Cyanobacterial Transposons Tn5469 and Tn5541 Represent a Novel Noncomposite Transposon Family

MICHAEL R. SCHAEFER* AND KATHERINE KAHN†

Division of Molecular Biology and Biochemistry, School of Biological Sciences, University of Missouri-Kansas City, Kansas City, Missouri 64110

Received 19 June 1998/Accepted 15 September 1998

A noncomposite transposon, designated Tn5541, was isolated from strain Fd33 of the filamentous cyanobacterium Fremyella diplosiphon UTEX 481. Sequence analysis showed that Tn5541 is structurally and genetically very similar to Tn5469, which is also endogenous to F. diplosiphon. Both Tn5469 and Tn5541 encode homologous forms of an unusual composite transposase and a protein of unknown function. DNA hybridization analysis showed that like Tn5469, Tn5541 was not widely distributed among cyanobacterial genera. A similar analysis showed that Tn5469 and Tn5541 were equally limited to and present as multiple genomic copies in three of six distinct strains comprising the Tolypothrix 1 cluster of heterocyst-forming filamentous cyanobacteria. These and other distinguishing features suggest that Tn5469 and Tn5541 represent a novel noncomposite transposon family.

A feature common to all examined bacterial genomes is the presence of transposable genetic elements. The simplest and most-common forms are insertion sequence (IS) elements, which range in size from 0.6 to 2.5 kbp and are found in multiplicities from a few to a few hundred per genome (7, 12). IS elements are characterized by one or a few genes required for transposition, which are flanked by terminal inverted repeat (IR) sequences. The more-complex transposable element forms are transposons that range in size from 2.5 to 60 kbp (3). Transposons and IS elements share many structural, functional, and genetic features associated with transposition, including generation of short directly repeated duplications of a target sequence. However, transposons usually possess one or more genes that confer a phenotype on the bacterial host, such as resistance to a specific antibiotic. Most bacterial transposons are typed as composite or noncomposite forms. Composite transposons are composed of two IS elements of the same type bracketing one or more genes. In noncomposite forms, the transposition and nontransposition genes are clustered and flanked by terminal IR sequences. Because transposition can be mutagenic, the activity of these transposable elements can significantly influence genome evolution.

Little is known about the transposable elements endogenous to the morphologically diverse and widely distributed cyanobacteria. With the exception of transposon Tn5469 (see below), only characteristic IS elements have been isolated from different strains of this large eubacterial group. Earlier reports described 14 cyanobacterial IS elements which were isolated from species in the filamentous Anabaena and Calothrix genera (1, 2, 5, 22). Recently, IS elements were isolated from two unicellular Synechocystis species, one from Synechocystis sp. strain BO 8402 (4) and three from Synechocystis sp. strain PCC 6803 (6). At least three additional IS element forms reported for Synechocystis sp. strain PCC 6803 (19) remain to be characterized.

We previously characterized transposon Tn5469 (16), which was isolated from strain Fd33 of the filamentous cyanobacterium Fremyella diplosiphon UTEX 481 (also referred to as Calothrix sp. strain PCC 7601). The 4,904-bp Tn5469 is a noncomposite transposon that encodes a large composite transposase, designated TnpA, and two unidentified proteins. Five copies of Tn5469 are present on the wild-type genome. Like most characterized cyanobacterial IS elements, Tn5469 is not widely distributed among morphologically distinct genera. Tn5469 has been implicated as the agent responsible for the phenotype of several different pigmentation mutants of strain Fd33 (15–17). Each of these mutants was found to possess an extra genomic copy of Tn5469, which facilitated isolation of the affected gene in a manner analogous to transposon tagging. During the course of analyzing a putative Tn5469-generated mutant, we identified a Tn5469-like transposon, which was designated Tn5541. Here we show that Tn5541 is very similar to Tn5469 and suggest that both represent a novel transposon family.

