Accepted Manuscript

Polyphasic identification of cyanobacterial isolates from Australia

Elvina Lee, Una M. Ryan, Paul Monis, Glenn B. McGregor, Andrew Bath, Cameron Gordon, Andrea Paparini

Please cite this article as: Lee, E., Ryan, U.M, Monis, P., McGregor, G.B, Bath, A., Gordon, C., Paparini, A., Polyphasic identification of cyanobacterial isolates from Australia, *Water Research* (2014), doi: 10.1016/j.watres.2014.04.023.
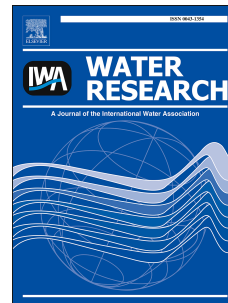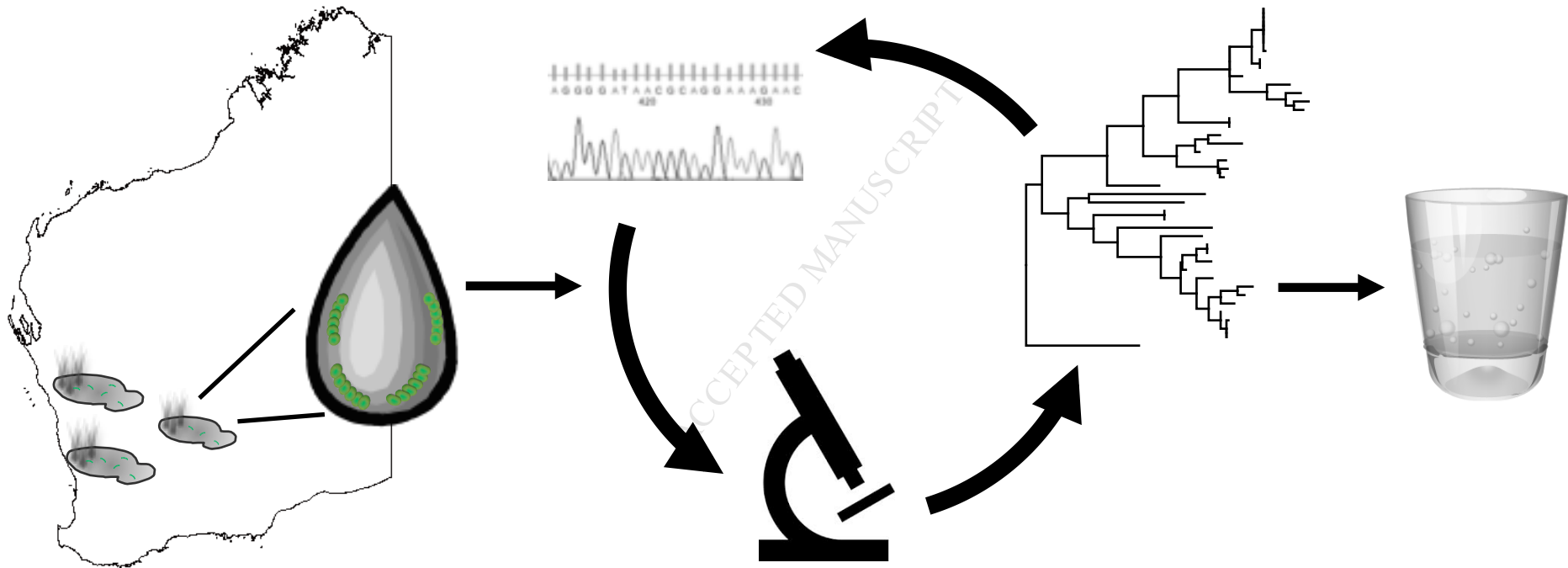
1   **Polyphasic identification of cyanobacterial isolates from**

2   **Australia**

3

4   Elvina Lee[a], Una M Ryan[a], Paul Monis[b], Glenn B McGregor[c],

5   Andrew Bath[d], Cameron Gordon[d], Andrea Paparini[a,*]

6

7   [a] School of Veterinary and Life Sciences, , Murdoch

8   University, Murdoch, Western Australia, 6150

9   [b] Australian Water Quality Centre, South Australian Water

10   Corporation, 250 Victoria Square, Adelaide, 5000

11   [c] Department of Science, Information Technology, Innovation

12   and the Arts, GPO Box 5078, Brisbane Queensland 4001

13   [d] Drinking Water Quality Branch, Water Corporation, 629

14   Newcastle Street, Leederville, Western Australia 6007

15   Corresponding author. Mailing address: School of Veterinary

16   and Life Sciences, Murdoch University, 90 South Street,

17   Murdoch, Western Australia 6150, Australia. Tel.: +61 89360

18   6650; Fax: +61 89360 6628. E–mail address:

19   a.paparini@murdoch.edu.au

20

Abstract

Reliable identification of cyanobacterial isolates has significant

socio-economic implications as many bloom-forming species

affect the aesthetics and safety of drinking water, through the

production of taste and odour compounds or toxic metabolites.

The limitations of morphological identification have promoted

the application of molecular tools, and encouraged the adoption

of combined (polyphasic) approaches that include both

microscopy– and DNA–based analyses. In this context, the

rapid expansion of available sequence data is expected to allow

increasingly reliable identification of cyanobacteria, and

ultimately resolve current discrepancies between the two

approaches.

