Molecular identification, typing and traceability of cyanobacteria from freshwater reservoirs

Elisabete Valério,1,2 Lélia Chambel,1 Sérgio Paulino,2 Natália Faria,2 Paulo Pereira2 and Rogério Tenreiro1

1Universidade de Lisboa, Faculdade de Ciências, Centro de Biodiversidade, Genómica Integrativa e Funcional (BioFIG), Edifício ICAT, Campus da F CUL, Campo Grande, 1749-016 Lisboa, Portugal
2Laboratório de Microbiologia e Ecotoxicologia, Instituto Nacional de Saúde Dr Ricardo Jorge, Avenida Padre Cruz, 1649-016 Lisboa, Portugal

In order to assess the potential of several molecular targets for the identification, typing and traceability of cyanobacteria in freshwater reservoirs, molecular techniques were applied to 118 cyanobacterial isolates mostly sourced from Portuguese freshwater reservoirs and representative of three orders of cyanobacteria: Chroococcales (54), Oscillatoriales (15) and Nostocales (49). The isolates were previously identified by morphological methods and subsequently characterized by composite hierarchical cluster analysis of STRR and LTRR (short and long tandemly repeated repetitive sequences) PCR fingerprinting profiles. Representative isolates were selected from each cluster and their molecular identification, at the species level, was obtained or confirmed by phylogenetic positioning using 16S rRNA gene and rpoC1 phylogenies. A highly congruent association was observed between STTR- and LTRR-based clusters and taxonomic affiliation, revealing the usefulness of such PCR fingerprinting profiles for the identification of cyanobacteria. Composite analysis of hierarchical clustering of M13 and ERIC PCR fingerprints also appeared suitable for strain typing and traceability within a reservoir, indicating its potential for use in cyanobacterial monitoring, as a quality management control. Based on Simpson (D) and Shannon–Wiener (J) indices a high diversity was observed within all species, with Planktothrix agardhii showing the lowest diversity values (D=0.83; J=0.88) and Aphanizomenon flos-aquae the highest ones (D=J=0.99). A diagnostic key based on 16S-ARDRA, ITS amplification and ITS-ARDRA for identification purposes is also presented.

INTRODUCTION

Cyanobacteria are a morphologically diverse group of bacteria ranging from unicellular to colonial and filamentous forms. Taxonomically, cyanobacteria are grouped into unicellular forms that divide by binary fission (Order Chroococcales, or Bergey’s Subsection I) or multiple fission (Order Pleurocapsales, or Bergey’s Subsection II); and filamentous forms that are non-heterocystous (Order Oscillatoriales, or Bergey’s Subsection III) or differentiate heterocysts in non-branching (Order Nostocales, or Bergey’s Subsection IV) or branching filaments (Order Stigonematales, or Bergey’s Subsection V) (Castenholz, 2001). A sixth cyanobacterial order, Gloeobacterales, was proposed by Cavalier-Smith (2002) to accommodate the genus Gloeobacter, formerly included in the Chroococcales.

Traditionally, the classification of cyanobacteria has been based on morphological characters such as trichome width, cell size, division planes, shape and arrangement, pigmentation and the presence of characters such as gas vacuoles and a sheath (Baker, 1991, 1992; Bourrelly, 1970; Komárek & Anagnostidis, 1986, 1989). Beyond the considerable expertise required to identify species by such characters, subjective judgment by operators can lead to errors, resulting in incorrect assignment of isolates. Komárek & Anagnostidis (1989) estimated that more than 50 % of the strains in culture collections are misidentified. Moreover, some diagnostic features, such as gas vacuoles or akinetes, can show variations with different environmental or growth conditions and even be lost during cultivation (Rudi et al., 1997; Lyra et al., 2001). Such limitations of phenotypic characters have highlighted the requirement for...
more reliable methods and promoted molecular approaches in cyanobacterial taxonomy, including DNA base composition (Kaneko et al., 1996, 2001), DNA hybridizations (Kondo et al., 2000), gene sequencing (Nübel et al., 1997) and PCR fingerprinting (Rasmussen & Svenning, 1998; Versalovic et al., 1991). Cyanobacterial-specific methods not requiring axenic cultures are of utmost importance since such cultures are difficult to obtain (Choi et al., 2008).

Replicative sequences constitute an important part of the prokaryotic genome (van Belkum et al., 1998). Despite their unknown function, and lack of knowledge about how they are maintained and dispersed, the presence, widespread distribution and high conservation of these sequences make them methodologically important for DNA fingerprinting and allow their use as an alternative for the identification of species or strains and in diversity studies. In the particular case of cyanobacteria, a family of repetitive sequences, the short tandemly repeated repetitive sequences (STRRs), has been described (Mazel et al., 1990). These heptanucleotide sequences have been identified in several cyanobacterial genera and species, so far mostly in heterocystous cyanobacteria (Rasmussen & Svenning, 1998; Zheng et al., 1999; Nilsson et al., 2000; Wilson et al., 2000; Teaumroong et al., 2002; Lyra et al., 2005; Prasanna et al., 2006), but also in some non-heterocystous ones (Rasmussen & Svenning, 1998). Furthermore, a 37 bp long tandemly repeated repetitive sequence (LTRR) has also been identified in some cyanobacterial species (Masepohl et al., 1996; Rasmussen & Svenning, 1998; Prasanna et al., 2006). Analysis of STRRs and LTRRs has been described as a powerful tool for taxonomic studies (Mazel et al., 1990). Moreover, the specificity of these sequences has made STRRs useful even for non-axenic cyanobacterial cultures (Nilsson et al., 2000).

A universal marker for DNA fingerprinting is the oligonucleotide csM13. It has already been tested in a small number of cyanobacteria (Valério et al., 2005), and has a demonstrated ability even to discriminate strains of the same species. Techniques based on the enterobacterial repetitive intergenic consensus (ERIC) have also been used for identification and discrimination purposes in some cyanobacteria (Rasmussen & Svenning, 1998; Lyra et al., 2001; Valério et al., 2005; Bruno et al., 2006).