Identification of Tn5541. Strain Fd33 is a short-filament mutant of F. diplosiphon UTEX 481 (9) that provides for colonial growth on solid medium. Pigmentation mutant strain FdGM2 was derived from strain Fd33 and contains an extra (sixth) genomic copy of Tn5469 (18) localized to an uncharacterized open reading frame (ORF) temporarily designated rcaY. A recombinant library of Fd33 genomic DNA was constructed in λEMBL3 (13). Clone λUMC001 from this library harbors a 15-kbp fragment containing intact rcaY sequences. DNA sequence analysis of a region flanking rcaY on the λUMC001 insert indicated the presence of an ORF predicting a polypeptide that has significant sequence identity with the TnpA transposase encoded by Tn5469. To investigate whether this sequence represented a new transposable genetic element, corresponding DNA fragments were subcloned from the λUMC001 insert for sequencing and genetic analysis. This effort revealed that the λUMC001 insert harbors a 4.7-kbp noncomposite transposon, which was designated Tn5541.

General features of Tn5541. The 4,745-bp transposon Tn5541 contains 12-bp near-perfect (9 of 12 bp) terminal IRs
that do not match corresponding sequences from known IS elements or transposons. Based on the orientation of two ORFs (see below), the IRs were designated IR\text{L} for the left end and IR\text{R} for the right end of the element. On the \text{UMC001} insert, \text{Tn5541} is not flanked by a duplicated target sequence; IR\text{L} is flanked by the sequence 5'-TGTTTAT-3', whereas IR\text{R} is flanked by the sequence 5'-GTGTTAT-3'. However, subsequent analysis of four additional genomic copies of \text{Tn5541} in \text{Calothrix} sp. strain PCC 7601 (see below) showed that three were flanked by different, duplicated 5-bp target sequences. Presumably, transposition of \text{Tn5541} in \text{F. diplosiphon} is accompanied by a characteristic generation of a duplicated target sequence, which in some cases becomes altered over time.

The nucleotide sequence of \text{Tn5541} predicts two ORFs arranged in tandem on the element (Fig. 1A). The left ORF (designated \text{tnpA}; nucleotide positions 351 to 3066), which is preceded by an \text{E. coli}-like promoter, predicts a 904-residue protein with a molecular mass of 103.2 kDa and a pI of 8.91. A BLAST (tblastn) search of the GenBank database indicated that the 904-residue protein has sequence identity with a number of transposases, most significantly with the \text{TnpA} transposase from \text{Tn5469}. The right ORF (designated \text{orf1}; nucleotide positions 3069 to 4533) initiates 3-bp downstream of \text{tnpA} and predicts a 488-residue protein with a molecular mass of 55.0 kDa and a pI of 8.62. This proximity and the lack of distinguishable promoter sequences upstream of \text{orf1} suggest that \text{tnpA} and \text{orf1} may be cotranscribed. A BLAST comparison of the \text{ORF1} polypeptide sequence against the GenBank database revealed no significant matches except for the partially homologous \text{ORF1} polypeptide from \text{Tn5469}.

**Homology between \text{Tn5541} and \text{Tn5469}**

The \text{TnpA} protein encoded by \text{Tn5541} has 43.9% overall amino acid sequence identity with the \text{TnpA} transposase from \text{Tn5469}. An alignment of the \text{TnpA} sequences is presented in Fig. 2A. The two \text{TnpA} proteins are homologous composite of two different and widely distributed transposase forms; this composite for \text{TnpA} from \text{Tn5469} was defined earlier (16). The amino-terminal one-third of the shared \text{TnpA} protein most resembles the 226-residue \text{ORF1} transposase encoded by \text{Lactococcus lactis} insertion sequence ISS1 (14) (data not shown). The ISS1 transposase has significant sequence identity with transposases encoded by IS elements found in gram-negative and Gram-positive bacteria (27). The carboxyl-terminal two-thirds of the shared \text{TnpA} protein most resembles the transposases encoded by noncomposite transposons \text{Tn5090} (23), \text{Tn552} (25), and \text{Tn7} (11) (data not shown). Like the transposases from \text{Tn5090}, \text{Tn552}, and \text{Tn7}, the \text{TnpA} proteins contain the invariant D(D(35)E) motif (Fig. 2A) that characterizes the members of a superfamily of bacterial transposases and eukaryotic retroviral and retrotransposon integrase proteins (10, 21).