In the present study morphological and molecular

characterisations of cyanobacterial isolates (n=39), collected

from various freshwater sites in Australia, were compared.

Sequences were obtained for the small ribosomal subunit RNA

gene (16S rDNA) (n=36), the DNA-dependent RNA

polymerase gene (*rpo*C1) (n=22), and the phycocyanin operon,

with its intergenic spacer region (*cpc*BA-IGS) (n= 19).

Phylogenetic analyses identified three cyanobacterial orders:

the Chroococcales (n=8), Oscillatoriales (n=6), and Nostocales

(n=25). Interestingly, multiple novel genotypes were identified,

with 22% of the strains (17/77) having <95% similarity to

available sequences in GenBank.

47   Morphological and molecular data were in agreement at the

48   species level for only 26% of the isolates obtained (10/39),

49   while agreement at the genus level was obtained for 31%

50   (12/39). Confident identification of the remaining 44% of the

51   strains (17/39) beyond the order level was not possible. The

52   present study demonstrates that, despite the taxonomic

53   revisions, and advances in molecular–, and bioinformatics–

54   tools, the lack of reliable morphological features, culture-

55   induced pleomorphism, and proportion of misidentified or

56   poorly described sequences in GenBank, still represent

57   significant factors, impeding the confident identification of

58   cyanobacteria species.

59

63 **1. Introduction**

64

65 Cyanobacteria are a group of ubiquitous photosynthetic

66 prokaryotes, found in all types of aquatic environments

67 (Whitton and Potts, 2000). Interest in this phylum has increased

68 due to (i) their ancient evolutionary origins (Tomitani et al.,

69 2006), (ii) their ecological role as oxygen producers, and

70 atmospheric nitrogen– and carbon–fixers (Reynolds, 2006),

71 (iii) the socio-economic impact on various industries (e.g.

72 water, tourism and food) of bloom–forming producers of toxins

73 and/or odorous metabolites (Granéli and Turner, 2006), and

74 (iv) their application as a source of biofuels and

75 pharmaceuticals (Borowitzka, 1995; Li et al., 2008).

76 Cyanobacteria identification, enumeration and classification

77 have traditionally been based on light-microscopy observations,

78 using morphological characteristics such as cell size, cell

79 fission type, trichome width, shape of the terminal cells, shape,

80 size and position of specialised cells such as akinetes and

81 heterocytes, presence of aerotopes etc. (Castenholz, 2001).

82 However, this approach requires considerable operator skill and

83 time; with distinctive phenotypic characteristics varying

84 significantly within species, or even being lost, due to

85 environmental or culture conditions, growth phase, use of

86 fixatives etc. (Lyra et al., 2001; Whitton and Potts, 2000).

87 Furthermore, manifestation of ecotypes, or microbial

88 pleomorphism during long-term cultivation, has resulted in a

4

89   large number of strains being misidentified, with disagreeing

90   nomenclature and morphological descriptions (Komárek, 2006;

91   Komárek, 2010).

92   The well-known limitations to morphology-based identification

93   promoted the development of DNA-based approaches, as a

94   means of reliably identifying cyanobacterial isolates (Valério et

95   al., 2009; Willame et al., 2006). The small ribosomal subunit

96   RNA gene (16S rDNA), together with its internal transcribed

97   spacer (ITS) region, have been widely used for taxonomic

98   purposes, to profile complex prokaryotic communities, and

99   infer phylogenetic relationships (Castenholz, 2001; Coenye and

100   Vandamme, 2003; Komárek, 2006). Other commonly used loci

101   include the more discriminatory protein-coding gamma subunit

102   of the DNA-dependent RNA polymerase (*rpo*C1) (Fergusson

103   and Saint, 2000), and the phycocyanin operon, consisting of the

104   two *cpc*B-*cpc*A genes with their variable intergenic region (PC-

105   IGS) (Neilan et al., 1995).

106   Previous studies have shown that identification of unknown

107   isolates can be hampered by incomplete (or unreliable)

108   morphological descriptions being provided for the sequenced

109   strains, incorrect identification of strains in culture collections,

110   and/or simply, by the lack of proper reference strains *tout court*

111   (Komárek, 2006; Komárek, 2010; Rajaniemi et al., 2005).

112   Much has been done to overcome these problems, such as the

113   proposal of the International Code of Nomenclature for algae,

5

114   fungi, and plants (Castenholz and Norris, 2005; McNeill et al.,

115   2012; Oren, 2004; 2011; Oren and Tindall, 2005), and the

116   International Code of Nomenclature of Bacteria which groups

117   cyanobacteria into subsections (Castenholz, 2001), and the

118   resulting revisions to cyanobacteria nomenclature and

119   classification (Anagnostidis, 2001; Otsuka et al., 2001;

120   Rajaniemi et al., 2005; Suda et al., 2002). However, this has

121   also resulted in new taxa being described using a combination

122   of both codes, causing further confusion (Komárek, 2010).

123   With time, the rapid expansion of available sequence data (Fig.

124   1) is expected to allow increasingly accurate molecular

125   identification of cyanobacteria, and help in resolving the

126   discrepancies between the microscopy and molecular

127   approaches. In light of the current maturity of sequence

128   databases, taxonomy and molecular tools (Komárek et al.,

129   2011; Siegesmund et al., 2008; Strunecky et al., 2011; Wacklin

130   et al., 2009), the aim of this study was to determine how well

131   morphological and molecular tools for the identification of

132   cyanobacteria isolates corroborate. To this end, cyanobacteria

133   were isolated from randomly selected locations in Western

134   Australia, from which there have been few cyanobacteria

135   studies, for use in this study.

