Form species *Nostoc commune* (Cyanobacteria)

^{1,2,3} Virginia Tech Center for Genomics (VIGEN)¹, Department of Biochemistry² and Fralin Biotechnology Center³, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

Deborah Wright,^{1,2} Todd Prickett,^{1,2} Richard F. Helm^{2,3} and Malcolm Potts^{1,2}

Author for correspondence: Malcolm Potts. Tel: +1 540 231 5745. Fax: +1 540 231 9070. e-mail: geordie@vt.edu

The form species concept for the Cyanobacteria was evaluated using a comprehensive set of *Nostoc* samples that were collected during the past two centuries, from all continents, including regions from the Tropics to the Poles. Phylogenies were constructed based upon the conserved regions of tRNA^{Leu} (UAA) group I intron DNA sequences. Thirty-four forms contained a tRNA^{Leu} (UAA) intron of 284 nt. These 284-nt introns contained 200 nt of conserved sequence that, in most cases, shared 100% sequence identity, they had three variable regions (I, II and III) amounting to 84 nt, contained no hypervariable region and formed a discrete cluster in phylogenetic analysis. These forms represented 31 independent populations in both hemispheres and constitute examples of form species *Nostoc commune*. Multiple introns were obtained from several of the populations. Ten populations contained introns of 287–340 nt with a hypervariable region, 8 to 59 nt in length, located between variable regions I and II. Alignments identified 15 examples where 5'-AAAAUCC-3' occurred at the hypervariable region-variable region II boundary; this sequence is identical to the conserved sequence at the 3' intron-exon boundary (splice site) within the tRNA^{Leu} (UAA) gene. The possibility that hypervariable regions were removed from the primary intron through secondary splicing was tested *in vitro* but proved to be negative under the experimental conditions used. Shared morphologies of genetically different strains, dissimilar morphologies in strains that share identical genetic markers, incorrect naming of culture collection strains and genetic drift in cultured strains emphasize that the successful delineation of cyanobacterial species requires the application of multiple taxonomic criteria.

Keywords: *Nostoc commune*, cyanobacteria, species concept

INTRODUCTION

The concept of a prokaryotic species and an understanding of the influence and consequence of environmental selection pressure on prokaryotic diversity and evolution remain intractable issues. Considerations of these issues are especially exacerbated when dealing with the Cyanobacteria (Castenholz, 1992). There are three taxonomic schemes for cyanobacteria in current use: *sensu* Geitler (1932), Stanier

(see Rippka *et al.*, 1979) and Komárek & Anagnostidis (1989). All three are phenetic schemes that place different emphasis on aspects of the morphology, biochemistry, physiology and DNA homologies of field communities and pure cultures. Depending on the context of usage, each scheme has its own shortcomings and advantages (Whitton & Potts, 2000).

The application of molecular tools has the potential to resolve much of the controversy over cyanobacterial taxonomy, evolution and the species concept (Wilmotte, 1994). Recent approaches include DNA–DNA hybridization (Wilmotte *et al.*, 1997), fingerprinting based upon PCR assay with primers from short and long tandemly repeated elements (Rasmussen & Svenning, 1998), classification of clone cultures based upon sequences from the variable

Abbreviations: DEPC, diethyl pyrocarbonate; PTB, phenacylthiazolium bromide.

The GenBank accession numbers for the tRNA^{Leu} (UAA) group I intron sequences determined in this study are AF204090, AF204098 and AF204064–AF204106.

regions V6, V7 and V8 of 16S rRNA (Rudi *et al.*, 1997), amplified rDNA restriction analysis of the internally transcribed spacer (Scheldeman *et al.*, 1999) and comparisons of the conserved structural and regulatory domains within divergent 16S rRNA–23S rRNA spacer sequences (Iteman *et al.*, 2000).

One question that remains particularly problematic is the role of gene exchange (lateral transfer), recombination and hyperbradytely (stasis; slow rate of change) in cyanobacterial evolution (Castenholz, 1992; Rudi *et al.*, 1998; Schopf, 2000). Other important questions are how far molecular studies will confirm the taxa recognized at present (from both bacteriological and classical methods) and how soon it will be possible to use molecular approaches routinely as a rapid means of characterizing cells from natural populations (Whitton & Potts, 2000). Unfortunately, the conclusions that derive from such molecular studies, and the various ‘phylogenetic’ trees illustrating similarities based on sequences from particular parts of the genome, often use results from strains whose generic and specific names are doubtful. In most cases, these strains have been maintained in culture for decades, they have lost morphological and physiological properties of determinative value and many are of unknown origin. This restricts the value of such trees and analyses for comparing possible evolutionary relationships with the taxonomic relationships suggested by classical names (Whitton & Potts, 2000).

One feature of cyanobacterial populations is the frequent dominance of a single form with very characteristic morphology. Examples of such morphotypes include *Trichodesmium erythraeum*, *Aphanothece halophytica*, *Microcoleus chthonoplastes*, *Arthrospira* sp. and *Nostoc commune* (Garcia-Pichel *et al.*, 1996; Potts, 2000; Scheldeman *et al.*, 1999). Because these cyanobacteria often have global distributions and can be readily identified in diverse geographical localities, the concept of a ‘form species’ developed (Castenholz, 1992). The concept raises important questions to do with dispersal, isolation, selection and response to environmental conditions. However, the ability to address these problems is compromised by the limitations of cyanobacterial taxonomy and by uncertainty over the spatial and temporal genetic consistency of ‘form species’ (Whitton & Potts, 2000). In the case of hot-spring communities, for example, recent studies suggest that their genetic diversity is much more complex than previously thought (Ward & Castenholz, 2000).

The potential of group I intron analysis as a genetic marker and taxonomic tool for cyanobacteria has been evaluated in several studies. Eubacterial group I introns were first identified in cyanobacteria (Kuhse *et al.*, 1990; Xu *et al.*, 1990); their distribution within this group appears to be broad and includes a range of morphologically diverse forms including *Microcystis*, *Scytonema*, *Gloeobacter* and *Nostoc* (Paquin *et al.*, 1997; Paulsrud & Lindblad, 1998; Rudi & Jakobsen,

1999). Group I introns interrupt protein-coding regions and RNA genes in eukaryotes and bacteriophages and are present in the anticodon loop of tRNA genes (Paquin *et al.*, 1999) as well as in 23S rRNA in eubacteria (Everett *et al.*, 1999). The acquisition, transfer and distribution of introns is of phylogenetic significance and the source of some controversy (e.g. Gilbert & Glynias, 1993). The presence of homologous tRNA^{Leu} (UAA) introns in cyanobacteria and plastids was considered evidence of an ancient origin for these sequences (Kuhse *et al.*, 1990; Xu *et al.*, 1990). Subsequent studies, however, provided evidence for lateral transfer of tRNA^{Leu} (UAA) introns in strains of *Microcystis* and thus a more recent, possibly polyphyletic, origin was proposed (Rudi & Jakobsen, 1997). From an investigation of the diversity of *Nostoc* cyanobionts in lichens, it was concluded that the tRNA^{Leu} (UAA) intron can be of great value when examining cyanobacterial diversity (Paulsrud & Lindblad, 1998; Costa *et al.*, 1999). A large, variable region in the *Nostoc* introns, which corresponded to bases 99–143 in *Anabaena* sp. PCC 7120 tRNA^{Leu} (UAA), exhibited either of two separate themes. It was suggested that the two classes of intron may correlate with different physiological capacities of the cyanobionts, although there were insufficient differences within the conserved regions of the intron to support or reject this proposal. In a recent study, these same authors used tRNA^{Leu} (UAA) intron analysis to study geographical patterns of diversity in *Nostoc*-containing lichens (Paulsrud *et al.*, 2000).

