *sepA1T1* operon in NS1 resulted in full segregation of pANL in a 2C10-transformed strain (AMC1584). WT, wildtype AMC06; 2C10, 2C10-transformed strain; AT1-1 & AT1-2, *sepA1T1*-complemented AMC1584 strains. D, primer pair specific for region deleted in the cosmid; S, primer pair specific for region shared by both pANL and the cosmid; M, 100-bp DNA ladder (Invitrogen); 1–12, boiled samples. Primer pairs used, in left-to-right order, 2ie3 $\times$ 614ie5 (D), 8e3 $\times$ 5e5 (D), and co4 $\times$ co8 (S). **(B)** Segregation of pANL is completed in 4F10-transformed strains (AMC1505) when a copy of *sepA2T2* or *sepA2* was inserted into NS1. WT, wild-type AMC06; 4F10, 4F10-transformed strain; AT2, *sepA2T2*-complemented AMC1505 strain; A2, *sepA2*-complemented AMC1505 strain. D, primer pair specific for deleted region in the cosmid; S, primer pair specific for region shared by both pANL and the cosmid; M, 100-bp DNA ladder (Invitrogen); 1–12, boiled samples. Primer pairs used, in left-to-right order: 614ie3 $\times$ gap43 (D), gap45 $\times$ 1ie5 (D), and gap35 $\times$ 10ie5 (S).

Table 1

Cyanobacterial strains and plasmids used in this study.

Strains	Genetic background	Plasmid/cosmid transformed	Antibiotic resistance ^a	Source or reference
AMC06	Wild-type	None	None	Lab collection
AMC1502	Wild-type	pAM3617	Gm ^R	This study
AMC1503	Wild-type	pAM3618	Gm ^R	This study
AMC1505	Wild-type	4F10	Km ^R	This study
AMC1529	Wild-type	pAM3715	Sp ^R Sm ^R	This study
AMC1530	Wild-type	pAM3716	Sp ^R Sm ^R	This study
AMC1584	Wild-type	2C10	Km ^R	This study
AMC1585	AMC1584	pAM3716	Sp ^R Sm ^R Km ^R	This study
AMC1586	AMC1505	pAM3617	Gm ^R Km ^R	This study
AMC1587	AMC1505	pAM3618	Gm ^R Km ^R	This study
Plasmid	Characteristics		Antibiotic resistance*	Source or reference
pAM1303	NS1 cloning vector		Sp ^R Sm ^R	Andersson et al. (2000)
pAM3511	Source of Gm-resistance gene (<i>accC1</i>)		Gm ^R Km ^R	This study
pAM3558	Gm resistant version of pAM1303		Gm ^R	This study
pAM3617	<i>sepA2T2</i> in pAM3558		Gm ^R	This study
pAM3618	<i>sepA2</i> in pAM3558		Gm ^R	This study
pAM3715	<i>sepA1</i> in pAM1303		Sp ^R Sm ^R	This study
pAM3716	<i>sepA1T1</i> in pAM1303		Sp ^R Sm ^R	This study

^aResistance to: Cm^R, chloramphenicol; Km^R, kanamycin; Sp^R, spectinomycin; Sm^R, streptomycin; Gm^R, gentamycin; Ap^R, ampicillin

Table 2

Primers^a used in this study.