The restriction fragment length polymorphisms (RFLPs) of particular PCR products can provide signature profiles specific to the genus, species, or even strain. Genetic characterization of cyanobacterial strains has been undertaken using RFLPs of the 16S rRNA gene (16S-ARDRA) (Lyra et al., 1997) and of the intergenic transcribed spacer region (ITS-ARDRA) (Lu et al., 1997; West & Adams, 1997). Furthermore, amplification of the 16S–23S rRNA ITS, which has been shown to be variable in length (Rocap et al., 2002; Iteman et al., 2002; Neilan, 2002, Laloui et al., 2002) and number (West & Adams, 1997; Iteman et al., 2002) in cyanobacteria, can also be used as an identification tool.

A sequential polyphasic approach was used in this study. Isolates were identified by observation of their morphological features. A hierarchical analysis with STRR and LTRR PCR fingerprinting patterns was performed and representatives of the clusters obtained were identified by a phylogenetic analysis using two genes, one coding for the small-subunit rRNA (16S rRNA gene) and the other for the DNA-dependent RNA polymerase subunit (rpoC1). Subsequent characterization of all isolates by M13 and ERIC fingerprints allowed the discrimination of strains, revealing also the traceability potential of these last methods for routine freshwater monitoring. Furthermore, a diagnostic key was constructed for the identification of cyanobacterial species, based on the use of 16S-ARDRA, ITS length and ITS-ARDRA.

**METHODS**

**Cyanobacterial strains.** A total of 118 unicellular cyanobacterial non-axenic cultures, belonging to the orders Chroococcales (54), Oscillatoriiales (15) and Nostocales (49), mainly isolated from Portuguese freshwater blooms and maintained in the LMECYA culture collection (Cyanobacteria Culture Collection Estela Sousa e Silva – INSA) using Z8 medium (Skulberg & Skulberg, 1990), were used in this study (Supplementary Table S1, available with the online version of this paper). Isolates were kept under a 14:10 h light:dark cycle (20 ± 4 μmol m⁻² s⁻¹) at 20 ± 1 °C. Lyophilized samples of the isolates were obtained as described by Saker et al. (2003).

**Morphological identification.** The morphology of cells and filaments was studied using an Olympus BX60 light microscope with a digital camera. The following parameters were analysed: length and width of vegetative cells; morphology of the terminal cell; the presence or absence of heterocysts, akinetes and gas vesicles; the distance between heterocysts and the distance between a heterocyst and the nearest akinete; and finally, the shape of filaments and their potential aggregation into colonies.

**DNA extraction from cyanobacteria and heterotrophic bacteria.** Genomic DNA of cyanobacterial isolates was extracted according to a previously described method (Valério et al., 2005).

For isolation of heterotrophic (non-obligately oligotrophic) bacteria, associated with some cyanobacterial cultures (randomly selected), an aliquot of the supernatant of isolates LMECYA 64, AQ5, LMECYA 79, LMECYA 13 and LMECYA 173 was grown on LB agar (tryptone 1 % w/v; yeast extract 0.05 %, w/v; agar 2 %, w/v) at 22 °C for 3 days. The strains obtained were subsequently purified and grown in LB medium (tryptone 1 %, w/v; yeast extract 0.05 %, w/v; NaCl 0.05 %, w/v) overnight at 22 °C for DNA extraction. Cells were recovered by centrifugation and total DNA extracted using the guanidium thiocyanate method, according to Pitcher et al. (1989).

**PCR fingerprinting and data analysis.** STRR PCR fingerprinting was performed using both STRR1F (5'-CCCCCARTCCCCC-C3') and STRR3F (5'-CAAAGTCTACAGT-C3') primers (Wilson et al., 2000). For LTRR PCR fingerprinting both LTRR1 (5'-GGATTITGGTTTGTAGTAAAAAC-C3') and LTRR2 (5'-CTATCAGGGGATTGAAAG-C3') primers (Rasmussen & Svenning, 1998) were used.

M13 PCR fingerprinting was performed using oligonucleotide csM13 (5'-GAGGTGTGCGGCTTC-C3') (Huey & Hall, 1989) as single primer. For ERIC PCR fingerprinting both ERIC1R (5'-ATGTAA-...
GCTCCTGGGGATTCAC-3') and ERIC2 (5'-AAGTAAGTGACT-GGGTGACGG-3') primers (Versalovic et al., 1991) were used.

PCRs were performed in 50 μl containing 1 × PCR buffer (Invitrogen), 0.4 mM of each of the four dNTPs (Invitrogen), 0.5 mM (for STRR and LTRR fingerprinting) or 1 mM (for M13 and ERIC fingerprinting) of each primer, 10–15 ng genomic DNA, 3 mM (for STRR, LTRR and M13 fingerprinting) or 2 mM (for ERIC fingerprinting) MgCl₂, 1 mg BSA ml⁻¹ and 1 U Taq DNA polymerase (Invitrogen). Amplification was performed in a T Gradient thermocycler (Biometra), consisting of an initial denaturation step at 95 °C for 10 min, followed by 35 cycles of 90 s at 95 °C, 2 min at 56 °C (for M13 fingerprinting), 40 °C (for ERIC fingerprinting) or 38 °C (for STRR and LTRR fingerprinting) and 2 min at 72 °C, and a final extension step of 5 min at 72 °C.

The PCR products were electrophoresed in 1.4 % (w/v) agarose gel in 0.5 × TBE buffer at 90 V for 3 h 30 min, using the 1 kb plus DNA Ladder (Gibco-BRL) as molecular size marker. After staining with ethidium bromide, the image obtained under UV transillumination was digitized using the Kodak 1D 2.0 system.

To assess reproducibility of the fingerprinting methods according to Sneath & Johnson (1972), at least 10 % of isolates were randomly selected and analysed in duplicate in independent experiments. The similarity between each pair of duplicates was obtained from the analysis based on a dendrogram computed with Pearson’s correlation coefficient and the unweighted pair group method with arithmetic average (UPGMA) as the agglomerative clustering, using BioNumerics software v4.0 (Applied Maths) (data not shown). The reproducibility value was determined as the average similarity for all pairs of duplicates.