The genus *Nostoc* is one of five genera in the family Nostocaceae of subgroup 4, section A, of the oxygenic phototrophic bacteria (Holt *et al.*, 1993). *Nostoc commune* is a conspicuous component of terrestrial microbial populations worldwide, especially those associated with nutrient-poor soils and limestones, where growths may achieve macroscopic proportions (Helm *et al.*, 2000; Hill *et al.*, 1994a, b; Potts, 2000; Shirkey *et al.*, 2000). Records indicate that *N. commune* has been collected and studied for at least the past 2000 years (Potts, 1997, 2000). Numerous samples of *Nostoc* spp. are present in herbaria, many are still viable and most are accompanied by detailed records of the time and place of collection. There are few other bacteria for which such detailed information is available and, with respect to cyanobacterial diversity, such collections constitute a valuable resource.

The complete genome sequence of *Anabaena* sp. PCC 7120 became available in February 2000 (unannotated format; available at <http://www.kazusa.or.jp/cyano/anabaena/>) and that of *Nostoc punctiforme* ATCC 29133 is being annotated (available at <http://genome.ornl.gov/microbial/npun/>). It is important to understand the phylogenetic relationships between these, and other, forms of ecologically relevant cyanobacteria. In this context, the present study provides the most comprehensive dataset to date on forms of *Nostoc*. A diverse and comprehensive set of *Nostoc* samples were used: (i) to evaluate the form species

concept in cyanobacteria; (ii) to evaluate the utility of group I intron analysis in the taxonomy of *Nostoc*; (iii) to understand the potential influence of geographical factors on intron dispersal and evolution; and (iv) to examine phylogenetic relationships between *Anabaena* and *Nostoc* species.

METHODS

Biological materials. Colonies of *N. commune* were collected by the authors and their colleagues and stored desiccated, in the dark, until analysis (Table 1). Many of the herbarium specimens were obtained in sealed paper envelopes that had not been opened since the time of collection. Cultures of *N. commune* DRH1, *N. commune* UTEX 584, *N. punctiforme* ATCC 29133 and *Anabaena* sp. strain PCC 7120 are maintained in laboratory culture. The latter strain is considered a nomen species of *Nostoc* sp. PCC 6705 and 6719 (Rippka, 1988). The designation *Anabaena* sp. PCC 7120 is retained in this study (see Discussion).

Isolation of genomic DNA

Large scale. Desiccated samples were lyophilized and ground to a powder under liquid nitrogen. The powder was rehydrated with TE (1 mM EDTA, 10 mM Tris/HCl, pH 7.5; approximately 1:10 powder/buffer). Extraction buffer [1.4 M NaCl, 20 mM EDTA, 1.5% (w/v) CTAB, 1% (v/v) β -mercaptoethanol, 100 mM Tris/HCl, pH 7.5; approximately 4 vols of TE added to the powder] was added and the mixture was incubated at 65 °C for 30 min. The slurry was homogenized, frozen under liquid nitrogen and thawed at 65 °C. The freeze-thaw cycle was repeated six times. The mixture was extracted with chloroform/isoamyl alcohol (24:1) and centrifuged and the aqueous phase was mixed with an equal volume of 2-propanol. DNA was recovered and stored in a minimal volume of TE buffer.

Small scale. When limited amounts of material were available, the desiccated colonies were ground to a powder under liquid nitrogen. Samples were then added directly to PCR tubes and heated at 95 °C for up to 4 h. Rehydrated cells and/or the cell-free supernatant fractions were used in PCR amplifications (see below).

Pre-treatment of modified DNA. After long-term desiccation, the DNA in *Nostoc* cells may become covalently modified through Maillard reactions (Potts, 2000) and this can contribute to subsequent failures in PCR amplification (data not shown). If PCR amplification proved negative in preliminary trials, the genomic DNA was treated with *N*-phenacylthiazolium bromide (PTB) before assay. The PTB reagent was synthesized as described previously (Poinar *et al.*, 1998; Vasan *et al.*, 1996). Desiccated colonies were ground to a powder under liquid nitrogen and resuspended in 50 mM EDTA. An equal volume of 1.5% (w/v) molten agarose (type 1-A, low EEO; Sigma cat. no. A-0169) was added to embed the cell material. The blocks were incubated overnight at 50 °C with 2 mg lysozyme ml⁻¹ (final concentration; Sigma grade 1) in 50 mM EDTA and then overnight at 50 °C with 2 mg proteinase K ml⁻¹ (final concentration; Sigma cat. no. P6556) in 0.1 M EDTA, 1% (w/v) *N*-lauroylsarcosine, 0.05% (w/v) SDS and 10 mM Tris/HCl, pH 8.0. Finally, the blocks were equilibrated in 10 mM sodium phosphate buffer, pH 7.0, prior to the addition of PTB at a final concentration of 10 mM. The blocks were incubated in PTB for 10 h at 27 °C (room temperature) and then equilibrated in TAE buffer (2 mM EDTA, 40 mM Tris/acetate). The blocks were transferred

to the wells of a 1% (w/v) agarose gel (type 1-A, low EEO) and the DNA was resolved at 75 mA for 1 h. DNA was excised from the gel and purified further using a QIAEX II gel extraction kit (Qiagen).

Oligonucleotide primers. Primers for intron amplification were based upon the sequence of tRNA^{Leu} (UAA) of *Anabaena* sp. PCC 7120 (Zaug *et al.*, 1993). Primer LEUI1 corresponds to a region of the D loop (bases 6–28; 5'-TGTGGCGGAATGGTAGACGCTAC-3'); primers LEUI2 and LEUI3 are based on the reverse complement of regions within the T ψ C and variable loops (bases 72–56; 5'-GACTTGAACCCACACGAC-3') and the T ψ C loop and acceptor stem (bases 67–82; 5'-GGGTGGAGGGA-CTTGA-3'), respectively.

PCR assay. Assays were performed in 50 μ l reactions, in pH 8.5 buffer [60 mM Tris/HCl, 15 mM (NH₄)₂SO₄]. Reactions contained 12.5 nmol of each dNTP (each dNTP at 250 μ M), 3.5 mM Mg²⁺ and 2.5 U *Taq* polymerase (Promega Biotec). The temperature was controlled with a Biometra thermocycler model T-3 (Labrepco). All assays began with a denaturation step at 95 °C for 2 min and ended with an elongation time of 10 min; in each cycle, denaturation occurred at 95 °C for 1 min and elongation was at 72 °C for 90 s. The annealing temperature (60 s) was 66 °C in the first cycle and dropped by 0.2 °C per cycle thereafter for an additional 39 cycles.

The experiments reported here were performed over a period of approximately 2 years. In view of the resistance of the cells (and their DNAs) to air-drying and their probable persistence in aerosols, stringent precautions were used to circumvent cross-contamination. Conditions for PCR assay were optimized in preliminary trials using genomic DNA from *Anabaena* sp. strain PCC 7120 and *N. commune* DRH1. These DNAs served as positive controls in every assay. Negative controls were used to evaluate possible contamination of the samples, reagents and buffers with occult DNA. Contamination was identified on two occasions. An ambiguity in the published sequence of the *Anabaena* sp. strain PCC 7120 tRNA^{Leu} (UAA) intron (R) at position 148 (GenBank accession no. M38692) was resolved as A. The group I intron sequence of *N. punctiforme* ATCC 29133 (= *Nostoc* sp. strain PCC 73102) was obtained to compare with the published sequence for control purposes (accession numbers AF019924 and U83254).

