Evolutionary affiliation of the marine nitrogen-fixing cyanobacterium \textit{Trichodesmium} sp. strain NIBB 1067, derived by 16S ribosomal RNA sequence analysis

Annick Wilmotte,† Jean-Marc Neefs and Rupert De Wachter

Author for correspondence: Rupert De Wachter. Tel: +32 3 820 23 19. Fax: +32 3 820 22 48. e-mail: DWACHTER@REK.UL.BE

Department of Biochemistry, University of Antwerp (UIA), Universiteitsplein 1, B-2610 Antwerp, Belgium

The 16S rRNA sequence of \textit{Trichodesmium} sp. strain NIBB 1067 was determined and used for the construction of a distance tree and bootstrap analysis. The tree shows that, among the available cyanobacterial 16S rRNA sequences, \textit{Trichodesmium} NIBB 1067 has \textit{Oscillatoria} PCC 7515 as its closest relative, presenting 94.9\% of sequence similarity with the latter strain. This is in contrast to a difference of 9 mol\% G + C in mean genomic DNA base composition between the two organisms. Nevertheless, the genotypic heterogeneity presented by a number of strains assigned to the genus \textit{Oscillatoria} hinders a taxonomic decision on the separate existence of the genera \textit{Trichodesmium} and \textit{Oscillatoria}. The sequence of the internal transcribed spacer (ITS) between the 16S and 23S rRNA genes was also determined, as a possible marker to study inter- and intraspecific variability. The ITS contains the genes coding for tRNA^Ii and tRNA^Aia and its total length is 547 nucleotides. In six out of eight sequenced clones, there is a duplication of 29 nucleotides, surrounding the 5’ end of the tRNA^Ii.

\textbf{Keywords:} \textit{Trichodesmium}, 16S rRNA sequence analysis, taxonomy, cyanobacteria, internal transcribed spacer

\section*{INTRODUCTION}

\textit{Trichodesmium} species are marine planktonic cyanobacteria which produce extensive blooms at the surface of tropical and subtropical oceanic waters. Their morphology is typical of the filamentous \textit{Oscillatoriaceae}, which possess simple trichomes with cells dividing in a single plane. The ultrastructure of \textit{Trichodesmium erythraeum} (Van Baalen & Brown, 1969) is quite similar to that of \textit{Trichodesmium thielii} (Gantt et al., 1984) and both exhibit the \textit{Oscillatoria}-type wall features (Drews & Weckesser, 1982). They contain gas vesicles to regulate their buoyancy (Walsby, 1981). The pigments present are phycocyanin and a phycourobilin-containing phycoerythrin (Fogg, 1987).

Although they do not form heterocysts, \textit{Trichodesmium} species are able to fix nitrogen aerobically in the light, and may be responsible for up to a quarter of the total nitrogen fixation in the oceans (Capone & Carpenter, 1982). The contiguous arrangement of nitrogenase genes in \textit{Trichodesmium} sp. strain NIBB 1067 is typical of non-heterocystous cyanobacteria (Zehr et al., 1991a).

The taxonomy of \textit{Trichodesmium} and its relation to the genus \textit{Oscillatoria} is problematic. Different opinions can be found in the literature, depending on the taxonomic weight given to the two morphological features which can be used to separate the two genera: (1) aggregation in macroscopic colonies forming bundles with parallel or radial arrangement of trichomes held together by a thin mucilaginous layer; and (2) gas vacuoles arranged as peripheral cylinders and made up of unusually strong gas vesicles (Walsby, 1981; Fogg, 1987). Umezaki (1974) observed that \textit{Trichodesmium} contains more fatty acids of low molecular mass than \textit{Oscillatoria} species usually do.
The overall morphological similarity of the two genera has prompted some authors (Geitler, 1932; Sournia, 1968) to transfer *Trichodesmium* to *Oscillatoria*, whereas others (Gomont, 1892; Golubić, 1977; Anagnostidis & Komárek, 1988; Castenholz, 1989) have supported the retention of the genus *Trichodesmium*. The evolutionary affilitation of the genus *Trichodesmium* and its systematics are thus still a matter of debate.

Cultivation of *Trichodesmium* species is extremely difficult and until now only the strain NIBB 1067, isolated from Kuroshio waters (Japan), has been grown successfully in laboratory culture (Okhi *et al.*, 1986). A sequence of 359 nucleotides from the *nifH* gene of this cultivated strain presents a similarity of 98% with the homologous sequence from a natural assemblage of *T. thiebautii* from the Caribbean Sea (Zehr *et al.*, 1990). To determine the genotypic relationship of this strain with other cyanobacteria, the 16S rRNA sequence of strain NIBB 1067 was determined and used to build a distance tree containing all available complete and partial cyanobacterial sequences having at least 686 positions. In addition, the sequence of the internal transcribed spacer (ITS) following the 16S rRNA gene was determined.

