INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

RESEARCH ARTICLE

Delbaje *et al.*, *Int. J. Syst. Evol. Microbiol.* 2021;71:004648
DOI 10.1099/ijsem.0.004648

MICROBIOLOGY SOCIETY

Phylogenomic analysis of *Anabaenopsis elenkinii* (Nostocales, Cyanobacteria)

Endrews Delbaje[1], Ana Paula D. Andreote[1], Thierry A. Pellegrinetti[1], Renata B. Cruz[1], Luis H. Z. Branco[2] and Marli F. Fiore[1,*]

Abstract

The saline-alkaline lakes (soda lakes) are the habitat of the haloalkaliphilic cyanobacterium *Anabaenopsis elenkinii*, the type species of this genus. To obtain robust phylogeny of this type species, we have generated whole-genome sequencing of the bloom-forming *Anabaenopsis elenkinii* strain CCIBt3563 isolated from a Brazilian soda lake. This strain presents the typical morphology of *A. elenkinii* with short and curved trichomes with apical heterocytes established after separation of paired intercalary heterocytes and also regarding to cell dimensions. Its genome size is 4495068 bp, with a G+C content of 41.98%, a total of 3932 potential protein coding genes and four 16S rRNA genes. Phylogenomic tree inferred by RAxML based on the alignment of 120 conserved proteins using GTDB-Tk grouped *A. elenkinii* CCIBt3563 together with other genera of the family Aphanizomenonaceae. However, the only previous available genome of *Anabaenopsis circularis* NIES-21 was distantly positioned within a clade of *Desikacharya* strains, a genus from the family Nostocaceae. Furthermore, average nucleotide identity values from 86–98% were obtained among NIES-21 and *Desikacharya* genomes, while this value was 76.04% between NIES-21 and the CCIBt3563 genome. These findings were also corroborated by the phylogenetic tree of 16S rRNA gene sequences, which also showed a strongly supported subcluster of *A. elenkinii* strains from Brazilian, Mexican and Kenyan soda lakes. This study presents the phylogenomics and genome-scale analyses of an *Anabaenopsis elenkinii* strain, improving molecular basis for demarcation of this species and framework for the classification of cyanobacteria based on the polyphasic approach.

INTRODUCTION

The cyanobacterial genus *Anabaenopsis* belongs to the family Aphanizomenonaceae of the order Nostocales [1]. According to Komárek [2] the first taxa were described by G. S. West in 1907 [3] as a planktic members of the genus *Anabaena* inhabiting lakes of Eastern Africa. After Wołoszyńska [4] established the section *Anabaenopsis* inside of the genus *Anabaena* based on similar populations from Indonesia (Java), Miller [5] raised this section to genus level and the species *Anabaenopsis elenkinii* was chosen as the type species. The main distinguishing morphological feature of *Anabaenopsis* is the formation of paired intercalary heterocytes after asymmetrical division of two neighbouring vegetative cells in a trichome [2, 5, 6]. Trichomes often disintegrate between paired heterocytes and give rise to short trichomes with apical heterocytes, which grow in irregular or regular spirals or screw-like coils [2]. Akinetes develop solitary or several in a short row, intercalary, arise always paraheterocytic, but usually a slightly distant from heterocytes [7].

Anabaenopsis has been recorded in Asia, North America, South America, Africa, Europe and Australia/Oceania [2, 8–10]. All *Anabaenopsis* species described so far are planktonic and are usually distributed in water bodies of tropical and subtropical regions, but occur commonly also during the summer in warmer areas of temperate zones [2]. The majority of species inhabit brackish and saline-alkaline waters [2, 8–13] and some of them can increase population abundance to excessive levels and form visible blooms in suitable environmental conditions [2, 10, 11, 13]. The type species *Anabaenopsis elenkinii* was firstly observed in a muddy ditch near the city of Iwanovo-Vosnesenskii in Central Russia [5]

Author affiliations: [1]University of São Paulo, Center for Nuclear Energy in Agriculture, Avenida Centenário 303, 13400-970 Piracicaba, SP, Brazil; [2]São Paulo State University, Institute of Bioscience, Languages and Exact Sciences, 15054-000 São José do Rio Preto, SP, Brazil.
***Correspondence:** Marli F. Fiore, fiore@cena.usp.br
Keywords: Cyanobacteria; Genome; Soda lake; Ecology.
Abbreviations: AAI, average amino acid indentity; ANI, average nucleotide identity; BI, Bayesian inference; GTDB-Tk, genome taxonomy database toolkit; ML, maximum-likelihood; NCBI, National Centre for Biotechnology Information; NJ, neighbour-joining; POCP, percentage of conserved proteins.
The GenBank accession number for the genome sequence of CCIBt3563 is CP063311.
Two supplementary tables and one supplementary figure are available with the online version of this article.

and has been shown to be a common inhabitant of soda lakes [8, 10, 13–16].

Anabaenopsis is a morphologically well–defined genus, which has been confirmed genetically through phylogeny of 16S rRNA gene sequences [8, 13, 17, 18]. Although the infrageneric diversity has been reported as difficult to distinguish and delimit due to low variation between populations and the existence of transitional forms between species [19], 37 morphologically described species are cited in the AlgaeBase [20]. Nevertheless, in the CyanoDB database only three species (*A. elenkinii*, *A. gangetica* and *A. philippinensis*) are mentioned [21]. *Anabaenopsis* species proposed so far were based on phenotypic description and therefore a careful taxonomic revision is required using a polyphasic approach that incorporates genotypic properties. Furthermore, as full-genome sequences are made available it will be possible to apply the whole-genome average nucleotide identity (ANI) method for delimiting cyanobacterial species [22]. The ANI of all orthologous genes shared between any two genomes offers robust resolution between strains of the same or closely related species as an alternative for the labour-intensive DNA–DNA hybridization technique [23, 24].

Sequences of ribosomal 16S RNA of *Anabaenopsis* are limited as they have only been determined for five species and in phylogenetic analysis the majority of species appeared intermixed [8, 13, 16], emphasizing the need for species revision. In the NCBI GenBank only one genome of *Anabaenopsis* (*Anabaenopsis circularis* NIES-21, GCF_002367975.1) is available, but in a phylogenomic analysis it grouped within a *Nostoc*-like clade [25]. Therefore, the aim of this study was to characterize a strain of *Anabaenopsis elenkinii* using a combination of morphological, ecological and genomic information in order to improve taxonomic resolution of the bloom-forming *Anabaenopsis* type species.