The hierarchical clustering of isolates, based on their genomic fingerprinting patterns, was performed with Pearson’s correlation coefficient and UPGMA using BioNumerics software v4.0. The accuracy of each dendrogram to represent the similarity matrix was assessed by the cophenetic correlation coefficient (ρ), which ideally should be between 0.6 and 0.9 (Priest & Austin, 1993). Composite analyses were performed by constructing dendrograms based on the patterns obtained with two different methods. Clusters of isolates were correlated with taxonomic positioning, geographical origin and available toxicity data.

The diversity within each cyanobacterial species comprising more than five isolates was evaluated with the indices of Simpson (Hunter & Gaston, 1988) and Shannon–Wiener (Zar, 1996). The Simpson index (D) is based on the number of types (fingerprinting profiles) and isolates for each type and measures the probability of two non-related strains, taken from the tested population, belonging to two different genomic types. The Shannon–Wiener index (J) also provides an evenness measure, expressing the observed diversity as the proportion of the possible maximum diversity and reflecting the homogeneity/heterogeneity of the distribution of isolates among the genomic types.

**PCR amplification and sequencing of 16S rRNA and rpoC1 genes.** Since universal primers for direct sequencing of 16S rRNA genes are usually designed to be used with axenic cultures, and available primers for rpoC1 amplification and sequencing are highly degenerate, specific primers were selected or designed in order to obtain clean sequences for both genes without the need for a cloning step. Primers used in the amplification and sequencing of the 16S rRNA and rpoC1 genes are listed in Table 1. The PCRs were performed in 50 μl containing 1 × PCR buffer (Invitrogen), 0.4 mM of each of the four dNTPs (Invitrogen), 0.25 mM of each of the two primers, 10–15 ng genomic DNA, 2 mM MgCl₂ (Invitrogen), 0.5 mg BSA ml⁻¹ (Invitrogen) and 1 U Taq DNA polymerase (Invitrogen). Amplification was performed in a T Gradient thermocycler.

**Table 1.** Primers used for 16S rRNA, rpoC1 and ITS region amplification and for 16S rRNA and rpoC1 sequencing.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
<th>Amplicon size (nt)</th>
<th>Target site</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>16S rRNA</td>
<td>CYA106F</td>
<td>CGGACGGGTGAGTAACGCGTGA</td>
<td>106–127</td>
<td>Nuβel et al. (1997)</td>
<td>Universal for cyanobacteria</td>
</tr>
<tr>
<td></td>
<td>CYAN1281R</td>
<td>GCAATTACTAGCGATTCCTCC</td>
<td>1260 1281–1302</td>
<td>Valeiro et al. (2005)</td>
<td>Universal for cyanobacteria</td>
</tr>
<tr>
<td>rpoC1</td>
<td>RPOCM61F</td>
<td>GGAAATGGATGGGTTATTCTGC</td>
<td>580 61–82</td>
<td>This study</td>
<td>Universal for cyanobacteria</td>
</tr>
<tr>
<td></td>
<td>RPOCM624R</td>
<td>TAAAACCATCCATTCTGCCTC</td>
<td>624–643</td>
<td>This study</td>
<td>Specific for Microcystis</td>
</tr>
<tr>
<td></td>
<td>RPOC683R</td>
<td>AARTTRTCAATTACCCGCA</td>
<td>555 683–701</td>
<td>This study</td>
<td>Except for Microcystis</td>
</tr>
<tr>
<td></td>
<td>RPOC1006R</td>
<td>TGCTTACCTTCAATAATGTC</td>
<td>880 1006–1025</td>
<td>This study</td>
<td>Universal for cyanobacteria</td>
</tr>
<tr>
<td>ITS region</td>
<td>ITSCYA236F</td>
<td>CTGGTTCRAGTCCAGGAT</td>
<td>Variable</td>
<td>This study</td>
<td>Universal for cyanobacteria</td>
</tr>
<tr>
<td></td>
<td>ITSCYA225R</td>
<td>TGCAGTTKTCAAGGTTCT</td>
<td></td>
<td></td>
<td>Universal for cyanobacteria</td>
</tr>
</tbody>
</table>

* F, Forward primer; R, reverse primer.

16S rRNA of *Synechocystis* PCC 7001. (NC000911), 16S rRNA of *Synechococcus* sp. WH7803 (WH7803 (CT971583)), 16S rRNA of *Microcystis* sp. WH7803 (WH7803 (CT971583)).
(Biometra) using the following conditions: 10 min of initial denaturation at 95 °C, followed by 35 cycles at 94 °C for 45 s, 55 °C for 45 s and 72 °C for 1 min. The amplification was completed by holding for 5 min at 72 °C to allow the complete extension of the PCR product. PCR products were visualized as described above but using 1 % (w/v) agarose gel. Amplification reaction products were purified with a Jet Quick-PCR Purification kit (Genomed) as described by the manufacturer. The purified PCR products were sequenced in both directions using an automated ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems) with dye terminators, using standard protocols and the cyanobacterial-specific primers used for amplification, as well as primers CYAN738F and CYAN738R (Table 1) for the 16S rRNA gene sequencing.

Phylogenetic analysis and identification at species level. A two-step strategy was used to obtain a 16S rRNA gene phylogeny for the cyanobacterial strains under analysis. Firstly, a separate phylogeny was constructed for each order (Chroococcales, Oscillatoriales and Nostocales), including the sequences from 71 LMECYA strains and reference sequences from well-characterized and reliably identified strains of validated species: sequences from completely sequenced genomes; sequences from the strains used by Suda et al. (2002) for the taxonomic revision of the genus Plankothrix; Limnothrix redekei sequences determined by Gkelis et al. (2005); and Microcystis spp. sequences determined by Otsuka et al. (2001).

In the next step, a subgroup of LMECYA isolates and reference organisms was selected from the phylogeny of each order as representatives of each tree structure and phylogenetic distances and a global 16S rRNA gene phylogeny was constructed to include the six cyanobacterial orders. As indicated in the figure legends, the alignments used encompassed 1182 or 1203 nucleotides of the 16S rRNA gene.