Subcloning and DNA sequencing. A minimum of three separate PCR assays were performed on each sample. Further assays were performed on samples that either provided more than one type of intron sequence or failed to generate a product in initial trials. Reaction mixtures were resolved in agarose gels and putative amplification products were excised from the gel, the agar was crushed in a minimal volume of water and the DNA was allowed to diffuse passively overnight. The DNA solutions were used directly in ligation reactions with pCR2.1-TOPO (Invitrogen) that were then used to transform *Escherichia coli* TOP10 (Invitrogen). Purified plasmid templates were subjected to DNA sequence analysis using the BigDye Terminator chemistry of Perkin-Elmer Applied Biosystems and *Taq* DNA polymerase on an Applied Biosystems 377 Prism DNA sequencer. For each organism, replicate samples of the colony were used, the intron DNAs that derived from three separate PCR amplifications were sequenced completely on both strands and all six sequences for a given sample were aligned (see below). In only a few cases, there were between one and three discrepancies within the aligned DNA sequence of one of the three clones of a given sample. These

Table 1. Origin of materials and source histories

With the exception of the WH and NSW series of samples, reference numbers include a mnemonic of the place of origin and the year the sample was desiccated. Samples obtained from the Wien Herbarium, Austria, and the Victoria Herbarium, Australia, have the designations WH and VH, respectively, together with the reference number provided by the herbarium. The NSW series of samples were provided by Stephen Skinner. For *N. punctiforme* ATCC 29133, see Rippka *et al.* (1979); for *N. commune* UTEX 584 and *N. commune* DRH1, see Potts (2000). NK, Not known.

Designation	Reference/strain no.	Origin	Source	Year of collection
Strains (in culture)				
<i>N. punctiforme</i>	ATCC 29133	Macrozamia, Australia	ATCC	1973
<i>N. commune</i>	UTEX 584	UK?	T. Gibson	NK
<i>N. commune</i>	DRH1	CHEN/1986	D. R. Hill	NK
<i>Anabaena</i> sp.	PCC 7120	USA?	J. Elhai	NK
Desiccated samples				
Europe				
<i>N. commune</i> Vaucher	MAL/1989	Malham Tarn, UK	M. Potts	1989
<i>N. commune</i> Vaucher	BER/1988	Bermatingen, Switzerland	S. Scherer	1988
<i>N. commune</i> Vaucher	TAG/1988	Tägerwilen, Switzerland	S. Scherer	1988
<i>N. commune</i> Vaucher	WH001	Cojocna, Romania	WH3132	1923
<i>N. commune</i> Vaucher	WH002	Cahul, Romania	WH3132	1939
<i>N. commune</i> Vaucher	WH003	Caras, Romania	WH3132	1941
<i>N. commune</i>	WH004	Heidelberg, Germany	WH2622	1849
<i>N. commune</i>	WH005	Konstanz, Germany	WH2622	1859
<i>N. commune</i>	WH006	Germany	WH2622	1878
<i>N. commune</i>	WH007	Trieste, Italy	WH2922	1866
<i>Nostoc pellucidum</i>	WH016	Sesia a Vercelli, Italy	WH670	1865
<i>Nostoc alpicola</i> Kützing	WH011	NK	WH2525	1860
The Americas				
<i>N. commune</i>	WH009	Montevideo, Uruguay	WH2486	~ 1880s
<i>N. commune</i> var. <i>flagelliforme</i>	WH015	Zacatecas, Mexico	WH2407	1878
<i>N. commune</i> var. <i>flagelliforme</i>	WH008	San Pedro, TX, USA	WH2525	NK
<i>N. commune</i>	WH013	Winchester, MA, USA	WH	NK
<i>N. commune</i>	WH014	Lawrence Co., IN, USA	WH2009	1948
<i>N. commune</i> Vaucher	TOP/1993	Topsail Is., SC, USA	S. Scherer	1993
<i>N. commune</i> Vaucher	TEN/1988	Knoxville, TN, USA	W. Herndon	1988
<i>N. commune</i> Vaucher	BBC/1990	Blacksburg, VA, USA	M. Potts	1990
<i>N. commune</i> Vaucher	BBC/1992	Blacksburg, VA, USA	M. Potts	1992
<i>N. commune</i> Vaucher	ENG/1996	Blacksburg, VA, USA	M. Potts	1992
Antarctica				
<i>N. commune</i> Vaucher	MOA/1997	Mars Oasis, Alexander Is.	A. George	1997
<i>N. commune</i> Vaucher	RIS/1979	Ross Ice Shelf	M. Potts	1979
<i>Nostoc</i> sp.	LBP/1995	Lake Bonney Pond, Dry Valleys	H. Paerl	1995
Asia				
<i>N. commune</i> Vaucher	HUN/1982	Chiyang-Hunan, China	S. Scherer	1982
<i>N. commune</i> Vaucher	HUN/1986	Hunan, China	W.-Q. Xie	1986
<i>N. commune</i> Vaucher	CHEN/1986	Hunan, China	S. Scherer	1986
<i>Nostoc sphaericum</i>	SPH/1998	Health store, China	Z. Huang	1998
Australasia				
<i>N. commune</i>	WH012	Surabaya, Java, Indonesia	WH447	1880
<i>N. commune</i> Vaucher	MEL/1968	Victoria, Australia	VH698677	1968
<i>Nostoc</i> sp.	PORT/1999	New South Wales, Australia	PMC003	1999
<i>N. commune</i> Vaucher	RNP/1999	New South Wales, Australia	NSW 435425	1999
<i>N. commune</i> Vaucher	ARM/1999	New South Wales, Australia	NSW 435426	1999
<i>N. commune</i> var. <i>flagelliforme</i>	PARA/1979	New South Wales, Australia	NSW A3124	1979
<i>N. commune</i> Vaucher	NZE/1997	South Island, New Zealand	P. Broadie	1998

Table 1 (cont.)

Designation	Reference/strain no.	Origin	Source	Year of collection
Indo-Africa				
<i>N. commune</i> var. <i>lesler</i> Bornet	WH010	Colombo, Sri Lanka	WH2525	~ 1880s
<i>Nostoc</i> sp.	ALD776D	Aldabra Atoll, Indian Ocean	A. Donaldson	1973
<i>N. commune</i> Vaucher	ALD779D	Aldabra Atoll, Indian Ocean	A. Donaldson	1973
<i>N. commune</i> var. <i>flagelliforme</i>	ALD857D	Aldabra Atoll, Indian Ocean	A. Donaldson	1973
<i>Nostoc</i> sp.	ALD8122	Aldabra Atoll, Indian Ocean	M. Potts	1974

Table 2. Structural features of *Nostoc* introns

Values are lengths in nucleotides. For further details, see Fig. 2(b).

Designation	Intron	Conserved	Variable	Hypervariable
<i>N. commune</i> form species	284	200	84	—
ALD857DA/1973	298	200	65	33
ALD857DB/1973	298	200	65	33
WH9/1880	340	200	81	59
TOP/1993	340	200	81	59
ALD776DA/1973	325	201	91	33
ALD776DB/1973	325	201	91	33
ALD8122/1973	325	201	91	33
MALA/1989	300	200	66	34
MALB/1989	300	200	66	34
PARA/1979	287	200	60	27

transversions/transitions were attributed to infidelity in the PCR assay and the consensus sequence of the other two clones was thus reported.

In vitro transcription and intron splicing. All plasticware, glassware and solutions were treated with diethyl pyrocarbonate (DEPC) before use. Plasmids pDJW13 (containing the intron from *N. commune* UTEX 584) and pDJW37 (containing the intron from *N. commune* TOP/1993) were linearized with either *Hind*III or *Xba*I and dephosphorylated. Run-off transcription reactions contained (final concentrations): 5 µg template DNA, 40 mM Tris/HCl, pH 7.5, 5 mM MgCl₂, 10 mM DTT, 2 mM spermidine, NTPs (5 mM), 80 units RNAsin (Promega Biotec), 800 U T7 or SP6 RNA polymerase (Promega Biotec) and DEPC-treated water to a final volume of 50 µl. Reactions were incubated at 37 °C for 2 h. In order to increase transcript production, extra NTP mix (final 9 mM), MgCl₂ (final 5 mM) and balance of DEPC-treated water were added to the reaction mixture and reactions were incubated for a further 2 h. RQ1 RNase-free DNase (approx. 3–5 units; Promega) was added to the reaction and, after 15 min incubation at 37 °C, the mixture was extracted with phenol/chloroform and RNA was precipitated with ethanol.