METHODS

Cyanobacterium strain and culture conditions

The *Anabaenopsis elenkinii* strain CCIBt3563 was isolated from a water bloom sample collected on 6 May 2012, at the soda lake 'Salina da Reserva' located at the Nhumirim Farm, municipality of Corumbá, sub-region of Nhecolândia, Mato Grosso do Sul state, Brazil (18°57′35″S, 56°37′18″W). Salinity, electrical conductivity and pH of 'Salina da Reserva' water were measured *in situ* using a multiparameter probe WTW 340i by Santos *et al.* [13]. The CCIBt3563 strain is kept at the Centre for Nuclear Energy in Agriculture/University of São Paulo in 125 ml Erlenmeyer flasks containing 50 ml of Z8 liquid medium [26] modified by adjusting the medium to pH 9.4 with NaOH and salinity with NaCl 7.5 g l^{-1}. The cultures incubated in a growth room are illuminated with 40–50 µmol photons·µm^{-2}·s^{-1} under white fluorescent light, using a 14:10 h light:dark cycle, at 21±1 °C and humidity of 60±5%. Subsamples of cultured material were preserved in 4% formaldehyde (*v/v*) and deposited in the 'Maria Eneyda P. Kauffman Fidalgo' Herbarium (SP) of the Institute of Botany, São Paulo state, Brazil (CCIBt3563 – voucher SP 428475).

Morphological analyses

The morphology of CCIBt3563 was evaluated using an Olympus BX53 microscope (Olympus Optical Co., Tokyo, Japan) equipped with differential interference contrast device (DIC). Microphotographs and measurements were taken using a DP71 digital camera (Olympus Optical Co., Tokyo, Japan) coupled to the optical system and the cellSens image analysis system (Olympus). Important taxonomic features for this genus, such as filament structure, presence or absence of mucilage, shape and dimensions of vegetative cells and heterocyte and presence or absence of aerotopes, were evaluated. Quantitative parameters were taken based on 30 measurements minimum.

High-throughput genome sequencing and *de novo* genome assembly

In order to reduce associate bacteria growing in the unicyanobacterial culture, 50 ml of 30 days old cultured cells were subjected to a serial washing procedure adapted from Heck *et al.* [27]. The modification of Heck *et al.* [27] procedure consisted of increase EDTA to 5 mM and ethanol to 60% in the washing solution and the introduction of a final step with the resulting pellet being washed with a solution of 5 ml of 0.1% Extran and 20 ml of 0.9% NaCl by vacuum filtration through a 8 µm nitrocellulose membrane.

Total genomic DNA was extracted using AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) and a mate-pair library was prepared from 5 to 8 Kbp inserts using the Nextera Mate Pair Library Prep Kit (Illumina), according to the manufacturer's protocols. Sequencing was carried out in the HiSeq 2500 platform using the HiSeq v4 Reagent Kit (Illumina) following the instructions provided by the manufacturer.

The quality of the reads obtained were checked using graphics generated by FastQC 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Adapters from mate-pair library and reads shorter than 30 bp were removed from the datasets using NxTrim v0.4.2 [28] and Cutadapt v1.18 [29], respectively. Libraries were quality filtered using a *phred* score above 25. *De novo* assembly was obtained with SPAdes v3.11.0 [30], using error correction and automatic k-mer estimation size. Subsequently, the generated contigs were improved using SSPACE [31] to merge scaffolds, Pilon [32] for variant detection and GapFiller [33] for filling of gaps. The quality assessment of the assemblies was done with Quast 5.0.2 [34] and the level of completeness and contamination of the sequence was measured with CheckM [35]].

The genome from the studied strain was deposited in the NCBI GenBank database (National Centre for Biotechnology Information; http://www.ncbi.nlm.nih.gov/), under the accession number CP063311.

16S rRNA phylogenetic and 16S-23S ITS secondary structure analysis

The 16 S-23S ITS rRNA sequences was recovered from the assembled genome. This gene and the closest related

sequences from other cyanobacteria were aligned using MUSCLE [36] with default parameters. Phylogenetic trees were inferred using Bayesian (BI), Maximum-Likelihood (ML) and Neighbour-Joining (NJ) methods. *Chroococcidiopsis thermalis* SAG 42.79 was used as the outgroup (GenBank accession number KM020000.1). The ML and NJ trees were reconstructed using MEGA X [37], applying the best fitted model GTR+I+C and Kimura two, respectively. The robustness of the phylogenetic tree was estimated via bootstrap analysis using 1000 replications. The Bayesian inference was conducted by MrBayes v.3.2.1 [38], applying GTR+G+I model, in two runs of four chains Markov Chain Monte Carlo, each one with 5×10^6 generations. Phylogenetic trees were visualized in ITOL 5.3 [39]. Sequence identity matrix was estimated by BioEdit v. 7.2.5 [40].

To provide a taxonomic resolution at the species level, the 16 S-23S ITS region of the studied sequence was used for secondary structure folding. The secondary structures of the D1-D1' and BoxB sub regions were obtained using the Mfold WebServer v.3.6 with the default conditions, except for the application of the structure draw mode with an untangle loop fix [41].

Phylogenomic and genomic similarity analysis

The maximum-likelihood phylogenomic tree of *A. elenkinii* CCIBt3563 was inferred with RAxML v8.0.0 [42] with 1000 bootstraps using the PROTGAMMAIGTR model, assigned as the best by ProtTest 3.4.2 [43]. The protein sequences obtained were based on 120 bacterial single-copy conserved marker proteins selected and aligned with GTDB-Tk v0.3.2 [44].

Species were delineated by in silico genome sequence comparisons using the bioinformatic approaches average nucleotide identities (ANI) and average amino acid identity (AAI) calculated with OrthoANI v1.4 [45] and AAI calculator (http://enve-omics.ce.gatech.edu/aai/), respectively. Genus boundary was delineated using the percentage of conserved proteins (POCP) [46].

RESULTS AND DISCUSSION

Morphological characterization and ecology

The morphology of the CCIBt3563 strain fitted in the description of the species *Anabaenopsis elenkinii* made by Miller [5]. Trichomes solitary, isopolar, deeply curved, commonly circular, usually not completing a full coil, with up to 10 cells, around 30 µm long maximum, constricted, without mucilaginous envelope. Cells cylindrical or ellipsoid with rounded ends, barrel-shaped immediately after division, (1.6-) 2.1–4.0 µm long (3.0 µm in average; *n*=40), (1.8-) 2.0–3.4 µm wide (2.7 µm in average; *n*=40), length/width ratio (0.7-) 0.9–1.8 (1.1 µm in average; *n*=40). Cell content blue-green, granulated, aerotopes present. Heterocytes terminal, single, spherical, 2.3–3.2 µm long (2.7 µm in average; *n*=30), 2.5–2.3 µm wide (2.8 µm in average; *n*=30), length/width ratio 0.9–1.1 (1.0 in average; *n*=30). Akinetes not observed (Fig. 1). Cell dimensions of this strain fall in the lower limit usually described for the species, but the general morphological characteristics observed, as short and curved trichomes, cylindrical to ellipsoid cells, heterocytes mainly spherical and up to eight cells between heterocytes, correspond to many descriptions of *A. elenkinii* in the literature [2, 8–10, 13, 47]. Akinetes were not observed in the samples studied, but their production is extremely dependent of environmental conditions. This result corroborates earlier findings of absence of akinete cells in *A. elenkinii* cultured strains, whose presence was visualized only in natural populations of Nhecolândia soda lakes [13]. These specialized cells are resting cells that preserve genotypes under harsh conditions, contributing to the species' adaptability and survival [48]. It should be mentioned that the strain *A. elenkinii* CCIBt3563 has been kept in culture for more than 8 years, which may have prevented akinete cells differentiation. Although the genetic regulation of akinete formation is completely unknown, an akinete marker protein, AvaK, has been identified [49]. An open reading frame encoding a protein of 69.55% (100% query coverage) amino acid sequence similarity to AvaK of *Aphanizomenon ovalisporum* ILC-164 (KJ725138) was found in the genome of *A. elenkinii* CCIBt3563 (data not shown), indicating its potential to produce akinete cells.