Since rpoC1 sequences were determined for only 38 LMECYA isolates and fewer reference sequences are available, a global phylogeny including the sequences of all these isolates and reference sequences was constructed using an alignment of 452 nucleotides. All alignments were made with CLUSTAL_X version 1.83 and visually corrected. Phylogenetic trees were constructed based on a Bayesian approach using MrBayes software version 3.0b4 (Huelsenbeck & Ronquist, 2001). Each analysis consisted of 2 000 000 generations from a random starting tree and four Markov chains (with default heating values) sampled every 100 generations. The first 8000 sampled trees were discarded, resulting in a set of 12 000 analysed trees sampled after the chains became stationary. The eubacterial Escherichia coli sequence was used as an outgroup. Neighbour-joining and maximum-parsimony trees were also constructed using the same alignments and PAUP 4.0b10 software (Swofford, 2003).

The BLAST tool of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) was used to find homologous and other close sequences (97–100 % identity and E-values < 10^{-20}) to be included in the phylogenies as reference sequences.

To achieve the final identification of the isolates, the following criteria were used. When the phylogenetic positioning of an isolate retrieved identification at the species level, this prevailed; if the phylogenetic positioning allowed only genus determination or no molecular identification was possible, then the identification at the species level obtained by morphological analysis was retained.

16S rRNA gene and ITS analysis. To select the most suitable restriction endonucleases to provide an easily identifiable 16S rRNA gene restriction profile, a theoretical study was performed with NEBcutter v2.0 software (Vincze et al., 2003), using cyanobacterial 16S rRNA gene sequences available in databases, resulting in the selection of AvaI and BanII enzymes. For 16S-ARDRA, approximately 1 µg of the amplicon of 1260 bp, obtained with CYAN106F and CYAN128R primers (Table 1), was digested without further purification with AvaI (BioLabs) and BanII (Takara) restriction endonucleases in separate reactions, according to the manufacturer’s instructions.

The ITS region was amplified using DNA of non-axenic cultures and two cyanobacterial-specific primers designed in this study, ITSCYA236F and ITSCYA225R (Table 1). The PCRs were performed in 50 µl containing 1 × PCR buffer (Invitrogen), 0.4 mM of each of the four dNTPs (Invitrogen), 0.5 mM of each primer, 10–15 ng genomic DNA, 2.5 mM MgCl2, 0.5 mg BSA ml^{-1} and 1 U Taq DNA polymerase (Invitrogen). Amplification was performed in a T Gradient thermocycler (Biometra), consisting of an initial denaturation step at 95 °C for 6 min, followed by 35 cycles of 45 s at 95 °C, 45 s at 52 °C and 1 min at 72 °C, and a final extension step of 5 min at 72 °C.

Also in this case, a theoretical study was performed with NEBcutter v2.0 software (Vincze et al., 2003), using cyanobacterial ITS sequences available in databases, resulting in the selection of TaqI enzyme.

For ITS-ARDRA, amplified DNA (10 µl, approximately 1 µg) was digested without further purification with TaqI restriction endonuclease (Fermentas) according to the manufacturer’s instructions.

ITS amplicons and restriction products for 16S-ARDRA and ITS-ARDRA were visualized as described above, using 1.2 % (w/v) and 2 % (w/v) agarose gels, respectively.

RESULTS

Morphological identification

Identification of isolates at the species or genus level was previously assessed by their morphology (Supplementary Table S2). Out of the 118 isolates, 99 could be identified at the species level and the remaining 19 only at the genus level. These results suggest that the cyanobacterial species present in Portuguese freshwaters, most of which are responsible for bloom formation, are diverse, and belong to three of the six cyanobacterial orders: Chroococcales, Oscillatoriales and Nostocales. Of the 118 isolates under study, 54 (45.8 %) belong to Chroococcales, 15 (12.7 %) to Oscillatoriales and 49 (41.5 %) to Nostocales. The species predominantly found were Microcystis aeruginosa (52 isolates; 44 %) and Aphanizomenon flos-aquae (13 isolates; 11 %).

Assessment of PCR fingerprinting reliability

The average reproducibility of the tested fingerprinting techniques, estimated by the similarity average value for all pairs of duplicates, was 92 %, representing a good reproducibility (Sneth & Johnson, 1972), since the Pearson’s correlation coefficient used is very sensitive to the band intensity, leading to lower final similarity values. In all dendrograms the cophenetic value (ρ) is above 0.85, which corresponds to a good representation of the original data.

In order to assess the effect of the non-axenic nature of the uni-cyanobacterial cultures, the PCR fingerprinting...
methods to be tested (STRR, LTRR, M13 and ERIC) were applied to a set of cyanobacterial cultures and their co-cultured heterotrophic (non-obligately oligotrophic) bacteria. From the same culture (Fig. 1 underlined group of lanes), cyanobacteria (in bold) and cultivated heterotrophic bacteria displayed M13 (Fig. 1c) and ERIC (Fig. 1d) PCR fingerprints with very few co-migrating fragments (white boxes). However, isolates LMECYA 79 and LMECYA 173 displayed identical fingerprinting profiles with each technique, despite the fact that their corresponding cultivated bacteria showed M13 and ERIC profiles different from each other but with some co-migrating fragments with the cyanobacteria (Fig. 1, lanes 8–11). On the other hand, STRR and LTRR PCR fingerprinting (Fig. 1a, b) showed specificity for cyanobacteria, since no amplification was obtained for the cultivated bacteria.

**STRR and LTRR PCR fingerprinting**

A composite analysis of hierarchical clustering of STRR and LTRR patterns was performed with all isolates (data not shown) and subsequently for each order as presented in Fig. 2. It can be observed that the isolates grouped according to their genomic similarity. To confirm the species identification, a set of representative isolates (marked with bold italics in Fig. 2), including members of different clusters and members of the same cluster, was selected and their identification was obtained or confirmed by phylogenetic positioning using 16S rRNA gene and rpoC1 phylogenies.