### METHODS

Genomic DNA was kindly provided by Dr K. Okhi, National Institute of Basic Biology, Okazaki, Japan. Amplification by the polymerase chain reaction (PCR) of a fragment of the rRNA operon comprising the 16S rRNA, the ITS and the 5' end of the 23S rRNA, was carried out with primers 1 and 18 (Wilmotte *et al.*, 1993). In a total volume of 100 µl, the PCR mix contained about 5 ng genomic DNA, 2.8 µg of each dNTP, 25 U of Taq DNA polymerase, 10 mM Tris/HCl, 50 mM KCl, 1.5 mM MgCl₂, 0.01% (w/v) gelatin, pH 8.3. An initial denaturation of 1 min at 94°C was followed by 25 cycles comprising 1 min at 94°C, 1 min at 55°C and 3 min at 72°C. An incubation of 7 min at 72°C was added to the last cycle. The PCR product was purified by gel electrophoresis in agarose and cloned into vector pSK+ (Stratagene) to which a ddT-tail had been added at the 3' end by terminal deoxynucleotidyl transferase (Holton & Graham, 1991). Colony hybridization was performed after transformation into competent *Escherichia coli* strain DH5α, as described previously (Wilmotte *et al.*, 1993).

From 227 colonies which hybridized with the PCR product, eight recombinant clones were randomly chosen and pooled for the sequencing reactions. The sequence determination was carried out using the Sequenase 2.0 kit (USB) following the manufacturer’s instructions. Eighteen primers annealing to evolutionarily conserved areas (Wilmotte *et al.*, 1993), a primer with the sequence CGCTCTACCATCTAGCTA annealing to a region of the rRNA*α*-prime (primer 19), and two primers complementary to the M13 sites on the vector were used to sequence both strands of the 16S rRNA gene, the ITS and the 5' end of the 23S rRNA gene. In addition, the first 350 nucleotides of the ITS sequence were determined separately for all eight clones using the *Taq* cycle-sequencing kit (USB) with primer 14 (Wilmotte *et al.*, 1993). The latter kit was also used to sequence directly the first 490 nucleotides of the ITS with primers 14, 17 and 19 after purification of the PCR product by electrophoresis.

The new 16S rRNA sequence was fitted into an alignment (Neefs *et al.*, 1993) containing all published complete or nearly complete bacterial 16S rRNA sequences, plus a number of partial cyanobacterial 16S rRNA sequences. For each pair of aligned sequences, distance values were calculated, corrected for multiple mutations by means of the formula of Jukes & Cantor (1969) as previously described (Van de Peer *et al.*, 1990), and entered into a distance matrix. A distance tree was constructed by means of the neighbour-joining method of Saitou & Nei (1987). Confidence values for individual nodes of the resulting tree were determined by a bootstrap analysis (Felsenstein, 1985), in which 500 bootstrap trees were constructed using randomly resampled data. This data analysis was carried out with the software package TREECON (Van de Peer & De Wachter, 1993).

### RESULTS AND DISCUSSION

The complete 16S rRNA sequence was obtained, except for an estimated 27 nucleotides at the 5' end. This corresponds to 1471 nucleotides.

Fig. 1 shows a distance tree constructed from all cyanobacterial 16S rRNAs for which the complete sequence or an appreciable fraction of it is known. All available cyanobacterial sequences were used because selection of strains may cause changes in the topology of the tree. In this tree, *Trichodesmium* NIBB 1067 appears closely related to *Oscillatoria* PCC 7515, for which a partial 16S rRNA gene sequence is known (Giovannoni *et al.*, 1988). The grouping of the two strains is statistically supported at a level of 100% by the bootstrap analysis. *Oscillatoria* PCC 7515 is quite different in morphology and physiology from *Trichodesmium* sp., being a freshwater strain isolated from a greenhouse water tank in Sweden, with discoid cells that have a diameter of 15 µm. It is able to synthesize nitrogenase anaerobically (Rippka *et al.*, 1979), and is the reference strain of the genus *Oscillatoria* as defined in Bergey’s Manual of Systematic Bacteriology (Castenholz, 1989). By comparison, the average cell width of *Trichodesmium* NIBB 1067 is 7±5 µm and the cell length varies between 54 and 57 µm in culture (K. Okhi, personal communication). Although the mean DNA base composition of *Oscillatoria* PCC 7515 is 40 mol % G+C (Herdman *et al.*, 1979) and is greater than that (31 mol %) of *Trichodesmium* NIBB 1067 (Zehr *et al.*, 1991b), the base compositions of the determined 16S rRNA sequences are very similar (55-3 and 54-3 mol % G+C, respectively) and fall within the variation range observed for 12 cyanobacterial sequences (52-7±56 mol %) (unpublished data).