Protocols for the assay of intron splicing were based upon the method of Zaug *et al.* (1993). Reactions contained 25 mM HEPES/NaOH, pH 7.5, 15 mM MgCl₂, the total RNA from one *in vitro* transcription reaction (in 15 µl 20 mM HEPES/NaOH, pH 7.5) and DEPC-treated water to a final volume of 18 µl. The reaction mixture was pre-

incubated at 50 °C for 15 min and then incubated at 32 °C for 2–3 min, and the splicing reaction was then initiated by the addition of 2 µl 10 mM guanosine (in 100 mM HEPES/NaOH, pH 7.5). The reaction was mixed gently and incubated at 32 °C for 60 min. Control reactions contained the same components with the exception of guanosine.

Northern analysis. RNAs in splicing reactions were resolved in formaldehyde/agarose gels, transferred to a positively charged nylon membrane (Boehringer Mannheim) through capillary blotting in 50 mM NaOH and cross-linked to the membrane using UV radiation. Oligonucleotides that corresponded to a region of the intron, hereafter referred to as the hypervariable region, were 3'-end-labelled with digoxigenin-dUTP for use as hybridization probes. The primers were 5'-ACTCTTAACCTCCTAACTCCTA-3' (to detect the intron in *N. commune* UTEX584) and 5'-ATCGTTTCGACTGAGCGTTCG-3' (to detect the intron in *N. commune* TOP/1993) and were used at a concentration of 25 pmol ml⁻¹ in hybridization buffer. Hybridization was at 50 °C overnight with final stringent washes of 0.1 × SSC and 0.1 % (w/v) SDS at 60 °C (Xie *et al.*, 1995). DNA-RNA hybrids were visualized using alkaline phosphatase-conjugated anti-digoxigenin antibody and the chemiluminescent detection system of Boehringer Mannheim.

Phylogenetic analysis. An alignment was constructed using the conserved secondary structure of introns and the conserved DNA sequence corresponding to bases 1–98, 144–218 and 223–249 of the *Anabaena* sp. PCC 7120 intron (see Paulsrud & Lindblad, 1998). The remaining regions of the introns, referred to as variable regions I, II and III

(including an especially poorly resolved 'hypervariable' region), and the 34 nt extensions (Table 2) in introns ALD776DA/1973, ALD776DB/1973 and ALD8122/1973, were not used in phylogenetic analysis. DNA sequences were aligned using MEGALIGN within version 4.0 of the LaserGene software (DNASTAR).

Phylogenetic trees were constructed based upon parsimony analysis (ordinary parsimony) and distance methods using the phylogenetic inference package (PHYLIP version 3.573c) obtained from J. Felsenstein (Department of Genetics, University of Washington, Seattle, WA, USA). SEQBOOT was used to produce 100 datasets from bootstrap resampling (Felsenstein, 1985). Majority-rule strict-consensus analysis was performed with CONSENSE. In view of its lack of distinguishing morphological characteristics (in comparison with the other strains studied), *Anabaena* sp. PCC 7120 was arbitrarily designated as the outgroup form. Distance matrices were obtained with Kimura's two-parameter model using the default transition/transversion ratio (Kimura, 1980) and calculated with the DNADIST and NEIGHBOR (Saitou & Nei, 1987) programs of PHYLIP (Felsenstein, 1985). Unrooted trees were plotted in DRAWTREE or DRAWGRAM and edited in Adobe Illustrator version 9.0.

Sequence folding. Intron sequences were analysed with reference to the *Anabaena* sp. PCC 7120 structure (Chech *et al.*, 1994). The structures of sequences that were not readily alignable were predicted using the *mfold* software version 3.1 available on-line at <http://bioinfo.math.rpi.edu/~mfold/rna/form1.cgi> (Zuker *et al.*, 1999).

RESULTS

PCR amplification of tRNA^{Leu} (UAA) group I introns

Primers based upon regions flanking the intron of the tRNA^{Leu} (UAA) gene of *Anabaena* sp. strain PCC 7120 amplified consistently a single, discrete fragment of DNA from the different desiccated samples in PCR assays (Fig. 1). As expected, the two different combinations of primers generated fragments that differed in size by approximately 10 bp. Use of the primer combination that included LEUI3 (corresponding to a more conserved region of the *Anabaena* sp. strain PCC 7120 tRNA^{Leu} (UAA) intron; see Methods) resulted in more efficient amplification of the *Nostoc* intron sequences (Fig. 1). Preliminary sequence analysis identified the products as amplified group I intron sequences with portions of the flanking tRNA^{Leu} (UAA) exons. The position of insertion of the intron in the different tRNA^{Leu} genes was, in each case, between U and A within the anticodon loop [conserved 5'-GGACUU↓AA-3'; the same position identified for other cyanobacterial group I introns in previous studies (Paquin *et al.*, 1997)].

Three of the 41 desiccated samples, WH005, WH010 and WH011, failed to generate an amplification product in PCR assays in multiple trials, despite exhaustive purification of their DNAs, including treatments with PTB, which had proved effective with other recalcitrant samples. The same genomic DNAs from WH005, WH010 and WH011 did, however, generate products of the expected size when the PCR assay was performed with primers designed to amplify portions

Fig. 1. Sets of oligonucleotide primers based upon regions of tRNA^{Leu} (UAA) from *Anabaena* sp. PCC 7120 amplify a single product from *N. commune* WH014 desiccated for 50 years. DNA was purified using PTB reagent. Agarose gel stained with ethidium bromide. Lanes: 1, primers LEU1 and LEU2; 2, primers LEU1 and LEU3; M, 1 kbp ladder markers (LTI Inc.).

of two other genes (*phr* and 23S rDNA) identified in *N. commune* DRH1 (data not shown). Of the 38 desiccated samples that provided amplification products, a total of 45 independent intron sequences were obtained for analysis; i.e. some samples generated more than one size of fragment in repeat trials (see below). Introns were also amplified from liquid cultures of *N. commune* DRH1, *N. commune* UTEX 584, *N. punctiforme* ATCC 29133 and *Anabaena* sp. strain PCC 7120. Published sequences from strains identified as *Anabaena* or *Nostoc* were included in the analysis to provide a total dataset of 71 sequences.

Structure of conserved and variable regions in tRNA^{Leu} (UAA) group I introns

Intron sequences were ordered in several provisional subgroupings according to intron size, DNA sequence similarities and the presence or absence of hyper-variable regions (Table 2). The majority (34) of the amplified *Nostoc* tRNA^{Leu} (UAA) group I intron sequences were of identical sizes (284 bp) and shared very high sequence similarity or were identical. The sequence similarity between one given 284 bp sequence and any one of the 33 other 284 bp sequences varied from 95.8 to 100% and it was therefore possible to align these sequences without ambiguity. Sixty-seven of the introns had a conserved sequence of 200 bp.

Fig. 2. Structure of group I introns. (a) Secondary structure predictions of variable region 1, hypervariable region (h) and variable region II. Arrows indicate the 5' and 3' ends of the hypervariable region. A conserved G residue is indicated by asterisks. (b) Consensus sequence of the variable regions. Arrows under bases indicate sequence repeats or similarities.

In ALD776DA/1973, ALD776DB/1973, ALD8122/1973 and LBP/1995, the conserved region was 201 bp. In addition, ALD776DA/1973, ALD776DB/1973 and ALD8122/1973 each contained a 34 bp extension between the first conserved region and variable region I.

All introns conformed to a primary structure with three conserved regions and two variable regions. The latter were subdivided into variable region I, a hypervariable region and variable region II, and variable region III (Fig. 2b). Variable region III and parts of I and II for many samples were readily aligned and

secondary structure predictions suggest that many of the bases in variable region I, the hypervariable region and variable region II may be base-paired (Fig. 2a).