The identification presented in Fig. 2 is in agreement with the final identification of the isolates obtained by 16S rRNA gene sequencing (which in some cases was also confirmed by the rpoC1 sequencing, as presented below). It can be observed that the majority of the clusters correspond to different species, showing the taxonomic potential of STRR and LTRR fingerprints.

The three species (two of them represented by only one isolate) belonging to the order Chroococcales (Fig. 2a) are intermixed and could not be separated according to their STRR and LTRR fingerprinting patterns. In fact, the isolates of *Microcystis aeruginosa* show very diverse profiles (grouping in three distinct clusters) and the isolate of *Synechocystis* sp. is more similar (52.5%) to cluster II of *M. aeruginosa* than the *M. aeruginosa* clusters are to each other (15–27%). *Synechococcus nidulans*, together with one isolate of *M. aeruginosa*, has the lowest similarity (10%) with all the other isolates. Fingerprinting patterns of isolates belonging to the order Oscillatoriales (Fig. 2b) led to the formation of two clusters and four isolated strains at 40% similarity, while nine major clusters and four isolated strains were observed at 50% similarity for isolates of the order Nostocales (Fig. 2c). In both cases almost all the clusters correspond to different species, except for *Anabaena circinalis* and *Aphanizomenon gracile*, which formed two distinct clusters (genotypes).

![Fig. 1. Fingerprinting profiles of cyanobacteria and cultivable bacteria from each non-axenic culture. (a) STRR PCR fingerprinting; (b) LTRR PCR fingerprinting; (c) M13 PCR fingerprinting; (d) ERIC PCR fingerprinting.](image)
Fig. 2. Dendrograms obtained by composite hierarchical analysis of STRR and LTRR fingerprinting profiles using Pearson's correlation coefficient and the UPGMA clustering method for the orders used in the study: (a) Chroococcales, (b) Oscillatoriales and (c) Nostocales. The cophenetic correlations (ρ) of the dendrograms are shown. The scale corresponds to global percentage similarity. gen, genotype. A 'weight of 2' was attributed to the STRR patterns, and a 'weight of 1' to the LTRR patterns.
Phylogenetic analysis using the 16S rRNA gene

The phylogenetic tree presented in Fig. S1 corresponds to the partial 16S rRNA gene phylogeny of the order Chroococcales. This phylogeny confirmed the identification of the 27 sequenced isolates only at the genus level. In fact, although the genus Microcystis forms a monophyletic cluster, all five species are highly homologous and thus cannot be distinguished at the 16S rRNA gene level.

The partial 16S rRNA gene phylogeny of the order Oscillatoriales is depicted in Fig. S2; it allowed the identification at species level of seven sequenced isolates (out of 10), four of these (LMECYA 152, 153, 153E and 153F) as Planktothrix agardhii, one (LMECYA 203) as Planktothrix rubescens, one (LMECYA 162) as Planktothrix pseudoagardhii and one (LMECYA 145) as Limnothrix redekei. The three remaining isolates (LMECYA 79, 173 and 214) were grouped in a cluster containing sequences of Phormidium and Leptolyngbya isolates. However, taking into account their morphological characteristics, these isolates were identified as Phormidium sp. The distribution of Oscillatoria and Lyngbya isolates in several distinct clades also reveals the polyphyletic nature of these genera.

The partial 16S rRNA gene phylogeny of the order Nostocales (Fig. S3) allowed the identification of 21 out of 34 sequenced isolates. The isolate LMECYA 177 groups in a cluster where several Anabaena species are present. This phylogenetic positioning does not allow an identification of this isolate; however, its morphological features are similar to those of Anabaena planctonica, which is also present in the cluster. This is also the case with isolates 77A, 77B, 88, 99, 125, 129 and 141, grouped in a cluster with mainly Aphaniizomenon flos-aquae but also Aphaniizomenon gracile. However, morphological features of these isolates are typical of Aphaniizomenon flos-aquae. Isolates LMECYA 9 and 33 grouped in a cluster with Anabaena flos-aquae, Aphaniizomenon flos-aquae and Aphaniizomenon gracile sequences, but morphological features of these isolates are typical of Aphaniizomenon gracile. The isolates LMECYA 178, 182 and 185 did not group with any species or genus. In fact, their morphology is congruent with Anabaena, and LMECYA 178 is identical to Anabaena flos-aquae. This phylogeny also showed that Anabaena and Aphaniizomenon do not form monophyletic clusters, with isolates of both genera intermixed along the tree.

The global 16S rRNA gene phylogeny of representative species of the six cyanobacterial orders, using a Bayesian approach, is illustrated in Fig. 3(a). It shows that the Nostocales form a monophyletic cluster, except for the sole Scytonema sequence presented. Stigonematales are also monophyletic, containing the genera Chlorella and Nostocopsis, except the unique Scytonema sequence here presented. The order Pleurocapsales forms a monophyletic cluster, whereas the orders Chroococcales and Oscillatoriales are both divided into two distinct clades. Identical results were obtained using neighbour-joining and maximum-parsimony trees (Supplementary Fig. S4).

Phylogenetic analysis using rpoC1

The rpoC1-based phylogeny presented in Fig. 3(b) allowed the identification of six of the 38 sequenced isolates. Two of them (LMECYA 17 and 126) belong to Anabaena circinalis, one (LMECYA 134) to Cylindrospermopsis raciborskii, one (LMECYA 153E) to Planktothrix agardhii and one (LMECYA 68) to Synechocystis sp. All these identifications were congruent with those obtained by 16S rRNA gene sequencing.

Isolates LMECYA 9, 40, 64 and 148, identified both morphologically and by 16S rRNA gene sequencing as Aphaniizomenon gracile, clustered into two distinct groups, neither of them closely related to the available Aphaniizomenon gracile rpoC1 sequence.

Phormidium sp. isolates (LMECYA 79 and 173) grouped into a cluster containing another Phormidium sequence but also Leptolyngbya and Synechococcus sequences.

Some isolates did not group with a particular species, due to the low number of rpoC1 sequences available.