136   **2. Materials and Methods**

137   **2.1 Isolation and cultivation of cyanobacterial strains**

138   Freshwater samples (n=50; 1 L each) were collected between

139   November 2010 and June 2011, from various randomly

140   selected locations in Western Australia, including: (i) protected

141   freshwater reservoirs in the Great Southern region, with

142   restricted access and excellent water quality (GS samples;

143   n=6); (ii) urban lentic systems, in the Perth metropolitan area,

144   with free public access (n=3); (iii) shallow, rural, lotic systems,

145   which included a drain and a river (n=2) (Table 1). For sample

146   enrichment, water was pre-filtered through qualitative filter

147   paper Number 1 (Advantec, Japan), and sterile 0.2 µm and 0.45

148   µm mixed cellulose ester (MCE) filter membranes (Advantec,

149   Japan). Each filter was divided into thirds and placed into

150   ASM-1 medium (Gorham et al., 1964) (pH adjusted to 7.6), or

151   modified ASM-1 medium (with no, or 200 µM sodium nitrate),

152   and were allowed to grow for a fortnight. Strains were isolated

153   using traditional *ad hoc* methods such as: i) serial dilution to

154   extinction, ii) micromanipulation (Narishige, Japan), iii)

155   sequential centrifugation, iv) differential filtration, and v) agar

156   plate streaking. After repeated passages, isolates were finally

157   transferred into 75 cm^3 culture flasks (Greiner Bio-One,

158   Germany) for long-term cultivation, when a single morphotype

159   was stably observed.

7

160   In addition to the Western Australian isolates, strains from the

161   culture collection at the Australian Water Quality Centre

162   (AWQC), which were mostly collected in the early-mid 1990s

163   during surveys of Australian freshwater sources, were also

164   included in the study. The surveys included: protected and open

165   freshwater reservoirs from New South Wales, Victoria and

166   South Australia (n=6), open rural lentic systems from New

167   South Wales (n=2), and open rural lotic systems from New

168   South Wales and Victoria (n=2) (Table 1). AWQC isolates

169   were grown in 50 cm^3 tissue culture flasks with vented lids

170   (Greiner Bio-one, Germany), containing 20–30 mL of ASM-1

171   medium (pH adjusted to 7.6). Cultures were incubated at 20°C

172   under a photon irradiance of 20 μMm^{-1}s^{-1} provided by cool

173   white light (16 h/8 h light/dark cycle).

174   **2.2 Microscopic identification and molecular analyses**

175   After obtaining a stable homogenous phenotype and steady

176   growth, each isolate was subcultured into 50 cm^3 tissue culture

177   flasks, containing, approximately, the same number of

178   cells/filaments, and the same medium. Microscopic

179   identification of these sister cultures was performed, by two

180   independent laboratories. Observations were performed by light

181   microscopy at various magnifications on a large number of live

182   cells (i.e., unfixed), either directly in the culture flask, or by

183   making multiple fresh mounts of the cultures.

184 For molecular analyses, aliquots of cyanobacterial biomass (25

185 mL of cell suspension, or ~5 g of microbial mats) were

186 transferred into sterile 50 mL polypropylene tubes (Greiner

187 Bio-One, Germany). The tubes were centrifuged at maximum

188 speed (4,050 rcf, 30 *min*) using an Allegra X-15R (Beckman

189 Coulter, USA), and DNA was extracted from the pellet, using

190 commercially available DNA extraction kits (Promega, USA or

191 Qiagen, USA) according to the manufacturers' protocols,

192 except that the pellet was digested with proteinase-K overnight.

193 All PCR reactions were run on a G-Storm GS1 standard block

194 thermal cycler (Kapa Biosystems, USA). A partial fragment

195 (313 bp) of the 16S rDNA hypervariable region was amplified

196 using the cyanobacterial-specific PCR protocol previously

197 described (McGregor and Rasmussen, 2008), except that the

198 reverse primers 781R(a) and 781(b) (Nübel et al., 1997) were

199 used instead. Amplification of *rpo*C1 (409bp) was performed

200 using cyanobacteria-specific primers rpoC1-1 and rpoC1-T

201 (Palenik and Haselkorn, 1992). A partial fragment (423bp) of

202 the phycocyanin intergenic spacer region (*cpc*BA-IGS) and

203 corresponding flanking regions *cpc*B and *cpc*A was amplified

204 using the primers cpcBF (UPF) and cpcAR (URP) as described

205 by Robertson et al. (2001). Amplicons were visualised by 1%

206 agarose gel electrophoresis containing SYBR Safe Gel Stain

207 (Invitrogen, USA) in a dark reader trans-illuminator (Clare

208 Chemical Research, USA). Band products corresponding to the

9

209   expected length were excised from the gel, using a new scalpel

210   blade for each sample. These products were then purified using

211   a MO BIO UltraClean DNA purification kit (MOBIO

212   Laboratories, USA), and sequenced, bi-directionally, using an

213   ABI Prism Terminator Cycle Sequencing kit (Applied

214   Biosystems, USA), on an Applied Biosystem 3730 DNA

215   Analyzer. Sequencing chromatograms were then analysed by

216   FinchTV 1.4

217   (http://www.geospiza.com/Products/finchtv.shtml), and

218   imported into Bioedit Sequence Alignment Editor (Hall, 1999)

219   to generate one consensus sequence (per locus) for each isolate.