Although the hypervariable regions were not readily aligned, eight groups were identified on the basis of structural features. Short stretches of sequence were common within the members of the different groups. For example, the first eight (5'-TCAAAATC-3') and last six (5'-TAAAAT-3') bases of the hypervariable region in TOP/1993 correspond to the terminal regions in *Nostoc* AF019920, AF019912, AJ228712 and AF055660. The complete 38-nt hypervariable region

Fig. 3. Distance tree of conserved intron sequences. The tree (unrooted) was inferred using distance matrices obtained with Kimura's two-parameter model (Kimura, 1980), calculated with the DNADIST and NEIGHBOR programs of PHYLIP (Felsenstein, 1985). Bootstrap values were deduced from 100 replicates using BOOTSEQ (see Methods). The distance between two strains, expressed in substitutions per nucleotide, is obtained by adding the lengths of the horizontal branches connecting them. Intron designations consist of the sample number, a letter (if necessary) to indicate an independent amplified product/year the material was desiccated. Accession numbers refer to GenBank. *Nostoc* PCC 73102 = *N. punctiforme* ATCC 29133. Boxes and labels (I–IV) demarcate major clusters.

in *Nostoc* intron AF055660 shows 100% sequence identity to the collective 5' and 3' parts of TOP/1993 (and WH9/ND). Hypervariable regions in introns of strains that were collected as free-living colonies, e.g. *Nostoc* MAL/1989, *Nostoc* NZE/1997 and *Nostoc* PARA/1979, shared very high sequence similarity with those identified in introns of *Nostoc* cyanobionts (Paulsrud & Lindblad, 1998).

Multiple introns from single samples

Samples ALD776D, ALD857D, MAL/1989 and RIS/1979 provided multiple intron sequences. Introns ALD776DA/1973 and ALD776DB/1973 differ at two nucleotide positions (in conserved regions) and form part of group I (Fig. 3). They differ from ALD776DC/1973 (group IV) through indels of 2, 3, 4, 5, 6 and 11 nt within the variable regions. Similarly, ALD857DA/1973 and ALD857DB/1973 differ from one another at three nucleotide positions (in conserved regions) and from ALD857DC/1973 through a 14-nt indel; these three introns cluster in group IV. MALA/1989 and MALB/1989 (group III) differ at a single nucleotide and from MALC/1989 (group IV) through multiple indels within variable regions. RISA/1979 and RISB/1979 (group IV) differ at six nucleotide positions across conserved and variable regions.

Potential secondary splicing reaction

The 34 284-bp introns lack a hypervariable region and thus variable regions I and II are contiguous. The sequence of bases from position 47 to position 62 in these contiguous regions is 5'-TCCAAAATCCAA-AATT-3' (Fig. 2b); the central part of this sequence (5'-AAAATCC-3') corresponds exactly to the conserved sequence at the 3' intron-exon splice site within tRNA^{Leu} (UAA) (Fig. 2b). The hypervariable region interrupts variable region I at this position (Fig. 2b). In some introns with a hypervariable region, 5'-AAAATCC-3' appears within the hypervariable sequence itself (e.g. *Anabaena azollae*); in others, 5'-AAAAT-3' at the 3' end of the hypervariable region is immediately contiguous with 5'-CC-3' at the 5' end of variable region II (Fig. 2b; e.g. TOP/1993). In view of these data, including the secondary structure predictions of hypervariable regions, we questioned whether the hypervariable region underwent a secondary splicing reaction. We selected cloned intron TOP/1993, which has 5'-AAA AUCC-3' at the 3' boundary of the hypervariable region, and *N. commune* UTEX 584, which lacks this sequence in its hypervariable region. In multiple trials, no evidence was obtained for guanosine-catalysed splicing of the hypervariable region under experimental conditions that did lead to splicing of the complete introns from these two samples *in vitro*, as detected through Northern analysis (data not shown).

Phylogeny of *Nostoc* and *Anabaena*

Distance trees of the conserved intron sequences were constructed using a neighbour-joining algorithm. The robustness of each tree was tested using bootstrap analysis. Maximum-parsimony analysis was also used to build consensus trees.

All of the conserved sequences that corresponded with the provisional grouping of 284-bp introns (Table 2), with the single exception of WH2/1939, formed a single cluster, characterized by little or no divergence (Fig. 3; cluster IV). This cluster included two samples (WH008, WH015) identified originally as *N. commune* var. *flagelliforme*, one sample (SPH/1998) identified as *Nostoc sphaericum* and one sample (WH016) identified as *Nostoc pellucidum*. The relative ordering of sequences in cluster IV varied upon resampling because many sequences share 100% identity and cannot be discriminated from one another. *Nostoc* samples TOP/1993, WH009 and NZE/1997 have the classic colony morphology of *N. commune* and were otherwise indistinguishable from the samples represented by cluster IV. However, they were excluded from cluster IV on the basis of maximum-parsimony analysis (data not shown) and appear in cluster II with sequences from forms that enter into symbiotic associations. Four samples (WH015, WH008, PARA/1979 and ALD857D) were identified as *N. commune* var. *flagelliforme*. The morphology of this form consists of bundles of hair-like colonies. Three of the samples were included in cluster IV while *Nostoc* PARA/1979 (from Australia) grouped in cluster III with an independent isolate from China, AJ228710, which was also named as *N. commune* var. *flagelliforme* (Rudi *et al.*, 1998).

DISCUSSION

The growing realization of the potential breadth and extent of microbial diversity has prompted a renewed interest in the classification of prokaryotes and bacterial taxonomy. In studies that question phylogenies and evolutionary relationships, the concept of a bacterial 'species' is a long-standing issue that remains a topic of controversy and debate (Pace, 1997; Stencil, 2000). A satisfactory resolution of this issue, and a clearer understanding of the genetic basis for phenotypic plasticity, will depend on the identification of defined microbial populations and their interrogation with reliable genetic markers. Cyanobacterial populations have a number of advantages for such analyses: they have a cosmopolitan distribution; they dominate the microbial populations of many extreme environments; their colonies are often conspicuous and may assume macrophytic proportions; colonies of what appear to be morphologically identical forms occur in geographically isolated environments; growths, often attributed to a 'monospecific' population, may cover many square kilometres; and cyanobacteria have a

fossil record. Of those genetic markers that may resolve clusters and groupings within taxa below the level of genus, group I intron sequences have been investigated in a broad range of eubacterial taxa (Biniszkiwicz *et al.*, 1994; Paquin *et al.*, 1997, 1999; Rudi & Jakobsen, 1997; Edgell *et al.*, 2000). In this study, we focussed on a large set of desiccated samples of free-living *Nostoc* for which detailed records were available of time and place of origin and taxonomic assignment. The major objective was to test the form species concept as it pertains to cyanobacteria, to assess the utility of group I intron analysis as a determinative tool, to resolve possible relationships within the *Anabaena*–*Nostoc* lineage and to identify possible correlations between time of collection, geographical distribution, endemic isolation and intron sequence divergence. The possibility of secondary splicing events in some of the introns was also questioned.

Form species *Nostoc commune*

Texts such as Geitler (1932) emphasize the use of morphological features as determinative criteria for cyanobacterial taxa. Colonies of *N. commune* are cosmopolitan and occur as free-living, blackened, friable crusts when dry and as green, lobate masses with a cartilaginous to gelatinous texture when rehydrated (see Fig. 1b in Potts, 2000). The colony morphology is sufficiently characteristic that phylogenists and microbiologists felt able to assign the name *N. commune* with confidence, the taxonomic problems of the Cyanobacteria notwithstanding (see Whitton & Potts, 2000). Thirty of the samples used in this study were collected in 13 different countries, in both hemispheres, and were identified as *N. commune* by independent investigators over a period of some 200 years. Twenty-five of the forms assigned this species name each contained a 284-bp tRNA^{Leu} (UAA) group I intron that lacked a hypervariable region; collectively, these sequences were characterized by little or no divergence and formed cluster IV of the distance tree (Fig. 3). The remaining five samples had indistinguishable colony morphology but were excluded from cluster IV in phylogenetic analysis. On the basis of these data, the phylogenetic grouping represented by cluster IV was attributed provisionally to form species *N. commune*.