In the rpoC1 phylogeny, the Nostocales split into two clusters, a major one containing almost all species of the clade obtained with the 16S rRNA gene and a second grouping the remaining species, including Scytonema sp., with species from Stigonematales and Pleurocapsales. Oscillatoriales were separated in distinct clades, mostly equivalent to the 16S rRNA ones. Chroococcales presented a monophyletic cluster, except for the Prochloron and two Synechococcus spp. sequences, which were distributed in other clusters. When this rpoC1 Bayesian tree was compared with neighbour-joining and maximum-parsimony trees (see Supplementary Fig. S5), similar phylogenetic relationships were observed.

Strain discrimination and traceability

Representative M13 and ERIC PCR fingerprinting profiles of the genomic types of the species comprising more than five isolates are presented in Supplementary Fig. S6. As this figure shows, the richness and diversity of the DNA banding patterns allowed the discrimination of all distinct isolates and the detection of identical ones.

Individual dendograms were constructed for each of the seven species (data not shown). Based on the M13 and ERIC PCR fingerprints of all the Microcystis aeruginosa isolates evaluated by a composite analysis of hierarchical clustering, a cut-off level of 75% similarity was established for definition of a genomic type. Table 2 presents the percentage of types and values of Simpson and Shannon–Wiener diversity indices (for the species with more than five isolates). The lowest diversity values were obtained for Planktothrix agardhii, where four of the nine isolates have the same genomic type. The other four species whose diversity indices were determined revealed higher diversity values, with Aphaniizomenon flos-aquae showing the highest one.
Fig. 3. (a) Global MrBayes tree of representative species using 1182 nt of the 16S rRNA gene sequences in the alignment. (b) MrBayes tree of the rpoC1 sequences using 452 nt in the alignment. Posterior probabilities values above 0.5 are indicated to provide branch support. Bars: 10% sequence divergence. GenBank accession numbers are indicated after the species designation (names in bold correspond to sequences determined in this study).
In the dendrogram of all *Microcystis aeruginosa* isolates (Supplementary Fig. S7), most of those sourced from the same reservoir seem to group together and a general pattern of association also emerges in terms of microcystin-producing ability. In fact, two major clusters were formed at the 20% similarity level: cluster I with 22 toxic and five non-toxic isolates, and cluster II with 18 non-toxic and 2 toxic isolates. Furthermore, only for one type defined at 75% similarity did a toxic isolate (from Magos reservoir) group with a non-toxic one (from Roxo reservoir).

To test the traceability potential of these fingerprints, a dendrogram comprising a group of *Microcystis aeruginosa* isolates, recovered from the same reservoir (Montargil) on different dates, was constructed (Fig. 4a) and 10 genomic types were defined (at the 75% similarity level). Although the non-toxin-producing isolates (NT) were distributed into distinct clusters, all the toxin-producing ones (T) were also grouped together.

![Composite hierarchical analysis of M13 and ERIC fingerprints (a) and genomic and temporal relationships amongst isolates (b) of *Microcystis aeruginosa* sourced from Montargil reservoir on different dates. NT, non-toxic; T, toxic.](image)

Table 2. Genomic diversity of cyanobacterial species

<table>
<thead>
<tr>
<th>Species</th>
<th>% types*</th>
<th>D†</th>
<th>J‡</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Microcystis aeruginosa</em></td>
<td>59.6 (28/47)</td>
<td>0.96</td>
<td>0.93</td>
</tr>
<tr>
<td><em>Planktothrix agardhii</em></td>
<td>66.7 (6/9)</td>
<td>0.83</td>
<td>0.88</td>
</tr>
<tr>
<td><em>Anabaena circinalis</em></td>
<td>85.7 (6/7)</td>
<td>0.95</td>
<td>0.98</td>
</tr>
<tr>
<td><em>Aphanizomenon flos-aquae</em></td>
<td>92.3 (12/13)</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td><em>Aphanizomenon gracile</em></td>
<td>75.0 (6/8)</td>
<td>0.93</td>
<td>0.97</td>
</tr>
<tr>
<td><em>Aphanizomenon issatschenkoi</em></td>
<td>80.0 (4/5)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><em>Cylindrospermopsis raciborskii</em></td>
<td>80.0 (4/5)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*% types: (number of types/number of isolates) × 100.
†D, Simpson diversity index.
‡J, Shannon–Wiener diversity index.

Table 2. Genomic diversity of cyanobacterial species

<table>
<thead>
<tr>
<th>Species</th>
<th>% types*</th>
<th>D†</th>
<th>J‡</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Microcystis aeruginosa</em></td>
<td>59.6 (28/47)</td>
<td>0.96</td>
<td>0.93</td>
</tr>
<tr>
<td><em>Planktothrix agardhii</em></td>
<td>66.7 (6/9)</td>
<td>0.83</td>
<td>0.88</td>
</tr>
<tr>
<td><em>Anabaena circinalis</em></td>
<td>85.7 (6/7)</td>
<td>0.95</td>
<td>0.98</td>
</tr>
<tr>
<td><em>Aphanizomenon flos-aquae</em></td>
<td>92.3 (12/13)</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td><em>Aphanizomenon gracile</em></td>
<td>75.0 (6/8)</td>
<td>0.93</td>
<td>0.97</td>
</tr>
<tr>
<td><em>Aphanizomenon issatschenkoi</em></td>
<td>80.0 (4/5)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><em>Cylindrospermopsis raciborskii</em></td>
<td>80.0 (4/5)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*% types: (number of types/number of isolates) × 100.
†D, Simpson diversity index.
‡J, Shannon–Wiener diversity index.
grouped together at a similarity level above 55%, indicating the ability of M13 and ERIC fingerprints to discriminate toxic from non-toxic isolates. Two isolates, LMECYA 92C and LMECYA 110, recovered in different years, showed 92% similarity (reproducibility level; dotted line in Fig. 4a), indicating that they should be assumed to represent the same strain.

In spite of the reduced number of isolates under analysis, the cross-analysis of temporal occurrence and genomic relationships (Fig. 4b) points to the probable independent introduction of seven genomic types (two of them toxic), as well as the existence of resident and evolving genomic types (types 4-3, 6-7 and 8-9), revealing traceability potential for these fingerprints.