220

221   **2.3 Phylogenetic analysis**

222   Phylogenetic analyses were conducted on the sequences

223   obtained during the present study (GenBank accession numbers

224   JQ811771 to JQ811820) and retrieved from GenBank, using

225   the Basic Local Alignment Search Tool (BLAST). MEGA 5

226   (Tamura et al., 2011) was used for sequence manipulations,

227   alignments by CLUSTAL W (Larkin et al., 2007) and the *p*-

228   distance model (Kimura, 1980) was used for the calculation of

229   the pair-wise evolutionary distances. Phylogenetic analyses of

230   aligned sequences were conducted using distance (Neighbour-

231   Joining), maximum parsimony (MP) and maximum likelihood

232   (ML) methods, in MEGA 5 (Kimura, 1980); tree reliability was

233   evaluated with bootstrap analysis of 500 replicates. For the

10

234   purpose of molecular identification, a percentage molecular

235   similarity cutoff of 98% and 95% for species and genus

236   identification respectively was used for the 16S rDNA

237   (Stackebrandt and Goebel, 1994). As a result of the much

238   smaller number of sequences available (Fig. 1), 95% and 90%

239   similarity values respectively were used for species and genus

240   identification for the *rpo*C1 and *cpc*BA-IGS loci. These values

241   were chosen based on the genetic distance for species and

242   genus measured from the dataset used to generate the trees

243   (Fig. 4 and 5).

244   **3. Results**

245   From the 11 Western Australia sampling sites, 29 isolates were

246   obtained; of these, 12 were obtained from the protected

247   reservoirs, five from urban lentic systems, and 12 from rural

248   lotic waters. The Western Australian isolates were studied

249   together with the 10 cyanobacteria isolates from the AWQC

250   (Table 1).

251

252   **3.1 Comparison of morphological data**

253   All 39 isolates were examined and identified morphologically:

254   17 isolates were analysed by two independent taxonomists

255   (replicate identifications), while 22 isolates were analysed by

256   either one of the two taxonomists (unique identifications) (Fig

257   2, Table 2).

258   Overall, only 46% of the isolates (18/39) were morphologically

259   identified to species level by at least one taxonomist; the

260   remaining isolates (54%; 21/39) were identified only to the

261   genus level (Table 2). For the 17 isolates that were analysed by

262   both taxonomists, morphological identifications were in

263   complete agreement for three isolates (18%, 3/17): Buayanup

264   drain type 2 (*Anabaena torulosa*), Hyde Park type 1

265   (*Anabaenopsis elenkinii*), and Vasse River type 6

266   (*Sphaerospermopsis aphanizomenoides*) (Fig. 2, Table 2).

267   Baldwin Park type 3 was also identified to the species level by

268   both taxonomists, but as different species (*Anabaena*

269   *oscillaroides* vs. *Sp. aphanizomenoides*). For six further

270   isolates (35%, 6/17), morphological identifications were in

271   agreement at the genus level, while, with the exception of

272   Baldwin Park type 3, the lack of identifying features allowed

273   identification of the remaining seven isolates (41%), only at

274   either family, or order– level (Fig. 2, Table 2).

275

276   **3.2 Comparison of molecular identification and**

277   **phylogenetic analysis at different loci**.

278   Partial sequences were successfully obtained for the 16S rDNA

279   (n=36), *rpo*C1 (n=22) and PC-IGS (n=19) loci. Amplification

280   at all three loci was successful for 23% (9/39) of the isolates

281   (Table 2). Of the remaining isolates, 33% (10/30) amplified

282   successfully at both the 16S rDNA and *cpc*BA-IGS loci, 33%

283   (10/30) were successful at the 16S rDNA and *rpo*C1 loci, while

284   the remaining 33% (10/30) successfully amplified at only one

285   locus (Table 2). Tree topologies for all three loci were

286   relatively similar (Fig. 3, 4 and 5), with the isolates included in

287   distinct clusters according to their orders (Table 2, Fig. 3, 4 and

288   5).

289   Based on the percentage similarity threshold set (in the

290   Materials and Methods), when all three loci were successfully

291   sequenced from any given isolate and compared, molecular

292   identifications agreed at the species level for only one isolate

293   (Vasse River type 6), and at the genus level for one other

294   isolate (Vasse River type 13). Where successful amplification

295   was obtained for only two loci, agreement at the species level

296   for 25% (5/20) of the isolates was obtained. A further 50%

297   (10/20) of the isolates agreed at the genus level, while the

298   remaining 25% either agreed at the order level, or had no

299   agreement at the loci amplified (Table 2, Fig. 6).