Unfortunately, the ecological background of the *A. elenkinii* type species described by Miller [5] is unknown, however, the majority of *A. elenkinii* strains studied so far have been found in tropical lakes and ponds with alkaline and highly mineral or saline waters. The strain CCIBt3563 was isolated from a population found growing in the 'Salina da Reserva' lake water with pH 10.1, salinity 3.0 g l^{-1} and electric conductivity of 5435 µS·cm^{-1} according to the analyses performed *in situ* [13]. *Anabaenopsis elenkinii* is a common inhabitant of Nhecolândia lakes with pH 9.0–10.4, electrical conductivity 2870–19020 µS·cm^{-1}, salinity 1.4–11 g l^{-1} and temperature 21.7–36.1 °C [13]. Moderate *A. elenkinii* blooms are common in these lakes, however, in dry periods the blooms can be very intense [10, 13, 50]. Cultivation of this species is not trivial and successful *A. elenkinii* isolation was achieved only by adjusting the culture medium to pH 9.4, indicating that pH is an important growth limiting parameter for this species.

16S rRNA gene phylogeny and 16S-23S ITS secondary structure

Four copies of 16S rRNA gene were recovered from the *A. elenkinii* CCIBt3563 assembled genome and the phylogenetic analysis positioned them in a strongly supported (100%) clade of *Anabaenopsis* spp. strains clearly separated from all the other strains of the family Aphanizomenonaceae of the order Nostocales (Fig. 2). This major clade was subdivided into two subclusters (I and II), both supported by Bayesian inference, corroborating a previous study [51]. The *A. elenkinii* CCIBt3563 falls into the subcluster I formed only by strains of *Anabaenopsis elenkinii* and its closest related strain with pairwise identities of 99.73 and 99.79% was the *A. elenkinii* CCIBt3461 (Table S1, available in the online version of this article). This strain was also isolated from the

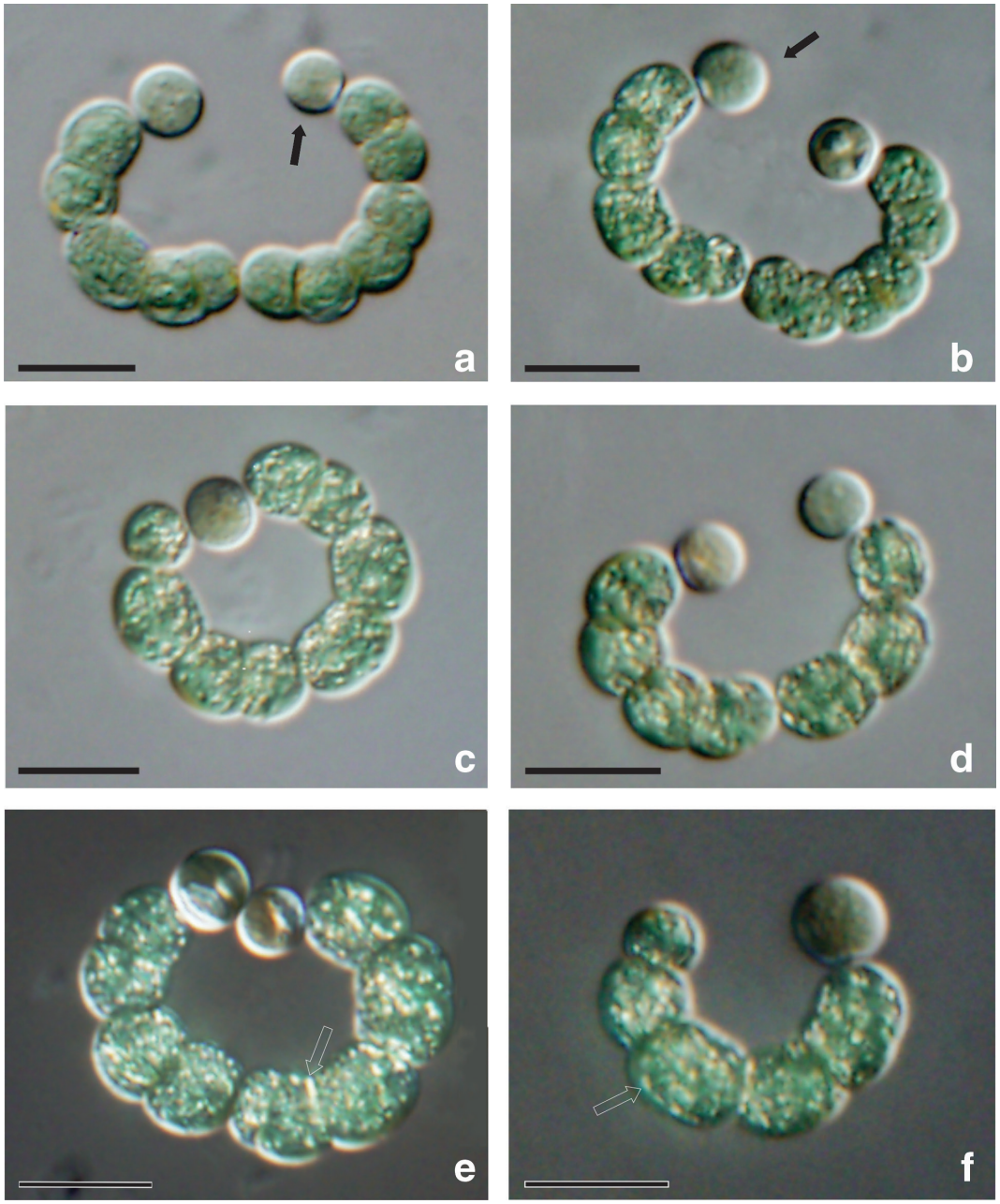

Fig. 1. (a–f) *Anabaenopsis elenkinii* – heterocytes (filled arrows) and aerotopes (empty arrows) are indicated in some figures. Aerotopes appear as small depressions due to Differential Interference Contrast (DIC) technique (e.g. fig. 1b–d) or as brilliant spots inside the cells due to light contrast (e.g. fig. 1e, f). Scale bars represent 5 µm.