**Diagnostic key for cyanobacterial identification**

With the aim of achieving fast and reliable identification of the isolates, PCR amplification and subsequent restriction of 16S rRNA gene and ITS region were tested.

The sizes of the fragments obtained after 16S rRNA gene digestion are summarized in Table 3. After restriction with *Ava*II four out of the 18 species are identified: *Limnothrix redekei* (630 bp, double band), *Cylindrospermopsis raciborskii* (1100 and 160 bp bands), *Phormidium* sp. (900 and 360 bp bands) and *Synechocystis* sp. (960 and 300 bp bands). For the remaining species, seven show a 1260 bp band (no restriction) and the others show a profile of 810 and 450 bp bands. Further restriction with *Ban*I enabled the identification of seven out of the 14 species not previously identified. However, *Microcystis aeruginosa*, *Anabaena circinalis*, *Anabaena flos-aquae*, *Anabaena planctonica*, *Anabaena spiroides*, *Aphanizomenon flos-aquae* and *Aphanizomenon gracile* cannot be distinguished, since they all show a profile with two bands of 815 and 445 bp.

The ITS amplicons obtained with the cyanobacterial-specific primers designed in this study retrieved the amplification fragments also summarized in Table 3. The number of amplified fragments varied between one and three, and their size ranged from 180 to 930 bp. Most of the species included in the Nostocales showed a 300 bp fragment. The range of ITS sizes for members of the genus *Microcystis* overlapped in some cases with *Anabaena* and *Aphanizomenon*. Some amplification profiles can also

<table>
<thead>
<tr>
<th>Table 3. 16S-ARDRA fragments generated with Avall and BanII endonucleases, ITS amplicon dimensions and ITS-ARDRA fragments generated with TaqI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>16S-ARDRA fragments (bp)</strong></td>
</tr>
<tr>
<td><strong>AvalI</strong></td>
</tr>
<tr>
<td>Chroococcales</td>
</tr>
<tr>
<td><em>Microcystis aeruginosa</em></td>
</tr>
<tr>
<td><em>Synechococcus nidulans</em></td>
</tr>
<tr>
<td><em>Synechocystis sp.</em></td>
</tr>
<tr>
<td>Oscillatoriales</td>
</tr>
<tr>
<td><em>Limnothrix redekei</em></td>
</tr>
<tr>
<td><em>Planktothrix agariphii</em></td>
</tr>
<tr>
<td><em>Planktothrix pseudoagariphii</em></td>
</tr>
<tr>
<td><em>Planktothrix rubescens</em></td>
</tr>
<tr>
<td><em>Phormidium sp.</em></td>
</tr>
<tr>
<td>Nostocales</td>
</tr>
<tr>
<td><em>Anabaena circinalis</em></td>
</tr>
<tr>
<td><em>Anabaena flos-aquae</em></td>
</tr>
<tr>
<td><em>Anabaena planctonica</em></td>
</tr>
<tr>
<td><em>Anabaena spiroides</em></td>
</tr>
<tr>
<td><em>Anabaenopsis circularis</em></td>
</tr>
<tr>
<td><em>Aphanizomenon flos-aquae</em></td>
</tr>
<tr>
<td><em>Aphanizomenon gracile</em></td>
</tr>
<tr>
<td><em>Aphanizomenon issatschenkoi</em></td>
</tr>
<tr>
<td><em>Aphanizomenon ovalisporum</em></td>
</tr>
<tr>
<td><em>Cylindrospermopsis raciborskii</em></td>
</tr>
</tbody>
</table>

*ND, not determined.*
provide an identification, as is the case for *Synechocystis* sp. (170 bp band), *Limnothrix redekei* (350 and 250 bp bands), *Anabaena flos-aquae* (420 and 300 bp bands), *Anabaenopsis circularis* (380 bp band), *Aphanizomenon ovalisporum* (315 bp band) and *Cylindrospermopsis raciborskii* (300 and 195 bp). Moreover, a 220 bp fragment was always obtained for the *Planktothrix* spp. isolates.

The ITS-ARDRA fragments obtained after restriction of the ITS with *TaqI* are summarized in Table 3. Some of the isolates show a profile that can be used for identification. The genotypic pattern produced using this restriction enzyme can provide a species identification marker: for instance *Planktothrix pseudoagardhii*, which showed a profile distinct from the other two *Planktothrix* species, *Planktothrix agardhii*, difficult to distinguish from the former by morphological characteristics, and *Planktothrix rubescens*, which can be identified by its cell dimensions and brownish colour.

Based on the 16S-ARDRA, ITS amplification and ITS-ARDRA a diagnostic key was constructed (Fig. 5) allowing the identification of 15 out of the 18 species analysed. Only *Anabaena planctonica*, *Anabaena spiroides* and *Aphanizomenon flos-aquae* could not be distinguished.

**DISCUSSION**

In this study, several molecular techniques were used to examine the level of genetic diversity of 118 cyanobacterial isolates for taxonomic purposes.

All the PCR fingerprinting methods showed good reproducibility, and the inter- and intra-specific variability, observed in all species analysed, was revealed to be appropriate for assessment of genomic relationships, isolate discrimination and traceability, in spite of the non-axenic character of the cyanobacterial cultures. STRR fingerprints were obtained for a larger number of non-heterocystous cyanobacteria than described so far (Rasmussen & Svenning, 1998) and the LTRR PCR fingerprinting was also tested in a large number of species. Beyond some discrimination ability, the specificity of the STRR and LTRR fingerprinting patterns revealed identification potential. In fact, all species from the Nostocales and Oscillatoriales could be discriminated by the composite hierarchical analysis of the STRR and LTRR fingerprints when only species/genera of the same order were analysed. For most species, a good correlation was found between morphological identification and genomic clustering. In the case of *Anabaena circinalis* two clusters were formed, corresponding to two genotypes whose 16S rRNA gene was 96.5% similar, whereas the isolates of the same genotype displayed 99.9% similarity. A similar finding was observed for *Aphanizomenon gracile* isolates, which were also divided into two clusters, showing 98% 16S rRNA gene similarity between them and 99–100% within the clusters. The composite hierarchical analysis of the STRR and LTRR fingerprints reflects the identification results obtained by 16S rRNA gene sequencing, and thus can be used as an identification tool, decreasing the need for 16S rRNA gene sequencing, often used in association with the morphological identification of a cyanobacterial strain (Nübel et al., 1997).