However, in comparison to \text{Tn5090}, \text{Tn552}, and \text{Tn7}, both \text{Tn5541} and \text{Tn5469} are simple in genetic structure and encode a single defined transposition protein. The contrasting genetic organization and unifying transposase suggest that all of these transposons arose independently but evolved from a common ancestor.

The \text{ORF1} protein encoded by \text{Tn5541} is partially homologous to the corresponding \text{ORF1} protein encoded by \text{Tn5469}. An alignment of the two \text{ORF1} sequences is presented in Fig. 2B. The amino-terminal 336 residues of the \text{Tn5541} \text{ORF1} protein have 31.3% sequence identity with the amino-terminal 329 residues of the \text{Tn5469} \text{ORF1} protein. Outside of this region, the different-sized \text{ORF1} proteins have limited sequence identity. A BLAST analysis of the CyanoBase database (19) showed that the \text{ORF1} proteins have domains of sequence identity with several \text{Synechocystis} sp. strain PCC 6803 proteins (data not shown). The amino-terminal one-third of \text{Tn5541} \text{ORF1} (and corresponding region of \text{Tn5469} \text{ORF1}) has greatest sequence identity with a putative ABC transporter component (CyanoBase designation slr0182) and the DnaX component of DNA polymerase III (CyanoBase designation slr1360). A feature common to these proteins is the presence of a P-loop ATP-binding motif ([AG]X4GK[S/T]) (26); this motif for the \text{ORF1} proteins is identified in Fig. 2B. The central one-third of \text{Tn5541} \text{ORF1} (and corresponding region of \text{Tn5469} \text{ORF1}) has significant sequence identity with two defined ClpB proteases (CyanoBase designations slr0156 and slr1641). Finally, the carboxyl-terminal one-third of \text{Tn5541} \text{ORF1} is unique and resembles an unidentifed protein (CyanoBase designation slr0188). The carboxyl-terminal region of \text{Tn5541} \text{ORF1}, as well as the slr0188 protein, contains many proline and threonine residues and aligns to a number of eukaryotic glycoproteins in the GenBank database. The fact that the homologous \text{ORF1} protein encoded by \text{Tn5541} and \text{Tn5469} has domains of identity with these unrelated cyanobacterial enzymes is interesting; however, a functional unification of these features is not obvious.

A comparison of terminal nucleotide sequences revealed a structural similarity between the IR regions of \text{Tn5541} and \text{Tn5469}. Among the terminal 25 bases containing their respective IR sequences, 11 are identical between IR\text{L} and IR\text{R} for \text{Tn5541} (Fig. 3A), whereas 23 are identical between IR\text{L} and IR\text{R} for \text{Tn5469} (Fig. 3B). The IRs of both transposons possess the terminal dinucleotide 5'-TG, which is a distinguishing feature of the prokaryotic and eukaryotic mobile genetic elements encoding transposases characterized by the D(D(35)E) motif described above. Curiously, the sequence of the \text{Tn5541} IR\text{R} region shares 17 of the 25 bases comprising the IR\text{R} of \text{Tn5469}. In contrast, the \text{Tn5541} IR\text{L} and IR\text{R} regions share only 11 of the terminal 25 bases. Thus, the \text{Tn5541} IR\text{R} region more closely resembles the terminal IR structure of \text{Tn5469} than its
corresponding IR region. The functional significance of this 
IR sequence similarity is unknown.

Among the transposon sequences deposited in the GenBank 
database, Tn5461 most resembles Tn5469. Both transposons 
are approximately 5 kb in length and possess similar terminal 
IRs flanking similar tandemly arranged ORFs (Fig. 1). A sig-
nificant structural difference between the two elements is that 
Tn5469 encodes a second unidentified ORF (designated orf2) 
immediately downstream of orf1 (Fig. 1B). Despite this differ-
ence, their collective structural and sequence similarities sug-
gest that Tn5469 and Tn5541 arose from a common ancestral 
transposon. Primarily on the basis of their unprecedented com-
posite transposase, but taking into account their uncharacter-
istic genetic structure and unidentifiable orf1, we propose that 
Tn5541 and Tn5469 represent a novel noncomposite transpo-
sion family.