same soda lake ('Salina da Reserva'), but from a water sample collected 2 years earlier [13]. The newly sequenced strain CCIBt3563 grouped tightly with other strains isolated from different soda lakes of the same Nhecolândia sub-region (Fig. 2, green mark), i.e. *A. elenkinii* CCIBt1059, *A. elenkinii* CCIBt 3462 and *Anabaenopsis* sp. CENA549 [12, 13], and with *A. elenkinii* AB2006/20 isolated from the soda lake Texcoco in Mexico [8]. This subcluster I aggregated *A. elenkinii* species from Brazilian (Nhecolândia sub-region lakes), Mexican (Texcoco lake) and Kenyan (Sonachi, Elmenteita and Nakuru lakes) soda lakes. Specific information and origin of the *Anabaenopsis* sp. 1A that was also within this subcluster I of *A. elenkinii* are not available (52). The separation of *Anabaenopsis elenkinii* strains inhabiting soda lakes from the other *Anabaenopsis* species is obvious in the phylogenetic tree. Subcluster II with several *Anabaenopsis* species originated from fresh water also encompassed two strains of *Anabaenopsis elenkinii* (NIVA-CYA 494 and NIVA-CYA 501) that were isolated from a freshwater lake in Uganda [8], a distinct environment from the usual habitat of the species,

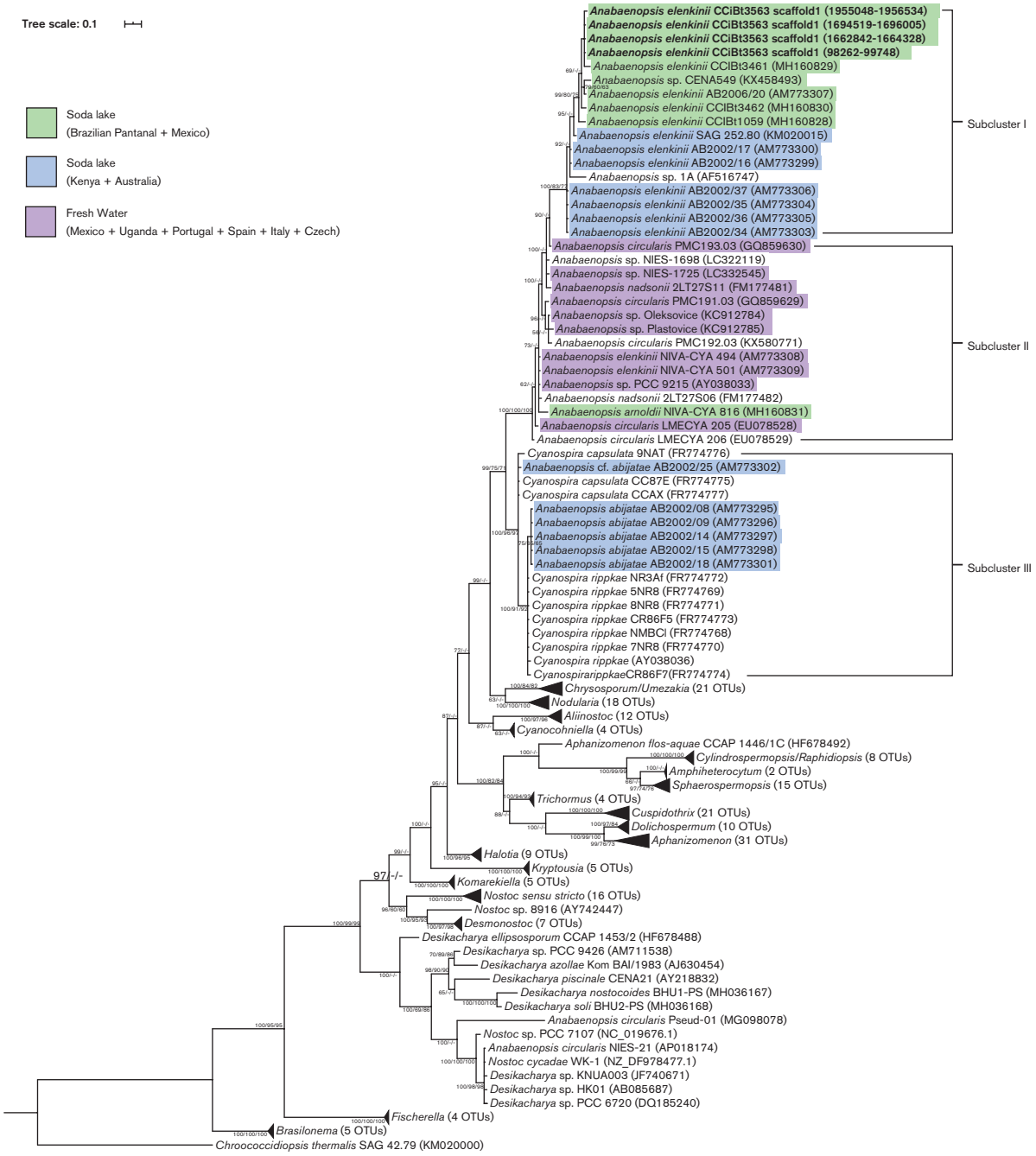

Fig. 2. Phylogenetic positioning the strain *Anabaenopsis elenkinii* CCIBt3563 based on 16S rRNA gene inferred by the Bayesian inference tree. Percentages of scores/bootstrap values above 50% are presented using Bayesian inference, maximum-likelihood and neighbour-joining, respectively. In parentheses are the Anabaenopsis *elenkinii* CCIBt3563 16S rRNA coordinates and for the remained strains are the GenBank accession numbers. Bar: 0.1 changes per nucleotide position.

and deserve further investigation. Only as more populations of *Anabaenopsis* species are collected and confirmed through molecular sequencing it will be possible to solve this intermixed subcluster II. In general, the pattern of sequence clustering observed was consistent with previous phylogenetic trees of *Anabaenopsis* spp. [8, 13, 51]. The exception was the positioning of *Anabaenopsis abijatae* strains in the *Anabaenopsis* spp. clade found by Ballot *et al.* [8]. These strains clustered with well-supported bootstrap values (100% Bayesian posterior probability, 96% ML and 97% NJ) within the *Cyanospira* subcluster III in our study and a previous one [51]. *Cyanospira* is an *Anabaenopsis* sister taxa according to 16S rRNA gene phylogenetic analysis and its members are also found inhabiting hyperhalkaline environments [51].