The genus *Microcystis* forms a monophyletic group distant from the other genera in both 16S rRNA and *rpoC1* phylogenies. All five *Microcystis* species showed a high homology (>98.5% with both genes), preventing their discrimination. This has already been shown by Otsuka et al. (2001), who proposed the unification of all five species, with *Microcystis aeruginosa* having nomenclatural priority. However, despite the high sequence homology and also the high DNA–DNA reassociation levels found among these species, Kondo et al. (2000) suggested their maintenance as distinct taxa based on their distinct morphological features.

Both 16S rRNA and *rpoC1* gene phylogenies showed that the genera *Synechococcus* and *Synechocystis* form distinct clades. Nevertheless, our results recommend a revision of the genus *Synechococcus* as already suggested by Robertson et al. (2001).

Regarding the order Oscillatoriales, we also found that the genus *Planktothrix* forms a monophyletic cluster closer to Nostocales than to other Oscillatoriales, as already described by Suda et al. (2002). It is also evident that the genus *Oscillatoria* still needs major revision, its species being widely distributed in 16S rRNA gene trees. In fact, this genus has recently undergone revisions (Suda et al., 2002) and its polyphyly was also inferred by others (Ishida et al., 2001; Marquardt & Palinska, 2007; Suda et al., 2002). Similarly to Joyner et al. (2008), a polyphyletic origin of the genus *Lyngbya* was also inferred from the 16S rRNA gene tree, justifying its taxonomic revision. This might be related to the fact that traditional criteria used for classification of the Oscillatoriales at the genus level predominantly rely on the characteristics of external sheaths and colony formation, rather than on cellular features (Marquardt & Palinska, 2007). Other 16S rRNA analyses have already demonstrated the polyphyletic origin of members of the orders Chroococcales (Litvaitis, 2002; Seo & Yokota, 2003) and Oscillatoriales (Ishida et al., 2001; Litvaitis, 2002).

We also can verify from the *rpoC1* phylogeny that the order Nostocales is not monophyletic and that its members are closely related to those of the Stigonematales, giving additional support to the statement already made by Gugger & Hoffmann (2004) based on 16S rRNA gene phylogeny. *Anabaena* and *Aphanizomenon* isolates were genetically heterogeneous and intermixed in both gene trees, confirming previous results obtained with other strains (Lyra et al., 2001; Gugger et al., 2002; Rajaniemi et al., 2005). Furthermore, the possible presence of incorrectly identified sequences in databases may also distort the inferred conclusions. Although the results obtained with *rpoC1* phylogeny seem to indicate possible lateral gene transfer events in *Anabaena* and *Aphanizomenon* species, the hypothesis that the morphological features are
Fig. 5. Diagnostic key based on 16S-ARDRA with Avall and BanII, ITS amplification and ITS-ARDRA with TaqI, for cyanobacterial identification purposes. C, Chroococcales; O, Oscillatoriales; N, Nostocales.
insufficient to delineate the true taxonomic structure of these genera cannot be ruled out. *Cylindrospermopsis raciborskii* forms a monophyletic group, more distant from the other species of the Nostocales with both markers.

The identification of genomic types by hierarchical clustering analysis of M13 and ERIC fingerprints allowed the assessment of intra-specific diversity for the representative species of this study and also revealed this approach to be a useful tool for traceability purposes within a freshwater reservoir.

Although a rather low number of isolates was analysed for the majority of the species, except for *Microcystis aeruginosa*, a high diversity was observed within all species, with *Planktothrix agardhii* showing the lowest diversity values ($D=0.83; J^*=0.88$) and *Anabarizomenon flos-aquae* the highest ones ($D=J^*=0.99$). Furthermore, and except for *Planktothrix agardhii*, a high diversity of genomic types was usually found amongst isolates of the same species collected from the same reservoir.

Cross-analysis of temporal occurrence and genomic relationships, performed with *Microcystis aeruginosa* isolates collected over several years from Montargil reservoir, allowed inference of the dynamics of genomic types within a reservoir, illustrated the traceability potential of M13 and ERIC PCR fingerprinting and, as already observed for *Cylindrospermopsis raciborskii* isolates (Valério et al., 2005), highlighted its usefulness for ecological studies and evaluation of populations, when high resolution is needed.

A method to identify cyanobacterial isolates was developed through the construction of a diagnostic key based on the amplification of the 16S rRNA gene and further digestion with one or two restriction endonucleases. For those cases where no identification could still be obtained, the amplification or even further restriction of the ITS region with *TaqI* provided the identification of most species under study. This diagnostic key provides a faster, easier and reliable method to perform cyanobacterial identification for 15 out of the 18 species studied, without the need for expert experience to perform morphological identification, thus avoiding a misidentification of isolates, as frequently happens (Komárek & Anagnostidis, 1989). As an example of the application potentiality of this method, if the key here presented had been applied to the 118 isolates included in this study 84.7 % of them (100 out of 118) would have been immediately identified at species level. However, for the three species not discriminated by this key, as well as for other species where morphological features prevail in terms of species delimitation, a morphological analysis will be more reliable or, at least, will probably be required. Therefore, and as envisaged by Rita Colwell almost four decades ago, a polyphasic approach merging traditional identification and molecular characterization will certainly be the most powerful option to detect, identify and characterize cyanobacterial species and strains, including the harmful ones.

**ACKNOWLEDGEMENTS**

The authors are deeply indebted to the anonymous journal referees for their thorough review and helpful comments on both scientific issues and terminology. Elisabete Valério is recipient of Fundação para a Ciência e a Tecnologia grant SFRH/BD/8272/2002.

**REFERENCES**


Molecular diagnostics for cyanobacteria


Edited by: D. J. Scanlan