Distribution of Tn5469 and Tn5541. The distribution of 
Tn5541 among several morphologically distinct strains of cya-
nobacteria was examined by DNA hybridization analysis. The 
examined strains were from the American Type Culture Col-
lection or the Pasteur Culture Collection and are referred to by 
their genus names followed by the collection numbers. In a 
low-stringency hybridization analysis of total DNA isolated 
from Anabaena sp. strain PCC 7120, Anabaena sp. strain 
ATCC 29413, F. diplosiphon UTEX 481, Nostoc sp. strain PCC 
8009, Synechococcus sp. strain PCC 7942, and Synechocystis 
sp. strain PCC 6803, a probe for Tn5541 hybridized only to 
DNA from F. diplosiphon UTEX 481 (data not shown). This result 
suggests that like most of the characterized cyanobacterial IS 
elements (1, 5, 22), neither Tn5469 nor Tn5541 is widely dis-
tributed among morphologically distinct genera.

We similarly examined the distribution of Tn5469 and 
Tn5541 among six distinct strains comprising the Tolypothrix 
1 group of heterocyst-forming filamentous cyanobacteria (24), 
which includes the F. diplosiphon UTEX 481 culture equiva-
lent, *Calothrix* sp. strain PCC 7601. Total DNA from *Calothrix* sp. strain PCC 7601 and *Tolypothrix* sp. strains PCC 7101, PCC 7504, PCC 7708, PCC 7710, and PCC 7712 was digested with SpeI and hybridized to a probe for Tn5469 (3.8-kbp EcoRI-XbaI fragment from pUMC227) or Tn5541 (1.8-kbp EcoRI fragment from pUMC334). Among the strains examined, DNA from *Calothrix* sp. strain PCC 7601 and *Tolypothrix* sp. strains PCC 7504, PCC 7710, and PCC 7712 hybridized to the Tn5469 probe (Fig. 4A). In comparison, only DNA from *Calothrix* sp. strain PCC 7601 and *Tolypothrix* sp. strains PCC 7710 and PCC 7712 hybridized to the Tn5541 probe (Fig. 4B). The weaker hybridization signal obtained for *Tolypothrix* sp. strain PCC 7504 with the Tn5469 probe (Fig. 4A, lane 3) most likely reflects the presence of a heterologous, cross-hybridizing transposon or transposase gene. If correct, Tn5541 and Tn5469 appear to be limited to the same three strains within the *Tolypothrix* I cluster. The simplest explanation for this distribution is that *Calothrix* sp. strain PCC 7601 and *Tolypothrix* sp. strains PCC 7710 and PCC 7712 have a common ancestor that acquired both Tn5541 and Tn5469 prior to diverging from *Tolypothrix* sp. strains PCC 7101, PCC 7504, and PCC 7708.