300   Analysis of the 16S rDNA data (313 characters; 110 parsimony

301   informative sites) showed the existence of six major clusters,

302   within the three major cyanobacterial orders: the Chroococcales

303   (subsection I; n=8), Oscillatoriales (subsection III; n=6) and

304   Nostocales (subsection IV; n=22) (Fig. 3) (Castenholz, 2001).

305   The Chroococcales included sequences from planktic,

306   unicellular coccoids, isolated mainly from the Vasse River, and

307   from two isolates (GS3-1 and MIC058-B), which were multi-

13

308   cellular planktic colonies. The Oscillatoriales included benthic,

309   non-heterocystous, filamentous (non-branching) strains, from a

310   variety of locations. All other 16S rDNA sequences were

311   obtained from cultures of filamentous (non-branching) strains

312   clustering within the Nostocales, of which several genera were

313   paraphyletic (Fig. 3).

314   At the *rpo*C1 locus (409 characters; 335 parsimony informative

315   sites), the genotypes identified belonged to the Nostocales

316   (n=17), Oscillatoriales (n=1), and Chroococcales (n=4) (Fig. 4).

317   Interestingly, GS2-1 grouped distinctly with *Pseudanabaena*

318   sp. (Oscillatoriales) (bootstrap value > 50%), while Baldwin

319   Park type 2, although on an isolated branch, was found well

320   within the Nostocales. This is in contrast to the 16S locus,

321   where GS2-1 and Baldwin Park type 2, respectively grouped

322   with *Nostoc commune* (bootstrap > 70%; well within the

323   Nostocales), or formed a clearly distinct branch, basal to the

324   order (Fig. 3). As with the 16S tree (Fig. 3), multiple clusters of

325   the Chroococcales were also evident from the *rpo*C1 tree (Fig.

326   4), with four sequences grouped within this order: MIC058-B,

327   Vasse River types 9, and 13 and GS6-1. Apart from GS2-1, the

328   remaining 17 sequences clustered within the Nostocales, which

329   was characterized by the distinct positions of Vasse River type

330   2, Buayanup drain type 2 and GS5-2.

331   Analysis of the *cpc*BA-IGS locus (423 characters; 396

332   parsimony informative sites) showed that, apart from GS2-1,

333 which grouped strongly (bootstrap value >80%) with

334 *Pseudanabaena* sp. (Oscillatoriales) on an isolated branch, the

335 overall topology was similar to that obtained for the 16S rDNA

336 (Fig. 5). As with the 16S and *rpo*C1 loci, the Nostocales (10

337 sequences) and the Chroococcales (2 sequences) formed

338 monophyletic groups, while the Oscillatoriales (7 sequences)

339 were paraphyletic (Fig. 5).

340 As the majority of *cpc*BA-IGS sequences available from

341 GenBank to date mainly belong to relatively few genera (e.g.,

342 *Arthrospira*, *Synechococcus*, *Phormidium* etc.), large distance

343 values and the presence of isolated branches were observed for

344 the tree based on this locus (Fig. 5). Baldwin Park type 2 could

345 not be successfully amplified at this locus, and only GS4-1,

346 GS4-2, Baldwin Park type 1 and Vasse River type 3 exhibited

347 almost complete (≥99%) homology with available sequences

348 (Fig. 5, Table 2).

349

350 **3.3 Comparison between morphological and molecular**

351 **identifications**

352 Discrepancies between morphological and molecular

353 identifications were observed for several isolates (Table 2). For

354 example, Hyde Park type 1 was identified, at the *cpc*BA-IGS

355 locus and morphologically, as *An. elenkinii*. However, at the

356 16S rDNA it was closest to *An. circularis*. Further molecular

357 identification of Hyde Park type 1 was hampered by the paucity

15

358   of *An. circularis* and *An. elenkinii* sequences at both the *rpo*C1

359   and *cpc*BA-IGS loci which prevented confident identification

360   at these loci.

361   The isolate Buayanup drain type 2, which was identified

362   morphologically as *A. torulosa* was most similar to *A.*

363   *oscillaroides* at the 16S rDNA locus, and to *A. spherica* at the

364   *cpc*BA-IGS locus (no *A. oscillaroides* sequences were available

365   at this locus). Isolates GS6-1 and Vasse River types 9, 12 and

366   13 were identified as *Aphanothece* sp. based on morphology.

367   However, using molecular methods, they were phylogenetically

368   more similar to *Synechococcus* sp. (HE975005) than

369   *Aphanothece minutissima* (FM177488) (Fig. 3).

370   Morphologically, ANA196-A was identified as

371   *Dolichospermum circinale*, but was phylogenetically placed

372   with *Aphanizomenon gracile*, using the 16S rDNA and *rpo*C1

373   sequence data (Table 2). This was also observed for GS4-2,

374   which was identified morphologically as a *Nostoc* sp. or *Sp.*

375   *aphanizomenoides*, but was found to be most closely related to

376   *An. bergii* (100% similarity) at both the 16S rDNA and *cpc*BA-

377   IGS loci. Similarly, although Baldwin Park type 1 and GS4-1

378   were identified morphologically as *Planktolyngbya* and

379   *Oscillatoriales/Geitlerinema* sp., they showed 100% similarity

380   to various *Limnothrix* spp. and *Planktothrix* spp., at the 16S

381   rDNA locus, and to *Geitlerinema amphibium* (FJ545644), at

382   the *cpc*BA-IGS locus.