Reliability of morphological criteria

The samples studied included those with designations of *Nostoc* sp. (four samples), *N. sphaericum* (one sample), *N. pellucidum* (one sample) and *N. commune* var. *flagelliforme* (four samples). In the cases of *N. sphaericum* and *N. commune* var. *flagelliforme*, the name designations were based, in large part, upon colony morphology. Colonies of the former cyanobacterium exist as gelatinous spheres when growing in standing water (Potts, 2000), while the desiccated thread-like and twine-like bundles of the latter are equally distinctive (the colloquial term for this cyano-

bacterium in China translates as 'hair-vegetable'; Potts, 1997). *N. sphaericum* (SPH/1998) and *N. commune* var. *flagelliforme* (WH015) were attributed to form species *N. commune* through intron analysis. Several of the samples of *N. commune* we have isolated in clonal culture can produce spheres or 'pearls' according to the growth conditions (Hill *et al.*, 1994a; unpublished data), which suggests that the use of this character as a taxonomic marker is unreliable. Whether the presence of a characteristic hair-like morphology justifies assignment to *N. commune* var. *flagelliforme* is harder to resolve. Three of the four samples grouped within cluster IV. The intron of the most divergent of the four samples (PARA/1979; collected in 1979 from Australia) was 94.5% identical to that of an independent isolate from China. Lateral intron transfer may be responsible for the assignment of WH015 and WH008 to cluster IV, but it is also possible that a characteristic growth habit (bundles of hair-like filaments) can be achieved by genotypically different forms of *Nostoc*. The latter possibility was discussed with regard to spherical colony morphology (above) and it can be emphasized further by noting that TOP/1993, with the classic *N. commune* lobate colony structure and cell morphology, is divergent from cluster IV on the basis of intron phylogeny.

A strain from China, identified as *N. commune* (in Rudi *et al.*, 1998), contained an intron that differed at only a single nucleotide from the intron sequence of *N. commune* UTEX 584 determined in this study. *N. commune* UTEX 584 is of unknown origin and the naming of this strain is considered erroneous (Potts, 2000). Unlike form species *N. commune*, [*N. commune*] strain UTEX 584 has a conspicuous life cycle and a pronounced production of the secondary metabolite geosmin, it lacks water stress protein *wsp* and contains at least two genes (*iphP*, *glbN*) that are absent from *N. commune* DRH1 (Potts, 2000). The *nif* genes of [*N. commune*] UTEX 584 and *N. punctiforme* ATCC 29133 (cluster II, IIa/IIb) are also more similar to one another than to those of *Anabaena* sp. strain PCC 7120 or any other cyanobacterium (T. Thiel, personal communication). It appears that strains from the University of Texas Culture Collection (UTEX) were donated to the Hydrobiological Institute of Wuhan, China, from where the strains (in Rudi *et al.*, 1998) were acquired (Z. Huang, personal communication). This example emphasizes the extreme caution that must be taken when interpreting phylogenetic relationships based upon names assigned to culture collection strains (Whitton & Potts, 2000).

A clear understanding of the diagnostic traits that can be used to discriminate between a strain of *Anabaena* and a strain of *Nostoc* is currently lacking. Based upon phylogenetic analysis of conserved regions, and irrespective of the sequence designated as the outgroup, the group I intron of *Anabaena* sp. PCC 7120 was the most divergent of the 71 sequences used in this study. In the taxonomic scheme of Stanier (see Rippka, 1988; Rippka *et al.*, 1979), the production of hormogonia

(motile filaments), if produced, is considered a good discriminatory character. The maturation of a hormogonium into a mature filament has been described in detail for a number of *Nostoc* species with complex life cycles (see Potts, 2000). Largely on the basis of DNA–DNA hybridization data, *Anabaena* sp. PCC 7120 is considered a nomenspecies of *Nostoc* sp. PCC 6705 and 6719, even though it produces no hormogonia. Hormogonium formation, as a negative character, is unreliable (Rippka, 1988). For example, strains derived from *N. commune* TEN/1988, BBC/1990 and TAG/1988 failed to generate hormogonia in liquid culture, but the parent materials did so when incubated on solid media. The complete genome sequence data of *Anabaena* sp. PCC 7120 and *N. punctiforme* ATCC 29133 provide a means to resolve this taxonomic problem.

Absence of group I introns

Three of the oldest samples, WH005 (1859), WH010 (1880) and WH011 (1860), failed to generate any amplification products in PCR assay. The samples were received with the names *N. commune*, *Nostoc alpicola* Kützinger and *N. commune* var. *lesleri* Bornet, respectively. It is unlikely that the tRNA sequences of these strains (i.e. the primer hybridization sites) were sufficiently different from those of other *Nostoc* strains to have prevented amplification using the primers based upon *Anabaena* sp. strain PCC 7120 tRNA^{Leu} (UAA). The latter is the most divergent strain in the study set, yet the primers allowed amplification from a diverse collection of samples. Even if introns were absent from these materials, PCR assay should have amplified part of the tRNA^{Leu} (UAA) genes. Given the age of the samples, the quality of the DNA may have prevented amplification, although introns were recovered from even older samples where PTB reagent proved effective. The small amounts of material available prevented Southern analysis, which might have resolved the issue.

Geographical correlations

The samples used in this study are not old enough (on a geological time scale) to permit a rigorous appraisal of intron evolution in dispersed populations of the same form species. However, the samples are old from the perspective of free-living populations of bacteria; many of the samples pre-date the Industrial Revolution and the onset of the widespread use of herbicides, pesticides and other xenobiotics, which have poorly understood effects on gene conversion and migration in terrestrial (and aquatic) ecosystems. The older samples also pre-date the widespread use of air travel and the increased frequency of travel to both populated and remote regions. Availability of inocula, transport and maintenance in aerosols are important, often overlooked, considerations in studies of microbial populations, especially those that tolerate

extreme desiccation. Some cyanobacteria, and also eukaryotic algae, are sufficiently desiccation-tolerant to survive long-distance transport in aerosols, even over Antarctica (Marshall & Chalmers, 1997). Viable cells of the non-heterocystous cyanobacterium *Microcoleus* were sampled from aerosols following nuclear bomb detonations (Kraus, 1969). These are important considerations, because there does seem to be some restriction on the global occurrence of cyanobacterial populations that may be circumvented by import of samples. Some species of *Synechococcus* are clearly restricted in geographical distribution. For example, all forms of thermophilic *Synechococcus* are absent from Icelandic hot springs even though numerous springs exist there that appear chemically suitable (Ward & Castenholz, 2000).

The samples used for analysis included the spectrum of longitude and latitude. If there is any correlation between sequence variation within the different phylogenetic trees and geographical distributions of *Nostoc*, it is not an apparent one. For example, introns ENG/1996 and BBC/1990, from independent populations growing approximately 4 km from one another in Virginia, share 100% identity and 98.9% identity with TEN/1988, from a population sampled from the roof of a schoolhouse several hundred kilometres to the south, in Tennessee. However, intron WH12/1880, from *N. commune* collected from Surabaya, Java (118 years previously and more than 15000 km distant), is 100% identical to ENG/1996 and BBC/1990. The significance of very small differences in sequence similarity is not clear, although Paulsrud *et al.* (2000) were able to use small differences in the variable I regions of tRNA^{Leu} (UAA) introns to compare geographical distribution patterns of *Nostoc* cyanobionts.

TOP/1993 was excluded from form species *N. commune* in all phylogenies that were constructed. Despite the characteristic colony morphology, these populations grow on coastal sand subject to sea spray, a feature that is atypical, considering the alkaline limestone environments typically colonized by *N. commune* (Potts, 2000).

Multiple introns

In samples that contained multiple introns, the different forms were either very similar, with only a few nucleotide substitutions, e.g. MALA/1989 and MALB/1989, or they differed more significantly and, in addition, had one or more indels; e.g. compare ALD776DA/1973 and ALD776DB/1973 with ALD776DC/1973. In the former situation, the introns clustered in the same phylogenetic grouping, and these data may reflect cases of intron evolution (in colonies that are stable over time). For the latter situation, the differences were sufficient to discriminate the introns into different phylogenetic groupings, which may suggest mixed populations (see Rudi & Jakobsen, 1999).