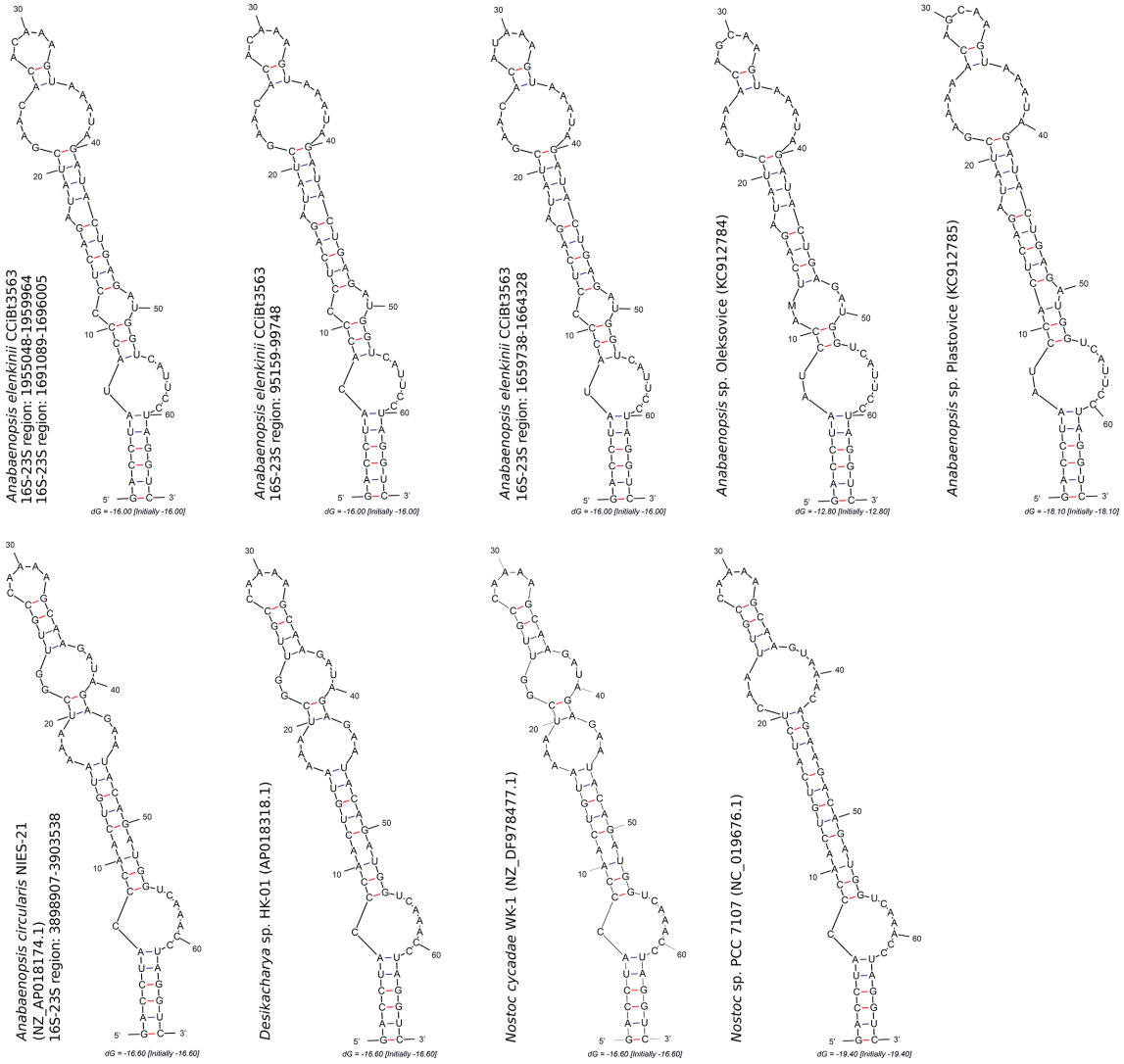

Fig. 3. Folded secondary structures of the D1–D1' helix regions from the 16 S-23S intergenic spacer of the members of the *Anabaenopsis* and *Desikacharya/Nostoc* clade. In parentheses are the *Anabaenopsis elenkinii* CCIBt3563 16S rRNA coordinates and for the remained strains are the GenBank accession numbers.

The 16S rRNA gene sequence of the *A. circularis* NIES-21 (accession numbers AF247595 and AP018174), the only *Anabaenopsis* strain with genome sequenced, fitted within the *Desikacharya* clade with strong support (100%) (Fig. 2). *Desikacharya* is a genus recently created and it was erected with members morphologically similar to the traditional and polyphyletic genus *Nostoc* from the family Nostocaceae [53, 54]. These authors also recommended that several existing members of *Nostoc*-like shall be reclassified into this proposed new genus. The 16S rRNA gene sequence of *A. circularis* NIES-21 showed 100% of identity with the sequence of *Desikacharya/Nostoc cycadae* WK-1 isolated from cyanobacterial colonies growing in the coralloid roots of the gymnosperm *Cycas revoluta* [55], and of 99.93% with *Desikacharya* sp. HK-01, a terrestrial cyanobacterium isolated from soil in Himeji, Hyogo, Japan [56] (Table S1).

Currently, only four 16 S-23S ITS sequences of different *Anabaenopsis* strains are available in public databases. Two of the four 16 S-23S ITS copies of *A. elenkinii* CCIBt3563 include the Isoleucine and Alanine tRNA genes (16 S-23S genomic regions 1691089–1696005 and 1955048–1959964). The D1-D1' region of the four 16 S-23S ITS copies showed three variants (Fig. 3). The differences occur due to changes of nucleotides in positions 7 and 29, however, these differences do not alter the secondary folded structures of the regions. The D1-D1' region of *A. elenkinii* CCIBt3563, *Anabaenopsis* sp. Plastovice (KC912785) and *Anabaenopsis* sp. Oleksovice (KC912784) showed similar secondary structures, clearly different from the structure of the *A. circularis* NIES-21 (NZ_AP018174.1). Secondary structure of the D1-D1' helix of this strain had markedly differences observed in the formation of the second and the third loop, but it was identical to the

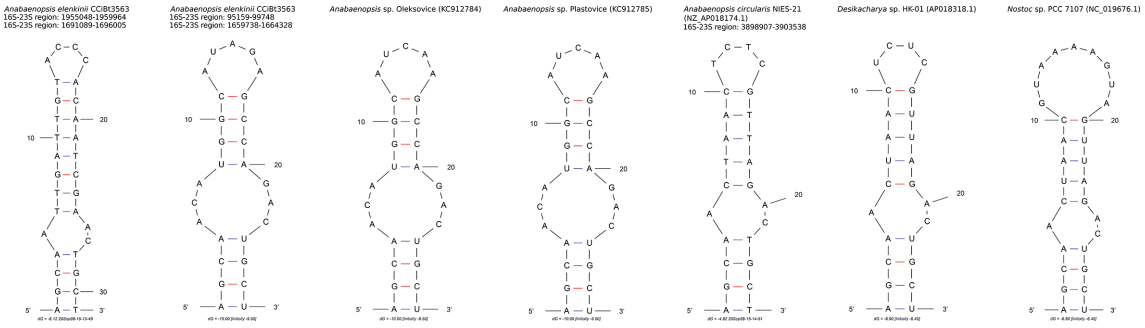

Fig. 4. Folded secondary structures of the BoxB helix regions from the 16S-23S intergenic spacer of the members of the *Anabaenopsis* and *Desikacharya/Nostoc* clade. In parentheses are the *Anabaenopsis elenkinii* CCIBt3563 16S rRNA coordinates and for the remained strains are the GenBank accession numbers.

strains *Desikacharya/Nostoc cycadae* WK-1 and *Desikacharya* sp. HK-01, sharing also similarity with *Nostoc* sp. PCC 7107, which showed the same structure observed in the first and fourth loop of the other *Nostoc*-like strains.