Primer	Sequence (5'→3')	Primer	Sequence (5'→3')
pan1	gcgacggcagatttcagacc	pan14	tgattggcgggtttctggtg
pan2	gcaatgcggaccgagcacc	<u>614ie3</u>	aggctccccgtttgaatcc
Co1	gatccagtcactcacgaggtg	<u>gap43</u>	cagctcagtgctgagacagg
2H8-B	ggaggatcgccatcgaagc	<u>lie5</u>	gtgcggcatttccttaatgc
pan3	gatcaaaggagcacgacggg	<u>gap45</u>	tggacatactccagcgtg
pan4-2	cggacttgctcgtggctgc	co28	ccggtactgaacgaggaaactgaggg
2H8-A	ctctcaaacgacgaaagcc	co29	cgccagtggagcgggtatctaccg
<u>lie3</u>	cctgggaattctcgtctgg	co24	gtctcaattactggatggcgagcc
<u>gap53</u>	agctattgctggaggccatg	co23	gctcaagagggaaggactgatttcgg
<u>gap55</u>	cgctactaccagctcgacatg	co27	cgttaaatgtcggattgggtgcagcc
<u>7ie3</u>	cttacaattaaagtaattctctcacactcc	co2	gtctgctcatgaatcgctgg
<u>7ie5</u>	agaggcgcattgtgagattg	co26	ttgtagagcctgtgctgatgcctgc
<u>gap13</u>	ttgatatgaagacgccaccg	co25	actcgtctccgctctttgctatggc
<u>3e3</u>	cccaagacttctctctttcacc	co6	cgggaccaagcagaaggactcagcg
<u>gap15</u>	tgtccgcaacattgagcg	co3	caaggcgcatcgctaacc
<u>gap23</u>	tggcatcgttgaactgaag	co7	tcgcttcttgaacatggcactgctc
<u>gap25</u>	tgaatctggcctatcgaaag	co8	gagacggaaggagagacctagccag
<u>411e5</u>	gaaccttgacgtgaccacgt	co4	ctttgaccttgacggctgc
<u>411e3</u>	gcatgacgcttgaagaaatgc	co5	gctgtgcagtttcttggagg
<u>gap33</u>	gggctgatcaatggtttgg	co13	ctgaagcgatgctgggctgtattgc
<u>10ie5</u>	ccaagccacacctgaacaga	co14	ctgggccagttctgattgcgttgc
<u>gap35</u>	tgtctctctctgcaaagacagc	co19	ccgctgtatctcgttgaggcaaccc
<u>8e5</u>	tcggccttgtagcaaaac	co20	acgagattggaggcccagggaagg
<u>10ie3</u>	gccactggccgaaaaagc	co21	tcgccatccccatctgctgttgc
<u>5e5</u>	ggtggaatcggaacaaatcg	co22	gtagcgacgcgaatggtagcaggc
<u>8e3</u>	acttcaacgcccgacatg	co30	gctctgctgcgttcttctgagggc
pan5	gcaactgccctgtccaatcg	co31	cagatatgggtaatgctgtagtgaggc
pan6	gactgctctgtctgtgcg	pem5	cactgagaacgagaagcggctcgac
pan7	agcaagcgcaattgagcagc	pemi3	gacgtagcttcgttcgactgaggtc
pan8	ggtaacggcgaaagaccctc	pemk3	tctgaggcagcgcttaaccagtaggc
pan9	cctcaccacagtagcttggc	add2-5	ctgcctgccatggaaagtgcgtc
<u>614ie5</u>	ggtaaggcattgttttaaccacc	add2-33	tcagggaagacagcctttgaggtgg
<u>2ie3</u>	cagagcagccgcagaatc	add2-34	gtgctactggatgtcagagagtcg
pan10-2	gccctctgtacctgatttgagctg	sci-1	cgaccgatgcccttgagagc
pan11	ccgcttctcaatcaagtcg	sci-2	gcttccattcaggtcgaggtgg
pan12	caccgcgtcagtttgaacg	sci-3	caacctatggaactgatgaatgggagc
pan13	gcctctcgtgtccgctgc	sci-4	cagtcgcttcacgtgcctgc

^aUnderlined primers were used for primer-walking.

Table 3

List of putative ORFs on pANL and their functional assignments.