Introns in remote populations

The question of intron stability in remote populations is an important one and it has some bearing on the controversy over ancient versus recent origins of group I introns. Aldabra (9° S) is an uninhabited, remote coral atoll situated in the extreme southwest corner of the Indian Ocean (see Potts, 2000 for references). Visually conspicuous colonies of *Nostoc* are scattered over the limestone surface and shallow soils of the atoll and become locally abundant in small depressions and solution-holes, which fill partially with water during the wet season (WNW monsoon, lasting from December to April). Samples ALD776D and ALD8122 were collected independently, respectively in 1973 and 1974, from pool W10 on West Island. At the time of collection, these samples were identified as *Nostoc* sp. (*sensu* Geitler, 1932). Sample ALD779D was collected from Bassin Cabri, a pool several hundred metres from W10, and was identified as *N. commune* upon collection. Sample ALD857D was collected from 'platin' (flat limestone pavement) in the vicinity of W10 and Bassin Cabri and was termed 'crumbly *Nostoc*' and identified as *N. commune* ± var. *flagelliforme* at the time of collection. The populations of these three environments are distinguished on the basis of tRNA^{Leu} (UAA) group I intron analysis. The group I intron sequences of ALD776D and ALD8122 are identical (with a 33-nt hypervariable region), which may suggest that the dominant population in pool W10 was 'stable' and monospecific, at least over the course of two wet seasons and one dry season. In contrast, the intron of ALD779D has the signature of the 284-nt group I intron (no hypervariable region) of form species *N. commune*. The hypervariable regions of introns from ALD857D and ALD8122 and ALD776D are very different. Given that there is no shortage of inoculum in any part of the atoll, the persistence of discrete populations, in very close proximity to one another, is significant.

Age correlations

N. commune strain DRH1 was brought into liquid culture in 1993 using CHEN/1986 as the primary source of inoculum (Hill *et al.*, 1994a). The intron sequences of these materials, determined in 1998, share 97.9% identity. Do the five nucleotide changes (one in the conserved region and four in the variable regions) reflect mutations derived from several years of liquid culture or does CHEN/1986 contain more than one form (and intron)? Variations due to PCR amplification are discounted, on the basis of the precautions taken to sample multiple times from each material for PCR assay and to sequence the cloned introns on both strands. For the same reason, the possibility of multiple forms of intron is unlikely; for other samples where more than one population of introns was in fact present, these were easily differentiated during PCR assay and subcloning. It is assumed, therefore, that the nucleotide substitutions reflect genetic drift in the

laboratory culture. It can be noted that this example contrasts with the data for strains *N. punctiforme* ATCC 29133 and *Nostoc* sp. PCC 73102. These strains have identical intron sequences, share a common origin and have been maintained for many years in separate culture collections.

Hypervariable regions and intron evolution

There is controversy as to whether tRNA^{Leu} (UAA) introns have an ancient or recent origin and whether they are stable or mobile (Besendahl *et al.*, 2000; Edgell *et al.*, 2000; Paquin *et al.*, 1997, 1999; Rudi & Jakobsen 1997, 1999). In fact, current opinion is that there is no single reason that can adequately explain the current distribution of group I and group II introns in bacteria (Edgell *et al.*, 2000). Recent work also suggests that the evolution of the intron in the tRNA^{Leu} (UAA) gene is considerably more complex than previously thought (Rudi & Jakobsen, 1999). The identification in this study of short sequences in many of the introns, often in tandem repeats and identical or very similar to the sequence of the conserved intron-exon boundary, raises questions about the origin of hypervariable regions. No evidence was obtained in this study for *in vitro* secondary splicing of these regions. It is interesting that all of the forms assigned to form species *N. commune*, representing a wide range of geographically isolated communities, have a group I intron of 284 nt that has no hypervariable region (Fig. 2a). The 284-nt group I intron of cluster IV appears to have withstood acquisition of a hypervariable region, be it from invasion by mobile elements and/or duplication events.

Our data confirm and emphasize the complexity of intron evolution and suggest that an intron phylogeny alone cannot be used with confidence to infer cyanobacterial species or strain relationships. The reasons for this include, but are not limited to, the morphological similarity of genetically different forms, morphological dissimilarities between forms that share identical genetic markers, misnaming of culture collection strains and possible genetic drift in strains maintained in continuous culture. Examples of all of these drawbacks were encountered during the study and they may pose greater obstacles to the successful definition of other cyanobacterial species that lack characteristic morphological criteria. Note, however, that most of the samples used in this study were never brought into culture and were assigned to *N. commune* on the basis of colony morphology *in situ* alone; significantly, the majority of these showed little to no divergence for independent genetic markers.

Conclusions

Populations that constitute global form species *N. commune* were identified through group I intron analysis. In many cases, the phylogenetic analysis supported identifications based solely upon morpho-

logical criteria. In contrast, for several samples, the analysis brought into question the reliability of certain morphological criteria. Information on the degree of genetic (genomic) diversity within the form species, and within the *Nostoc*–*Anabaena* lineage in general, is only fragmentary at this time. The phylogenetic tree based upon intron sequence analysis (Fig. 3) gives the impression of a continuum, with ill-defined borders, punctuated by clusters and groups.

ACKNOWLEDGEMENTS

We thank anonymous reviewers who pointed out important errors in the manuscript. The work was supported by NSF grant IBN9513157 (M.P.) and DARPA grant N00173-98-1-G005 (M.P. and R.F.H.). We thank all colleagues who provided samples for this study, especially Walter Till of the Wien Herbarium, Austria, and Stephen Skinner.