16

383   Overall, for the nine isolates that amplified at all loci studied,

384   microscopic and molecular data from at least one locus, were in

385   agreement at genus level for all isolates, except Vasse River

386   type 13 (Table 2). However, agreement between morphological

387   and molecular identifications, from all three loci, was obtained

388   for only one isolate (Vasse River type 6) (Table 2). When

389   morphology was compared with two loci for all isolates (16S

390   rDNA plus, either *rpo*C1, or *cpc*BA-IGS), species identities for

391   ANA150-A (*D. circinale*) and AWQC318 (*D. circinale*) were

392   in agreement. When morphological data was combined with

393   molecular identification from any one locus, a further seven

394   isolates could be identified to the species level. These included

395   ANA118-AR, ANA131-CR, ANA148-CR, ANA335-C (*D.*

396   *circinale*), Hyde Park type 1 (*An. elenkinii*), MIC-058-B

397   (*Microcystis flos-aquae*), and Vasse River type 1 (*D. flos-*

398   *aquae*). Of the remaining isolates, 31% (12/39) were in

399   agreement at the genus level, 28% (11/39) to the order level,

400   while no agreement was obtained for the remaining 15% (6/39)

401   (Table 2).

402

403   **3.4 Identification of novel isolates**

404   Based on their unique/variable phylogenetic positions and large

405   genetic distances from available sequences, two potentially

406   new members of the Nostocales (Baldwin Park type 2 –

407   morphologically *Anabaena* sp. 1, and GS2-1 – morphologically

17

408 *Nostoc* sp.) were identified. These new strains had no

409 particularly atypical morphology, and could only be

410 morphologically identified to genus level (*Anabaena* sp. 1, and

411 *Nostoc* sp., respectively). Noteworthy, 17 novel sequences with

412 <95% similarity to previously published sequences were

413 obtained during this study, with the majority of the novel

414 sequences being observed at the *rpo*C1 locus.

415

416 **4. Discussion**

417 This study has shown that even with the changes to

418 cyanobacteria taxonomy and the vast increase in sequences

419 available, identifying cyanobacteria isolates using a

420 combination of molecular and/or morphological methods still

421 remains problematic. This becomes even more evident when,

422 as done in this study, multiple loci and morphological

423 identifications of a single isolate are compared.

424 Despite major revisions to the taxonomy and systematics of

425 cyanobacteria, of the 17 isolates microscopically examined in

426 duplicate, agreement at species level was obtained for three

427 isolates, with another seven isolates agreeing to the genus level.

428 This clearly highlights the current difficulties in morphological

429 identification of cyanobacteria from environmental samples. As

430 such, when monitoring water bodies, it may be beneficial to

431 have more than one taxonomist. Furthermore, to maintain

432 consistency and integrity of the morphological identifications,

433   having morphological identifications done by the same

434   taxonomist for the duration of the study is also of importance.

435   Overall, tree topologies for the 16S rDNA and *rpo*C1 were

436   similar to previous publications (Fergusson and Saint, 2000;

437   Litvaitis, 2002; Lyra et al., 2001; Tomitani et al., 2006; Valério

438   et al., 2009). Although the 16S rDNA alignment was relatively

439   short (313 characters), using full-length reference sequences

440   from GenBank did not alter the clustering patterns and tree

441   topology (data not shown). Moreover, an *rpo*C1 amino acid

442   alignment produced a similar tree to that obtained in Fig. 5

443   (data not shown).

444   The strictly qualitative nature of this study discourages the

445   application of a statistically meaningful analysis or inference of

446   ecological and water quality parameters of the sampled sites. In

447   particular, the isolation methods implemented may have

448   favoured isolation of particular species, and cannot therefore be

449   used to comprehensively survey the original cyanobacteria

450   communities. Even so, 22% of the total number of sequences

451   obtained had less than 95% similarity to previously published

452   sequences (at any of the three loci studied), confirming the

453   presence of unexplored molecular varieties of freshwater

454   cyanobacteria in Western Australia.

455   Although there have been investigations on cyanobacterial

456   diversity in Australia (Fergusson and Saint, 2000; McGregor

457   and Rasmussen, 2008; Papineau et al., 2005; Saker et al.,

19

458   2009), this is the first study to compare morphological and

459   molecular identification of freshwater cyanobacteria from

460   Western Australia. In contrast to previous studies done in this

461   region, where Oscillatoriales and/or Chroococcales were found

462   to predominate (Garby et al., 2013; Gordon et al., 1981; Kemp

463   and John, 2006; Lund and Davis, 2000), majority of the isolates

464   obtained in this study belonged to the Nostocales.

465   The present study identified two potentially new strains of

466   cyanobacteria; of these, GS2-1 was the most interesting,

467   grouping with either the Nostocales or Oscillatoriales,

468   depending on the locus considered. Generally speaking, these

469   observed differences can be due to: (i) the lack of sequences

470   available in GenBank for the *rpo*C1 and *cpc*BA-IGS loci; (ii)

471   preferential amplification of contaminating strains; (iii)

472   horizontal gene transfer, or (iv) presence of habitat specific

473   gene pools (Komárek, 2010). Baldwin Park type 2 was also of

474   interest, as it was basal to the Nostocales at the well-studied

475   16S rDNA locus, but fell on an isolated branch on the *rpo*C1

476   tree. Minimum genetic distances for this isolate at either locus

477   were also considerable (6% and 14%, for the 16S rDNA and

478   *rpo*C1 locus respectively, from *A. oscillaroides* and *Anabaena*

479   sp. respectively), indicating a potentially previously

480   uncharacterized *Anabaena* species.