Putative ORF ^d	Coordinates	% (G+C)	Size (aa)	Putative product	Conserved domains (CD)	Region ^b (aa)	% CD aligned ^c	E-value
<i>anL01</i>	205–3204	57	999	Conserved hypothetical protein	COG1763 (MobB)	304–454	93	0.077
<i>anL02</i>	3314–4228 ^c	36	304	Conserved hypothetical protein	-	-	-	-
<i>anL03</i>	4484–4699 ^c	57	71	Conserved hypothetical protein	-	-	-	-
<i>anL04</i>	5077–5250	59	57	Hypothetical protein	-	-	-	-
<i>anL05</i>	5275–6699 ^c	54	474	Conserved hypothetical protein	COG5361	26–473	95	2e-71
<i>anL06</i>	7072–8031	50	319	Conserved hypothetical protein	-	-	-	-
<i>anL07</i>	8462–8722	56	86	Hypothetical protein	-	-	-	-
<i>anL08</i>	8807–9094	59	95	Hypothetical protein	-	-	-	-
<i>anL09</i>	9101–9418 ^c	50	105	Conserved hypothetical protein	-	-	-	-
<i>anL10</i>	9463–9732 ^c	56	89	C-terminal domain of recombinase	cd01182 (INT_REC_C)	1–55	33	5e-7
<i>anL11</i>	9877–10170	51	97	Conserved hypothetical protein	COG2929	1–86	96	2e-11
<i>anL12</i>	10136–10465	58	109	Conserved hypothetical protein	COG3514	35–108	81	2e-11
<i>anL13</i>	10610–10849	54	79	Hypothetical protein	-	-	-	-
<i>anL14</i>	10980–11411 ^c	56	143	Transcriptional regulator	COG3432	1–78	83	6
<i>anL15</i>	11601–11726	45	41	Conserved hypothetical protein	-	-	-	-
<i>anL16</i>	11835–12809 ^c	58	324	Zn-dependent oxidoreductase	COG0604 (Qor)	1–324	100	1e-9
<i>anL17</i>	13037–13762 ^c	47	241	Type 1 glutamine amidotransferase	cd03139 (GATase_L_PtPlL2)	6–184	100	1e-46
<i>anL18</i>	13855–14562	43	235	One-component signal transduction protein	Smart00065 (GAF)/Smart00421 (HTH_LUXR)	24–161/176–231	89/97	1e-10/ -
<i>anL19</i>	14905–15528	55	207	Conserved hypothetical protein	COG3544	49–207	78	8e-17
<i>anL20</i>	15804–17120 ^c	59	438	Histidine protein kinase	Smart00387 (HATPase_C)	330–433	100	1e-25
<i>anL21</i>	17089–17763 ^c	55	224	Response regulator	COG0745 (OmpR)	2–223	98	4e-51
<i>anL22</i>	17986–18468	55	160	Conserved hypothetical protein	-	-	-	-
<i>anL23</i>	18632–19222	56	196	Transcriptional regulator	Smart00420 (HTH_DEOR)	44–90	87	0.82
<i>anL24</i>	19197–19586	48	129	Conserved hypothetical protein	-	-	-	-
<i>anL25</i>	19841–20083	50	80	Conserved hypothetical protein	-	-	-	-
<i>anL26</i>	20113–21075 ^c	56	320	Site-specific recombinase	COG4974 (XerD)	24–318	92	2e-26
<i>anL27</i>	21165–21515 ^c	50	116	PemK-like toxin (SepT1)	Pfam02452 (PemK)	4–113	100	2e-4
<i>anL28</i>	21506–21721 ^c	49	71	Antidote (SepA1)	-	-	-	-
<i>anL29</i>	21765–22094 ^c	52	109	Nucleotidyltransferase	COG1669	18–101	80	6e-13
<i>anL30</i>	22197–23021 ^c	56	274	Partitioning protein ParB	Smart00470	36–76	45	0.87
<i>anL31</i>	23018–23629 ^c	56	203	Partitioning protein ParA	COG1192 (Soj)	4–201	99	1e-11
<i>anL32</i>	23763–24029	55	88	Antitoxin (SepA2)	-	-	-	-
<i>anL33</i>	24040–24423	52	127	VapC-like toxin (SepT2)	COG1487 (VapC)	1–123	97	3e-11
<i>anL34</i>	24420–24770 ^c	52	116	Rhodanese-related sulfotransferase	COG0607 (PspE)	17–112	88	1e-16
<i>anL35</i>	24860–26740 ^c	55	626	Cysteine desulfurase	COG1104 (NifS)	243–619	96	2e-48
<i>anL36</i>	26975–27895 ^c	55	306	Conserved hypothetical protein SrpI	-	-	-	-
<i>anL37</i>	27990–28949 ^c	57	319	Serine acetyltransferase SrpH	COG1045 (CysE)	134–301	82	9e-41
<i>anL38</i>	28946–29935 ^c	56	329	Cysteine synthase SrpG	COG0031 (CysK)	8–314	100	5e-97
<i>anL39</i>	30391–30750	48	119	Conserved hypothetical protein	-	-	-	-
<i>anL40</i>	31000–31545 ^c	53	181	Conserved hypothetical protein	COG1791	47–175	75	4e-11
<i>anL41</i>	31647–32081	54	144	Transcriptional regulator	COG1959	3–144	96	5e-21
<i>anL42</i>	32300–32878	53	192	Carbonic anhydrase	cd03379	28–182	100	6e-49
<i>anL43</i>	32980–33999	50	339	Catalase SrpA	Pfam00199 (catalase)	77–355	76	6e-31
<i>anL44</i>	34100–34648	56	182	Conserved hypothetical protein SrpB	Pfam02308 (MetC family)	18–73	43	3e-13
<i>anL45</i>	34800–35981	46	393	Chromate transporter SrpC	Pfam02417/COG2059	14–190/221–378	100/82	7e-46
<i>anL46</i>	36112–37101	57	329	Cysteine synthase SrpD	COG0031 (CysK)	4–304	100	8e-96
<i>anL47</i>	37098–38624	59	508	Gamma-glutamyltranspeptidase SrpE	Pfam01019	31–503	100	8e-121
<i>anL48</i>	38849–39376	57	175	Conserved hypothetical protein SrpF	-	-	-	-
<i>anL49</i>	39387–40448	58	353	Acyl-CoA dehydrogenase SrpI	COG1960 (CaiA)	4–324	87	1e-19