REFERENCES

- Besendahl, A., Qiu, Y. L., Lee, J. H., Palmer, J. D. & Bhattacharya, D. (2000). The cyanobacterial origin and vertical transmission of the plastid tRNA^{Leu} group-I intron. *Curr Gen* **37**, 12–23.
- Biniszkiewicz, D., Cesnaviciene, E. & Shub, D. A. (1994). Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. *EMBO J* **13**, 4629–4635.
- Castenholz, R. W. (1992). Species usage, concept, and evolution in the cyanobacteria (blue-green algae). *J Phycol* **28**, 737–745.
- Chech, T. R., Damberger, S. H. & Gutell, R. R. (1994). Representation of the secondary and tertiary structure of group I introns. *Nat Struct Biol* **1**, 273–280.
- Costa, J.-L., Paulsrud, P. & Lindblad, P. (1999). Cyanobiont diversity within coralloid roots of selected cycad species. *FEMS Microbiol Ecol* **28**, 85–91.
- Edgell, D. R., Belfort, M. & Shub, D. A. (2000). Barriers to intron promiscuity in bacteria. *J Bacteriol* **182**, 5281–5289.
- Everett, K. D. E., Kahane, S., Bush, R. M. & Friedman, M. G. (1999). An unspliced group I intron in 23S rRNA links *Chlamydiales*, chloroplasts, and mitochondria. *J Bacteriol* **181**, 4734–4740.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
- García-Pichel, F., Prufert-Bebout, L. & Muyzer, G. (1996). Phenotypic and phylogenetic analyses show *Microcoleus chthonoplastes* to be a cosmopolitan cyanobacterium. *Appl Environ Microbiol* **62**, 3284–3291.
- Geitler, L. (1932). Cyanophyceae. In *Rabenhorst's Kryptogamen-Flora*, vol. 14, pp. 1–1196. Leipzig: Akademische Verlagsgesellschaft.
- Gilbert, W. & Glynias, M. (1993). On the ancient nature of introns. *Gene* **135**, 137–144.
- Helm, R. F., Huang, Z., Edwards, D., Leeson, H., Peery, W. & Potts, M. (2000). Structural characterization of the released polysaccharide of desiccation-tolerant *Nostoc commune* DRH-1. *J Bacteriol* **182**, 974–982.
- Hill, D. R., Peat, A. & Potts, M. (1994a). Biochemistry and structure of the glycan secreted by desiccation-tolerant *Nostoc commune* (Cyanobacteria). *Protoplasma* **182**, 126–148.
- Hill, D. R. H., Hladun, S. L., Scherer, S. & Potts, M. (1994b). Water stress proteins of *Nostoc commune* (Cyanobacteria) are secreted with UV-A/B-absorbing pigments and associate with 1,4- β -D-xylanxylanohydrolase activity. *J Biol Chem* **269**, 7726–7734.
- Holt, J. G., Krieg, N. R., Sneath, P. H. A., Staley, J. T. & Williams, S. T. (editors) (1994). *Bergey's Manual of Determinative Bacteriology*, 9th edn. Baltimore: Williams & Wilkins.
- Iteman, I., Rippka, R., de Marsac, N. T. & Herdman, M. (2000). Comparison of conserved structural and regulatory domains within divergent 16S rRNA–23S rRNA spacer sequences of cyanobacteria. *Microbiology* **146**, 1275–1286.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120.
- Komárek, J. & Anagnostidis, K. (1989). Modern approach to the classification system of cyanophytes. 4 – Nostocales. *Arch Hydrobiol Suppl* **823** (*Algol Stud* **56**), 247–345.
- Kraus, M. P. (1969). Resistance of blue-green algae to ⁶⁰Co gamma radiation. *Radiat Bot* **9**, 481–489.
- Kuhse, M. G., Strickland, R. & Palmer, J. D. (1990). An ancient group I intron shared by eubacteria and chloroplasts. *Science* **250**, 1570–1573.
- Marshall, W. A. & Chalmers, M. O. (1997). Airborne dispersal of Antarctic terrestrial algae and cyanobacteria. *Ecography* **20**, 585–594.
- Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740.
- Paquin, B., Kathe, S. D., Nierzwicki-Bauer, S. A. & Shub, D. A. (1997). Origin and evolution of group I introns in cyanobacterial tRNA genes. *J Bacteriol* **179**, 6798–6806.
- Paquin, B., Heinfling, A. & Shub, D. A. (1999). Sporadic distribution of tRNA^{Arg}_{CCU} introns among α -purple bacteria: evidence for horizontal transmission and transposition of a group I intron. *J Bacteriol* **181**, 1049–1053.
- Paulsrud, P. & Lindblad, P. (1998). Sequence variation of the tRNA^{Leu} intron as a marker for genetic diversity and specificity of symbiotic cyanobacteria in some lichens. *Appl Environ Microbiol* **64**, 310–315.
- Paulsrud, P., Rikkinen, J. & Lindblad, P. (2000). Spatial patterns of photobiont diversity in some *Nostoc*-containing lichens. *New Phytol* **146**, 291–299.
- Poinar, H. N., Hofreiter, M., Spaulding, W. G., Martin, P. S., Stankiewicz, B. A., Bland, H., Evershed, R. P., Possnert, G. & Pääbo, S. (1998). Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* **281**, 402–406.
- Potts, M. (1997). Etymology of the genus name *Nostoc* (Cyanobacteria). *Int J Syst Bacteriol* **47**, 584.
- Potts, M. (2000). *Nostoc*. In *The Ecology of Cyanobacteria*, pp. 465–504. Edited by B. A. Whitton & M. Potts. Dordrecht: Kluwer Academic.
- Rasmussen, U. & Svenning, M. M. (1998). Fingerprinting of cyanobacteria based on PCR with primers derived from short and long tandemly repeated repetitive sequences. *Appl Environ Microbiol* **64**, 265–272.
- Rippka, R. (1988). Recognition and identification of cyanobacteria. *Methods Enzymol* **167**, 28–67.
- Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M. & Stanier, R. Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* **111**, 1–61.
- Rudi, K. & Jakobsen, K. S. (1997). Cyanobacterial tRNA^{Leu} (UAA) group I introns have polyphyletic origin. *FEMS Microbiol Lett* **156**, 293–298.

- Rudi, K. & Jakobsen, K. S. (1999). Complex evolutionary patterns of tRNA^{Leu} UAA group I introns in cyanobacterial radiation. *J Bacteriol* **181**, 3445–3451.
- Rudi, K., Skulberg, O. M., Larsen, F. & Jakobsen, K. S. (1997). Strain characterization and classification of oxyphotobacteria in clone cultures on the basis of 16S rRNA sequences from the variable regions V6, V7, and V8. *Appl Environ Microbiol* **63**, 2593–2599.
- Rudi, K., Skulberg, O. M. & Jakobsen, K. S. (1998). Evolution of cyanobacteria by exchange of genetic material among phylogenetically related strains. *J Bacteriol* **180**, 3453–3461.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
- Scheldeman, P., Baurain, D., Bouhy, R., Scott, M., Muhling, M., Whitton, B. A., Belay, A. & Wilmotte, A. (1999). *Arthrospira* ('Spirulina') strains from four continents are resolved into only two clusters, based on amplified ribosomal DNA restriction analysis of the internally transcribed spacer. *FEMS Microbiol Lett* **172**, 213–222.
- Schopf, J. W. (2000). The fossil record: tracing the roots of the cyanobacterial lineage. In *The Ecology of Cyanobacteria*, pp. 13–35. Edited by B. A. Whitton & M. Potts. Dordrecht: Kluwer Academic.
- Shirkey, B., Kovarcik, D. P., Wright, D. J., Wilmoth, G., Prickett, T. F., Helm, R. F., Gregory, E. M. & Potts, M. (2000). Active Fe-containing superoxide dismutase and abundant *sodF* mRNA in *Nostoc commune* (Cyanobacteria) after years of desiccation. *J Bacteriol* **182**, 189–197.
- Stencil, C. (2000). Microbial diversity; eyeing the big picture. *ASM News* **66**, 142–146.
- Vasan, S., Zhang, X. & Kapurniotu, A. & 10 other authors (1996). An agent cleaving glucose-derived protein crosslinks *in vitro* and *in vivo*. *Nature* **382**, 275–278.
- Ward, D. M. & Castenholz, R. W. (2000). Cyanobacteria in geothermal habitats. In *The Ecology of Cyanobacteria*, pp. 37–60. Edited by B. A. Whitton & M. Potts. Dordrecht: Kluwer Academic.
- Whitton, B. A. & Potts, M. (2000). Introduction to the cyanobacteria. In *The Ecology of Cyanobacteria*, pp. 1–11. Edited by B. A. Whitton & M. Potts. Dordrecht: Kluwer Academic.
- Wilmotte, A. (1994). Molecular evolution and taxonomy of the cyanobacteria. In *The Molecular Biology of Cyanobacteria*, pp. 1–25. Edited by D. A. Bryant. Dordrecht: Kluwer Academic.
- Wilmotte, A., Stam, W. & Demoulin, V. (1997). Taxonomic study of marine oscillatoriacean strains (Cyanophyceae, Cyanobacteria) with narrow trichomes. III. DNA–DNA hybridization studies and taxonomic conclusions. *Algol Stud* **87**, 11–28.
- Xie, W.-Q., Tice, D. & Potts, M. (1995). Cell water deficit regulates expression of *rpoC1C2* (RNA polymerase) at the level of mRNA in desiccation-tolerant *Nostoc commune* UTEX 584 (Cyanobacteria). *FEMS Microbiol Lett* **126**, 159–164.
- Xu, M.-Q., Kathe, S. D., Goodrich-Blair, H., Nierzwicki-Bauer, S. A. & Shaub, D. A. (1990). Bacterial origin of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria. *Science* **250**, 1566–1570.
- Zaug, A. J., McEvoy, M. M. & Cech, T. R. (1993). Self-splicing of the group I intron from *Anabaena* pre-tRNA: requirement for base-pairing of the exons in the anticodon stem. *Biochemistry* **32**, 7946–7953.
- Zuker, M., Mathews, D. H. & Turner, D. H. (1999). Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In *RNA Biochemistry and Biotechnology*, pp. 11–43. NATO ASI Series. Edited by J. Barciszewski & B. F. C. Clark. Dordrecht: Kluwer Academic.