481   Despite its wide usage, the 16S rDNA has been found to be too

482   conserved to reliably differentiate closely related bacterial

483   species (Coenye and Vandamme, 2003; Lyra et al., 2001), and

484   to have an evolutionary pattern that is not reflective of the

485   entire genome (Seo and Yokota, 2003). Consequently,

486   alternative loci such as the protein coding *rpo*C1, the *cpc*BA-

487   IGS, the nitrogenase genes (Kumari et al., 2009) and the 16-

488   23S ITS region (and its structure) have also been used for

489   phylogenetic reconstructions, identification and discrimination

490   of species (Johansen et al., 2011; Palenik and Haselkorn, 1992).

491   Although polyphasic approaches, combining morphological

492   and molecular identifications, have been proposed (Komárek,

493   2006; Litvaitis, 2002; Robertson et al., 2001; Seo and Yokota,

494   2003), a number of authors have demonstrated inconsistent

495   phylogenetic and morphological classifications (Litvaitis, 2002;

496   Robertson et al., 2001; Seo and Yokota, 2003). Thus,

497   incorporating data from multiple sources (e.g. morphology,

498   nucleotide sequences from multiple loci, biochemical

499   composition) for the identification of unknown environmental

500   genotypes has been recommended as a standard taxonomic

501   practice (Komárek, 2006). Furthermore, apart from the 97% –

502   98% sequence similarity for the 16S rDNA (Stackebrandt and

503   Goebel, 1994), there is no consensus percentage sequence

504   similarity for species delimitation using other loci. Hence, as a

505   reflection of the *rpo*C1 and *cpc*BA-IGS sequence numbers

506   available (Fig. 1), a less stringent criterion was used for the

507   determination of species and genus for these loci. Despite this,

508 molecular agreement between three loci was still lower than

509 between two loci. The combination of *cpc*BA-IGS and *rpo*C1,

510 however, showed no agreement among all pairs; conformity

511 was found only when a third locus (i.e. 16S rDNA) was

512 included (Fig. 6). These trends can be explained by the effects

513 of two intertwined factors (which are ultimately responsible for

514 the successful molecular identification of a given isolate): the

515 total number of sequences available at a specific locus, and the

516 number of species represented at that same locus.

517 Amplification efficiency of certain primer sets can potentially

518 be affected by large genetic variations, possibly explaining why

519 none of the isolates belonging to the Oscillatoriales

520 successfully amplified at the *rpo*C1 locus. Furthermore,

521 alignment length and gap treatment options adopted are known

522 to affect phylogenetic reconstructions including the resolution

523 of some areas of the tree (Lindgren and Daly, 2007). Where

524 alignments of protein-coding genes should present fewer gaps,

525 regions which are subject to less stringent genetic constraints

526 (e.g., 16S hyper-variable regions and intergenic spacers) may

527 produce numerous positions with gaps, requiring *ad hoc*

528 strategies different from protein-coding data sets (Talavera and

529 Castresana, 2007). This, potentially, accounts for some of the

530 topological differences observed in the trees produced. To

531 overcome such limitations stemming from sequence variability,

532 the use of cyanobacterial 16S-23S ITS secondary structure may

22

533   provide an alternative method to identifying intergenic, and

534   possibly intragenic, diversity which is congruent with that of

535   the 16S rDNA as demonstrated by Johansen et al. (2011).

536   Similarly, the choice of algorithms for generating trees and

537   alignments, as well as the nucleotide substitution models used,

538   can potentially affect the successful identification of isolates

539   (Eddy and Durbin, 1994; Lindgren and Daly, 2007). This could

540   explain, for instance, the high similarity of MIC058-B-derived

541   16S and *rpo*C1 sequences with *M. aeruginosa* and *M. flos-*

542   *aquae* reference sequences, despite the morphological

543   identification as *M. flos-aquae* (Fig 3 and 4). On the other

544   hand, however, this observation is in agreement with the

545   recommendation by Otsuka et al. (2001) that *M. flos-aquae* be

546   regarded as a morphological variant of *M. aeruginosa*.

547   In addition, it is well known that cyanobacteria frequently

548   undergo morphological changes during cultivation (Gugger et

549   al., 2002; Komárek and Anagnostidis, 1989; Lyra et al., 2001),

550   resulting in potential loss of taxon-defining features (Gugger et

551   al., 2002). Not only have the discrepancies between molecular

552   and microscopic characterizations been well documented

553   (Komárek, 2010; Willame et al., 2006), but it has also been

554   shown that incorrectly identified species exist in GenBank

555   (Komárek, 2010) (e.g. *A. variabilis* EF488831). Strains within

556   culture collections have also been incorrectly named and are

557   present in both GenBank and various culture collections under

558   different designations (e.g. *A. variabilis* ATCC29413 which

559   also appears as Nostoc sp. PCC7937, *A. variabilis* UTCC 105,

560   Anabaena PCC 7937, *A. flos-aquae* UTEX144 etc.).

561   Furthermore, with the many revisions to cyanobacteria

562   taxonomy and nomenclature (Fig. 1), there is no method for

563   these corrections to be easily incorporated into databases

564   (Komárek, 2006), causing further confusion when attempting to

565   identify environmental cyanobacteria isolates.

566   Finally, as pointed out by Castenholz and Norris (2005),

567   species identification is still blurred, this, together with the

568   difficulties discussed when identifying cyanobacteria, questions

569   the need for cyanobacteria identification to the species level.