The DNA hybridization analysis also showed that *Calothrix* sp. strain PCC 7601 and *Tolypothrix* sp. strains PCC 7710 and PCC 7712 harbor multiple genomic copies of both Tn5469 and Tn5541. *Calothrix* sp. strain PCC 7601 harbors five genomic copies of Tn5469; each copy is detected as a single fragment in the SpeI digest (Fig. 4A, lane 1). An identical Tn5469 hybridization profile was observed for *Tolypothrix* sp. strain PCC 7710 (Fig. 4A, compare lanes 1 and 5). The Tn5469 hybridization profile for *Tolypothrix* sp. strain PCC 7712 differed from that of *Calothrix* sp. strain PCC 7601 and *Tolypothrix* sp. strain PCC 7710 only by the absence of the hybridizing 15-kbp SpeI fragment (Fig. 4A, compare lane 6 with lanes 1 and 5). With the Tn5541 probe, five hybridizing fragments of high signal intensity and several fragments of low signal intensity were detected for *Calothrix* sp. strain PCC 7601 (Fig. 4B, lane 1). A similar analysis of the same DNA digested with Clal or Xbal (data not shown) supported five genomic copies of Tn5541 for this strain. For both *Tolypothrix* sp. strains PCC 7710 and PCC 7712, five genomic copies of Tn5541 were detected (Fig. 4B, lanes 5 and 6). The Tn5541 hybridization profiles for the three host strains appeared to have some common fragments (Fig. 4B, compare lanes 1, 5, and 6); however, in contrast to what was observed for Tn5469, the Tn5541 hybridization profile for *Calothrix* sp. strain PCC 7601 was not identical to that for *Tolypothrix* sp. strain PCC 7710 (Fig. 4B, compare lanes 1 and 5). In addition, whereas the Tn5469 hybridization profile for *Tolypothrix* sp. strain PCC 7712 differed from that of *Tolypothrix* sp. strain PCC 7710 by the presence of the 15-kbp hybridizing fragment (Fig. 4A, compare lanes 5 and 6), the Tn5541 hybridization profiles for these two strains differed only by the size of the largest hybridizing fragment (Fig. 4B, compare lanes 5 and 6).

The similar Tn5469 and Tn5541 hybridization profiles support a close phylogenetic relationship between the three host strains in terms of genome structure. We recently determined nearly identical DNA sequences for the five genomic copies of Tn5469 isolated from *Calothrix* sp. strain PCC 7601 (8), suggesting that all rapidly radiated from a single form. The multiple genomic copies of Tn5541 were presumably achieved by a similar process. An intriguing possibility is that Tn5541 and Tn5469 were simultaneously acquired by an ancestral host, and that because of their relatedness, the two elements multiplied in parallel via a common transposition mechanism; *Calothrix* sp. strain PCC 7601 and *Tolypothrix* sp. strains PCC 7710 and PCC 7712 would have subsequently arisen from this ancestral host.

The narrow host range of Tn5469 and Tn5541 supports the utility of either element as a genetic tool for mutagenic analysis of nonhost cyanobacterial strains. The development of such a tool would facilitate molecular genetic studies on strains which, like *F. diplosiphon*, are not amenable to characterized trans-
poson mutagenesis protocols. Toward this end, we have initiated the development of an assay for Tn5469 transposition in several morphologically distinct strains that may allow us to determine if a selectable form of the element can serve as an agent for transposon tagging.

Common transposition mechanism for Tn5469 and Tn5541? Little is known regarding the transposition of Tn5469 or Tn5541. For the better-characterized Tn5469, a replicative transposition mechanism was hypothesized (16) based on the following observations: (i) the source strain Fd33 harbors five genomic copies of the element, (ii) Fd33-derived primary and secondary mutants respectively harboring six and seven genomic copies of Tn5469 have been characterized, and (iii) no decrease in Tn5469 copy number or change in location of a preexisting element has been observed. However, a nonreplicative transposition event followed by the loss of the donor molecule can lead to transposon accumulation (20). Because F. diplosiphon harbors multiple cellular copies of its genome (28), a nonreplicative mechanism for Tn5469 cannot be ruled out. Tn5541 has in common with Tn5469 the property of genomic multiplicity; however, an increase in Tn5541 multiplicity has not been documented for any mutant or host strain in our collection. Nevertheless, their genetic and structural relatedness suggests that Tn5541 and Tn5469 have a common, and probably replicative, transposition mechanism. Of particular interest is whether the transposase from Tn5541 can sponsor transposition of Tn5469 or vice versa.

Nucleotide sequence accession number. The nucleotide sequence of Tn5541 has been deposited in the GenBank database under accession no. AF072896.

We thank P. Shubert and J. Salamisina for excellent technical assistance and N. Tandeau de Marsac and R. Rippka for providing cyanobacterial strains. We are grateful to W. K. Thomas for helpful discussions and comments.

This research was supported by National Science Foundation grant MCB-9513660. K.K. was supported by a UMKC Chancellor’s Interdisciplinary Ph.D. Fellowship.

REFERENCES
27. Tandeau de Marsac, N. Personal communication.