For the 994 nucleotides available for comparison, the 16S rRNA sequence similarity between the two strains is 94-9%. There is not yet a well-agreed relation between 16S rRNA sequence similarities and taxonomic levels, whereas DNA–DNA hybridization percentages are currently used to recognize conspecific and congeneric proteobacteria (Wayne *et al.*, 1987). However, a 16S rRNA sequence similarity of 94-9% is clearly too low to indicate conspecificity. We propose to retain the generic name *Trichodesmium* for strain NIBB 1067 because of the difference in mean DNA base composition and because it would not be practical to add this well-characterized genus to a taxon with a problematic phylogeny. Indeed,
the dispersal of the Oscillatoria strains in different lineages (Fig. 1) indicates that this genus is genotypically heterogeneous.

The ITS sequence given in Fig. 2 contains genes coding for RNA16S and RNA5S, as observed in several eubacteria including the cyanobacterium Synechoococcus sp. strain PCC 6301 (Tomioka & Suguri, 1984), and in plastids (Graf et al., 1980; Ohyama et al., 1986). In six out of the eight independently sequenced clones, it shows a duplication of 29 nucleotides involving 14 nucleotides upstream of the RNA16S and the first 15 nucleotides of the RNA5S. In addition, the direct cycle sequencing of the PCR product confirmed that there was a mixture of ITS sequences with and without duplication because the sequence obtained with primers situated upstream and downstream of the duplication became unreadable beyond one copy of the duplicated sequence. Other differences among the eight independently sequenced clones were found at positions 191, 328 and 339 (Fig. 2). All of these were observed in single clones, and they may represent microheterogeneities or misincorporations by the Taq polymerase, except for the difference at position 328 which was shared by the two clones without duplication.

Fig. 1. Distance tree based on partial 16S rRNA sequence analysis of Trichodesmium sp. strain NIBB 1067, and 45 other complete and partial cyanobacterial sequences published to date (Britschgi & Giovannoni, 1991; Giovannoni et al., 1988, 1990; Ligon et al., 1991; Tomioka & Suguri, 1983; Turner et al., 1989; Urbach et al., 1992; Weller et al., 1991, 1992; Wilmotte et al., 1992, 1993), Marchantia polymorpha plastid (Ohyama et al., 1986), Cyanophora paradoxa cyanelle (V.L. Stirewalt & D.A. Bryant, unpublished data) and Escherichia coli (Brosius et al., 1981) which was used to root the tree. Names beginning with SAR and OS designate organisms isolated directly from environmental samples (Britschgi & Giovannoni, 1991; Giovannoni et al., 1990; Weller et al., 1991, 1992). From the alignment, available from the authors, 686 positions common to the 49 sequences were used for tree construction. They correspond to Escherichia coli positions 254-507, 676-892, 1121-1312. The distance between two organisms, expressed in substitutions per nucleotide, is obtained by summing the lengths of the horizontal branches connecting them, using the scale at the top. Numbers at nodes indicate the frequency with which the cluster descending from that node was found in the 500 bootstrap trees. Only bootstrap percentages higher than 50% are given. The position of the root was chosen to equalize its distance to the outgroup, E. coli, and its average distance to all other included organisms.
ITS (Fig. 2) is noteworthy because it is also observed in seven other cyanobacterial ITS sequences (our unpublished data). This sequence conservation in the middle of a variable region may indicate that it has a function, such as in the processing of the rRNA precursor.

Sequence and length are more variable for the ITS than for the 16S rRNA gene, and species specific probes inferred from the ITS sequence have been used for bacterial identification (Rossau et al., 1992). Therefore, ITS sequences may be useful tools to study inter- and intraspecific variability.

ACKNOWLEDGEMENTS

We thank Dr Kaori Ohki (National Institute of Basic Biology, Japan) for the kind gift of genomic DNA from Trichodesmium NIBB 1067 and information, and two anonymous referees for constructive criticism. A. W. was senior research assistant of the National Fund for Scientific Research during this work.

This research was supported by the Fund for Joint Basic Research and performed in the framework of the Institute for the Study of Biological Evolution of the University of Antwerp.

REFERENCES


16S ribosomal RNA sequence of *Trichodesmium* sp.


Zehr, J. P., Limberger, R. J., Okhi, K. & Fujita, Y. (1990). Antiserum to nitrogenase generated from an amplified DNA fragment from


Received 19 November 1993; revised 11 February 1994; accepted 22 February 1994.