The Box-B regions of *A. elenkinii* CCIBt3563 showed two variants with distinct structures (Fig. 4). One variant showed the same structure of the Box-B regions of *Anabaenopsis* sp. Plastovice and *Anabaenopsis* sp. Oleksovice, with the only difference in the composition of the nucleotides in the positions 14 and 15, while the latter two *Anabaenopsis* showed identical sequences. *Anabaenopsis circularis* NIES-21 showed a divergent structure compared to other species of the genus, while it is identical to the *Desikacharya* sp. HK-01 and divergent from the structure of *Nostoc* sp. PCC 7107. The indistinguishable D1-D1' and Box-B regions among strains of the genus *Desikacharya* provide strong evidence that the *A. circularis* NIES-21 demands a revision of its generic identity.

Phylogenomic and genomic similarity analyses

The complete genome of the *A. elenkinii* CCIBt3563 has a total size of 4495068 bp assembled in one scaffold, with a 41.98% G+C content and a coverage of 59.89 times. Gene annotation revealed 3932 CDS with four rRNA genes and 45 tRNA genes. The evaluation of completeness and contamination level of *A. elenkinii* CCIBt3563 genome according to the CheckM analysis highlight its quality (completeness – 98.55% and contamination level – 1.57%). The main characteristics of *A. elenkinii* CCIBt3563 genome were compared to the *A. circularis* NIES-21 and others closely related cyanobacterial genomes available (Table S2).

The phylogenomic analysis based on the concatenated alignment of 120 orthologous proteins showed that the strain *A. elenkinii* CCIBt3563 groups together in a clade with strains of *Nodularia* and *Chrysosporum* (Fig. 5). These two genera also belong to the family Aphanizomenonaceae of the order Nostocales and also appeared as closely related genera in the 16S rRNA gene phylogenetic tree (Fig. 2). Unfortunately, genome sequence is not available yet for *Cyanospira*, the sister genus of *Anabaenopsis* as shown by the 16S rRNA gene sequence phylogenetic analysis. As also observed in the 16S rRNA gene phylogenetic tree, *A. elenkinii* CCIBt3563 was positioned in a major cluster containing only genera belonging to the family Aphanizomenonaceae (Fig. 5, green mark).

By contrast, the *A. circularis* NIES-21 clustered at a completely different node within the clade containing strains of the genus *Desikacharya*, with strong support (100%), confirming the result achieved by 16S rRNA gene phylogenetic analysis. Furthermore, the *A. circularis* NIES-21 closest related genome was *Desikacharya* sp. HK-01 (NIES-2109). Previous study also showed the *A. circularis* NIES-21 in a phylogenomic clade containing *Nostoc*-like strains [25]. These authors reported that this clade *Nostoc* I was formed by 'free-living aquatic *Nostoc* strains grouped together with *Anabaena* strains'. However, soil strains also were fitted in this clade as showed in our study.

The strain NIES-21 was identified as *A. circularis*, therefore, whole genome comparison analyses (ANI and AAI) were applied in order to assess species boundaries between its genome and the *A. elenkinii* CCIBt3563 genome. ANI and AAI represents the average nucleotide and amino acid identity, respectively, of all orthologous genes shared between any two genomes and offers robust resolution between strains of the same or closely related species [57–59]. The species delineation based on the ANI and AAI was unsuccessful for *A. elenkinii* CCIBt3563 and *A. circularis* NIES-21 genomes with an average similarity of 76.04 and 73.66%, respectively. However, *A. circularis* NIES-21 genome showed ANI and AAI values above 96% threshold [57] with *Desikacharya/Nostoc cycadae* WK-1 and *Desikacharya* sp. HK-01 genomes, indicating that these genomes belong to the same species [58, 59] (Fig. S1a, b). POCP analyses between these strains showed values above 76% (Fig. S1c), surpassing 50% that is considered a boundary for grouping prokaryotes of the same genus [46]. The POCP between *A. elenkinii* CCIBt3563 and *A. circularis* NIES-21 genomes resulted in 58.92%, a value also above the genus threshold. This incongruence reveals the difficulties in using a single stationary threshold for all prokaryotes and may be solved