Putative ORF ^a	Coordinates	% (G+C)	Size (aa)	Putative product	Conserved domains (CD)	Region ^b (aa)	% CD aligned ^c	E-value
<i>anL53</i>	40445–41257	58	270	Nitrate/sulfonate/bicarbonate ABC transporter, ATPase subunit SrpK	COG11116 (Taub)	17–264	97	3e–68
<i>anL54</i>	41283–42506	56	407	Nitrate/sulfonate/bicarbonate ABC transporter, periplasmic subunit SrpL	COG0715 (Taub)	12–350	97	4e–44
<i>anL55</i>	42508–43308	57	266	Nitrate/sulfonate/bicarbonate ABC transporter, permease subunit SrpM	COG0600 (Taub)	54–266	82	4e–30
<i>anL55.5</i>	43355–43579	59	74	Conserved hypothetical protein	-	-	-	-
<i>anL56</i>	43634–43912	55	92	Conserved hypothetical protein	COG2929	1–85	95	6e–11
<i>anL56.5</i>	43866–44117	57	83	Conserved hypothetical protein	COG3514	22–78	61	2e–4
<i>anL57</i>	44122–44325	54	67	Conserved hypothetical protein	-	-	-	-
<i>anL58</i>	44375–45211 ^c	38	278	Conserved hypothetical protein	-	-	-	-
<i>anL59</i>	45309–45881 ^c	58	190	Conserved hypothetical protein	-	-	-	-

^aORFs that have at least one paralogue (>30% protein sequence identity over 70% of the total length) on the chromosome of *S. elongatus* are highlighted.

^bRegion of the pANL protein that shows significant similarity to the relevant conserved domain in GenBank.

^cPercent of the length of conserved domain that can be aligned with pANL-encoded protein in the CD search (rpsblast).

Table 4

Large repetitive sequence pairs on pANL.

Repeat Sequence Pair (RSP) ^a	Type	Coordinates (ORF) Sequence 1	Coordinates (ORF) Sequence 2	Size (bp)	% identity
RSP-1	direct	9517–9762c (<i>anL10</i>)	20119–20364c (<i>anL28</i>)	246	84
RSP-2	direct	43412–43560 (<i>anL55.5</i>)	44161–44309 (<i>anL57</i>)	149	87
RSP-3	direct	19474–19531 (<i>anL26</i>)	19956–20013 (<i>anL27</i>)	58	82
RSP-4	direct	10398–10423 (<i>anL12</i>)	44062–44087 (<i>anL56.5</i>)	26	92
RSP-5	inverse	9150–9341c (<i>anL09</i>)	19828–20019 (<i>anL27</i>)	192	75

^a Only repeat sequence pairs larger than 20 bp are listed.

Table 5
Summary of cosmid-based deletion analysis of pANL.

Cosmids Introduced a	SuperCos I vector ^b	Region shared by both pANL and cosmid ^c	pANL region absent from cosmid ^d	ORFs not in cosmid	Coordinates of region not in cosmid (bp)
9H1	+	+	–	01–03	1776–4739
8A9	+	+	–	01–15	696–11492
2H7	+	+	–	06–23	6841–17582
2C10	+	+	+	18–30	13446–22164
9D9	+	+	+	19–30	13891–22171
2F10	+	+	–	21–29	14662–21293 <
2D12	+	+	+	22–50	15765–37343
2G5	+	+	+	23–34	17734–24298
9G2	+	+	+	31–36	22960–25049
4D1	+	+	+	31–42	22149–30520
4F10	+	+	+	32–36	23346–25770
2B10	+	+	–	33–39	23818–28499
4H2	+	+	–	33–46	23818–33455 <
3F2	+	+	+	38–56	27497–43702
10F3	+	+	+	43–16	31466–12193
2A8	+	+	+	45–59	32085–45552
2B4	+	+	–	46–48	33890–35060
2D11	+	+	–	48–50	35811–37343 <
9E12	+	+	+	48–56.5	35196–43990
9H7	+	+	+	48–59	35060–45705
6H1	+	+	–	49–55	36765–42913 <
2D1	+	+	+	52–01	39665–886
2F9	+	+	+	54–05	42003–6375
7G4	+	+	+	53–02	40591–4172
9B9	+	+	+	54–59	41738–45913
4F8	+	+	–	57–06	44307–7174
7F8	+	+	–	59–12	45658–10775

^a Cosmid that was introduced into *S. elongatus* wild-type strain AMC06.
^b PCR results using SuperCos I specific primer pairs (sci-1 × sci-2 & sci-3 × sci-4) to detect the presence of the cosmid.
^c PCR results using primer pairs specific for the shared region present in both pANL and the cosmid.
^d PCR results using primer pairs specific for the pANL region absent from the cosmid. +, Positive PCR result from boiled samples of *S. elongatus* cells transformed with cosmid; –, highlighted negative PCR result from those boiled samples (see Results and discussion). Four of the – cosmids that do not overlap next – cosmid are indicated by arrows. The resulted gaps are also marked by arrows in Fig. 3A.