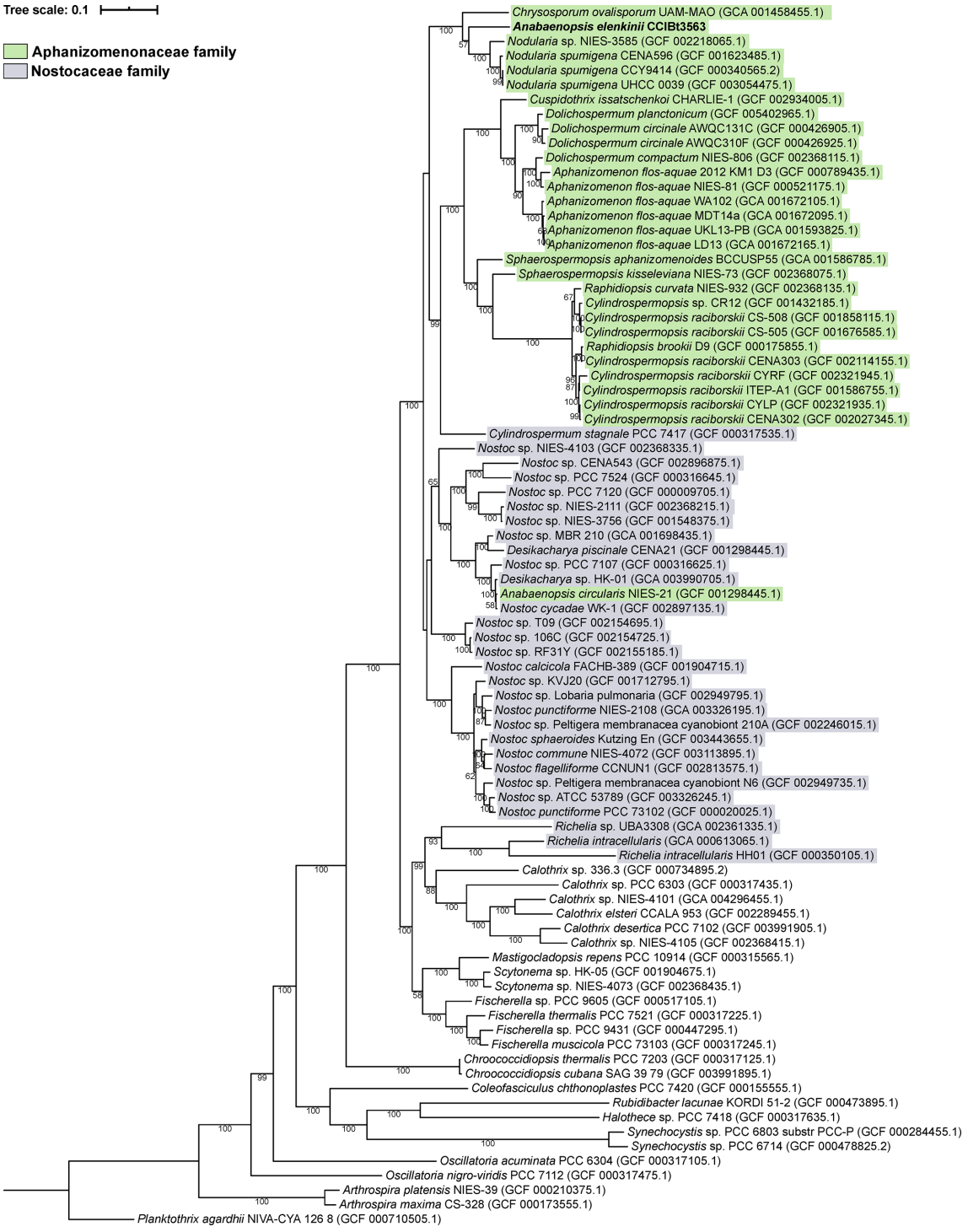

Fig. 5. Maximum-Likelihood phylogenomic tree based on 120 single-copy conserved proteins in cyanobacterial genomes. GenBank accession numbers are in parentheses. Bar: 0.1 substitutions per position.

by defining different thresholds in accordance to well established taxa [60].

In this study a typical *Anabaenopsis elenkinii* strain was fully described on the basis of morphology, ecology and phylogeny of single gene 16S rRNA as well as of the complete genome. Furthermore, the current phylogenetic position of *A. circularis* NIES-21 indicates that it actually is a member of the genus *Desikacharya*. The strain NIES-21 is the only supposed *Anabaenopsis* with genome available in the NCBI, therefore, the clarification of its taxonomy is

especially important, considering that the only reference genome for the genus so far has led to misleading analyses. For example, the classification based on the GTDB (release version 95) method [61] classifies the genomes of the strains *Desikacharya* sp. HK01 (GCA_003990705.1), *Nostoc* sp. PCC 7107 (GCF_000316625.1) and *Nostoc cycadae* WK-1 (GCF_002897135.1) as *Anabaenopsis*, due to their close proximity to NIES-21 (GCF_002367975.1). This issue was solved in this study, demonstrating that researches describing type strains must be continuously developed in order to improve the taxonomy of Cyanobacteria. Moreover, this investigation provides important information regarding the genus *Anabaenopsis*, nevertheless, studies must be carried out to understand its intraspecific diversity and evolution and to validate generic distinctions based on genomic relatedness within the order Nostocales.

Funding information
This research was supported by the São Paulo Research Foundation (FAPESP #2016/14227-5) and the National Council for Scientific and Technological Development (CNPq, 433166/2018-5).

Acknowledgements
E. D., is thankful to the Brazilian Federal Agency for the Support and Evaluation of Graduate Education (CAPES) for providing graduate scholarship (Finance code 001). T. A. P., was supported by graduate fellowship from FAPESP (#2017/12644-0). MFF and LHZB received a research fellowships (306803/2018-6 and 306851/2017-2, respectively) from the National Council for Scientific and Technological Development (CNPq). We gratefully acknowledge Dr Célia L. Sant'Anna for generously providing the strain *Anabaenopsis elenkinii* CCIBt3563.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Komárek J, Kastovsky J, Mares J, Johansen JR. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia* 2014;86:295–335.
2. Komárek J. Phenotypic diversity of the heterocytous cyanoprokaryotic genus *Anabaenopsis*. *Czech Phycol Olomouc* 2005;5:1–35.
3. West GS. Report of the freshwater algae INCL. phytoplankton of the third Tanganyika expedition. *J Linn Soc Bot* 1907;38:81–197.
4. Wołoszyńska J. Das Phytoplankton einiger javanischer Seen, mit Berücksichtigung des Sawa-Planktons. *Bull Acad Sci Cracovie, mat nat. ser* 1912:649–709.
5. Miller VV. K sistematike roda *Anabaena* Bory [Zur Systematik der Gattung *Anabaena* Bory]. *Arch russk Protistol Obšč* 1923;2:116–126.
6. Hindák F. Planktic species of two related genera *Cylindrospermopsis* and *Anabaenopsis* from Western Slovakia. *Algol Stud* 1988;50-53:283–302.
7. Komárek J. Cyanoprokaryota -3. Teil/ 3rd Part: Heterocytous genera. In: Büdel B, Gärtner G, Krienitz L, Schagerl M (editors). *Süsswasserflora von Mitteleuropa 19/3. Elsevier/Spektrum*. Heidelberg; 2013. pp. 1–1130.
8. Ballot A, Dadheech PK, Haande S, Krienitz L. Morphological and phylogenetic analysis of *Anabaenopsis abijatae* and *Anabaenopsis elenkinii* (Nostocales, cyanobacteria) from tropical inland water bodies. *Microb Ecol* 2008;55:608–618.
9. Jeeji-Bai N, Hegewald E, Soeder CJ. Taxonomic studies on the genus *Anabaenopsis*. In: Desikachary TV (editor). *Taxonomy of Algae*. Chennai: University of Madras; 1980. pp. 115–142.
10. Santos KRS, Sant'Anna CL. Cyanobacteria from different types of lakes ("salina", "salitrada" and "baía") representative of the Pantanal da Nhecolândia, MS, Brazil. *Rev Bras Bot* 2010;33:61–83.
11. Aguilera A, Komárek JIŘÍ, Echenique RO. *Anabaenopsis* morphospecies (cyanobacteria, Nostocales) from Los Patos shallow lake (Province of Buenos Aires, Argentina). *Phytotaxa* 2016;272:173–183.
12. Genuário DB, Andreote APD, Vaz MGMV, Fiore MF. Heterocyte-forming cyanobacteria from Brazilian saline-alkaline lakes. *Mol Phylogenet Evol* 2017;109:105–112.
13. Santos KRDES, Hentschke GS, Andreote ANAPD, Laughinghouse IV HD, Ballot A *et al*. Polyphasic characterization of newly isolated *Anabaenopsis* (Cyanobacteria) strains from tropical Brazil and Mexico. *Phytotaxa* 2018;367:1–12.
14. Andreote APD, Dini-Andreote F, Rigonato J, Machineski GS, Souza BCE *et al*. Contrasting the genetic patterns of microbial communities in soda lakes with and without cyanobacterial bloom. *Front Microbiol* 2018;9:244.
15. Đorđević NB, Simić SB. Cyanobacterial blooms in oligosaline and alkaline microaccumulation before and after rehabilitation. *Pol J Environ Stud* 2014;23:1975–1982.
16. Krienitz L, Dadheech PK, Kotut K. Mass developments of the cyanobacteria *Anabaenopsis* and *Cyanospira* (Nostocales) in the soda lakes of Kenya: ecological and systematic implications. *Hydrobiologia* 2013;703:79–93.
17. Iteman I, Rippka R, Tandeau de Marsac N, Herdman M. rDNA analyses of planktonic heterocystous cyanobacteria, including members of the genera *Anabaenopsis* and *Cyanospira*. *Microbiology* 2002;148:481–496.
18. Komárek J. Modern taxonomic revision of planktic nostocacean cyanobacteria: a short review of genera. *Hydrobiologia* 2010;639:231–243.
19. Komárek J, Mareš J. An update to modern taxonomy (2011) of freshwater planktic heterocytous cyanobacteria. *Hydrobiologia* 2012;698:327–351.
20. Guiry MD, Guiry GM. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. http://www.algaebase.org; searched on 07 August 2020. http://www.algaebase.org/search/genus/detail/?genus_id=43591.
21. Hauer T, Komárek J. CyanoDB 2.0 - On-line database of cyanobacterial genera. - World-wide electronic publication 2020; Univ. of South Bohemia & Inst. of Botany AS CR, http://www.cyanodb.cz.
22. Willis A, Woodhouse JN. Defining cyanobacterial species: diversity and description through genomics. *CRC Crit Rev Plant Sci* 2020;39:101–124.
23. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P *et al*. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007;57:81–91.
24. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:1–8.
25. Gagunashvili AN, Andrésson Ólafur S. Distinctive characters of *Nostoc* genomes in cyanolichens. *BMC Genomics* 2018;19:434.
26. Kotai J. *Instructions for Preparation of Modified Nutrient Solution Z8 for Algae. Publication B-11/69*. Oslo: Norwegian Institute for Water Research; 1972.
27. Heck K, Machineski GS, Alvarenga DO, Vaz MGMV, Varani AdeM *et al*. Evaluating methods for purifying cyanobacterial cultures by qPCR and high-throughput illumina sequencing. *J Microbiol Meth* 2016;129:55–60.
28. O'Connell J, Schulz-Trieglaff O, Carlson E, Hims MM, Gormley NA *et al*. NxTrim: optimized trimming of illumina mate pair reads. *Bioinformatics* 2015;31:2035–2037.
29. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J* 2011;17:10–12.
30. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
31. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 2011;27:578–579.

32. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A *et al*. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;9:e112963.
33. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. *Genome Biol* 2012;13:R56.
34. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
35. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;25:1043–1055.
36. Edgar RC. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–1797.
37. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018;35:1547–1549.
38. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A *et al*. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 2012;61:539–542.
39. Letunic I, Bork P. Interactive tree of life (iTOL) V4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.
40. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Window 95/98/NT. *Nucleic Acids Symposium Series* 1999;41:95–98.
41. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003;31:3406–3415.
42. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
43. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 2011;27:1164–1165.
44. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 2019;36:1925–1927.
45. Lee I, Ouk Kim Y, Park S-C, Chun J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 2016;66:1100–1103.
46. Qin Q-L, Xie B-B, Zhang X-Y, Chen X-L, Zhou B-C *et al*. A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol* 2014;196:2210–2215.
47. Jeeji-Bai N, Hegewald E, Soeder CJ. Revision and taxonomic analysis of genus *Anabaenopsis*. *Algol Stud* 1977;1977:25–32.
48. Kaplan-levy RN, Hadas O, Summers ML, Sukenik A. Dormancy and resistance in harsh environments. *Topics Curr Genet* 2010;21:189–202.
49. Zhou R, Wolk CP. Identification of an akinete marker gene in *Anabaena variabilis*. *J Bacteriol* 2002;184:2529–2532.
50. de Souza Santos KR, Rios Jacinavicius F, Leite Sant'Anna C, Santos KRS, Sant'Anna CL. Effects of the pH on growth and morphology of *Anabaenopsis elenkinii* Miller (Cyanobacteria) isolated from the alkaline shallow lake of the Brazilian Pantanal. *Fottea* 2011;11:119–126.
51. Sili C, Mascalchi C, Ventura S. Evolutionary differentiation of the sister cyanobacterial genera *Cyanospira* Florenzano, Sili, Pelosi et Vincenzini and *Anabaenopsis* (Woloszynska) Miller in response to extreme life conditions. *Fottea* 2011;11:107–117.
52. Neilan BA, Saker ML, Fastner J, Törökné A, Burns BP. Phylogeography of the invasive cyanobacterium Cylindrospermopsis raciborskii. *Mol Ecol* 2003;12:133–140.
53. Kabirnataj S, Nematzadeh GA, Talebi AF, Saraf A, Suradkar A *et al*. Description of novel species of *Aliinostoc*, *Desikacharya* and *Desmonostoc* using a polyphasic approach. *Int J Syst Evol Microbiol* 2020;70:3413–3426.
54. Saraf AG, Dawda HG, Singh P, gen D. Desikacharya gen. nov., a phylogenetically distinct genus of cyanobacteria along with the description of two new species, *Desikacharya nostocoides* sp. nov. and *Desikacharya soli* sp. nov., and reclassification of *Nostoc thermotolerans* to *Desikacharya thermotolerans* comb. nov. *Int J Syst Evol Microbiol* 2019;69:307–315.
55. Kanesaki Y, Hirose M, Hirose Y, Fujisawa T, Nakamura Y *et al*. Draft genome sequence of the nitrogen-fixing and hormogonia-inducing Cyanobacterium *Nostoc cycadae* strain WK-1, isolated from the coralloid roots of *Cycas revoluta*. *Genome Announc* 2018;6:e00021–18.
56. Katoh H, Shiga Y, Nakahira Y, Ohmori M. Isolation and characterization of a drought-tolerant cyanobacterium, *Nostoc* sp. HK-01. *Microb Environ* 2003;18:82–88.
57. Ciufo S, Kannan S, Sharma S, Badretdin A, Clark K *et al*. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int J Syst Evol Microbiol* 2018;68:2386–2392.
58. Konstantinidis KT, Tiedje JM. Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* 2005;187:6258–6264.
59. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 2009;106:19126–19131.
60. Barco RA, Garrity GM, Scott JJ, Amend JP, Nealson KH *et al*. A genus definition for bacteria and *Archaea* based on a standard genome relatedness index. *mBio* 2020;11:e02475–19.
61. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A *et al*. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;36:996–.