570   This is especially true in water monitoring situations where

571   identifications to the genus level would usually be sufficient for

572   initiation of remediation strategies.

573   In recent years, the use of mass spectrometry – namely matrix

574   assisted laser desorption/ionisation- time of flight (MALDI-

575   TOF) has seen increasing use in various fields. Its applicability

576   to detecting cyanotoxins (Welker et al., 2002), chemotypes

577   (Welker and Erhard, 2007) and determining cyanobacteria

578   spatial distribution has also been shown, allowing for the

579   possible future use of this technology for cyanobacteria

580   identification, with the development of a standardised and

581   regularly updated database.

582

24

583 **5. Conclusion**

584 As highlighted by various authors (Castenholz and Norris,

585 2005; Komárek, 2006; Oren, 2011), a polyphasic approach is

586 still currently the most reliable option for identifying

587 cyanobacteria. The present study however has highlighted that

588 despite the application of molecular techniques and the

589 subsequent increase in publicly available sequences, the ability

590 to accurately and definitively identify environmental

591 cyanobacteria isolates is still challenging. It is expected that

592 whole genome sequencing data from an ever increasing number

593 of species, should make misidentified sequences in GenBank

594 increasingly recognisable, allowing for a prompter exclusion or

595 revision. This combination of whole genome sequences,

596 together with alternative methods of cyanobacteria

597 identification (e.g. mass spectrometry) will facilitate the

598 process of cyanobacteria identification. However, until then,

599 accurate identification of these organisms will remain

600 problematic.

601

602 **Acknowledgments**

613

References

Anagnostidis, K., 2001. Nomenclatural changes in cyanoprokaryotic order Oscillatoriales. Preslia (Prague) 73(4), 359-375.

Borowitzka, M.A., 1995. Microalgae as sources of pharmaceuticals and other biologically active compounds. Journal of Applied Phycology 7(1), 3-15.

Castenholz, R.W., 2001. Phylum BX. Cyanobacteria. In: Boone, D.R., Castenholz, R.W., Garrity, G.M. (eds), Bergey's Manual of Systematic Bacteriology 2nd edn. Springer, New York. p. 473-599

Castenholz, R.W., Norris, T.B., 2005. Revisionary concepts of species in the cyanobacteria and their applications. Archiv für Hydrobiologie Supplement 159, 53-69.

Coenye, T., Vandamme, P., 2003. Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. FEMS Microbiology Letters 228(1), 45-49.

Eddy, S.R., Durbin, R., 1994. RNA sequence-analysis using covariance-models. Nucleic Acids Research 22(11), 2079-2088.

Fergusson, K.M., Saint, C.P., 2000. Molecular phylogeny of *Anabaena circinalis* and its identification in environmental samples by PCR. Applied and Environmental Microbiology 66(9), 4145-4148.

Garby, T.J., Walter, M.R., Larkum, A.W.D., Neilan, B.A., 2013. Diversity of cyanobacterial biomarker genes from the stromatolites of Shark Bay, Western Australia. Environmental Microbiology 15(5), 1464-1475.

Gordon, D.M., Finlayson, C.M., McComb, A.J., 1981. Nutrients and phytoplankton in three shallow, freshwater lakes of different trophic status in Western Australia. Australian Journal of Marine and Freshwater Research 32(4), 541-553.

Gorham, P.R., McLachlan, R.W., Hammer, U.T., 1964. Isolation and culture of toxic strains of *Anabaena flos-aquae*. Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie 19, 796-804.

Granéli, E., Turner, J.T., 2006. Ecology of harmful algae, Springer, Germany.

Gugger, M., Lyra, C., Henriksen, P., Couté, A., Humbert, J.-F., Sivonen, K., 2002. Phylogenetic comparison of the cyanobacterial genera *Anabaena* and *Aphanizomenon*. International Journal of Systematic and Evolutionary Microbiology 52(5), 1867-1880.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95-98.

662   Johansen, J.R., Kovacik, L., Casamatta, D.A., Fucikova, K.,
663   Kastovsky, J., 2011. Utility of 16S-23S ITS sequence and
664   secondary structure for recognition of intrageneric and
665   intergeneric limits within cyanobacterial taxa: *Leptolyngbya*
666   *corticola* sp. nov (Pseudanabaenaceae, Cyanobacteria).
667   Nova Hedwigia 92(3-4), 283-302.
668   Kemp, A., John, J., 2006. Microcystins associated with
669   microcystis dominated blooms in the southwest wetlands,
670   Western Australia. Environmental Toxicology 21(2), 125-
671   130.
672   Kimura, M., 1980. A simple method for estimating
673   evolutionary rates of base substitutions through
674   comparative studies of nucleotide sequences. Journal of
675   Molecular Evolution 16(2), 111-120.
676   Komárek, J., 2006. Cyanobacterial taxonomy: Current
677   problems and prospects for the integration of traditional and
678   molecular approaches. Algae 21(4), 349-375.
679   Komárek, J., 2010. Recent changes (2008) in cyanobacteria
680   taxonomy based on a combination of molecular background
681   with phenotype and ecological consequences (genus and
682   species concept). Hydrobiologia 639(1), 245-259.
683   Komárek, J., Anagnostidis, K., 1989. Modern approach to the
684   classification system of Cyanophytes 4 - Nostocales.
685   Archiv für Hydrobiologie - Algological Studies,
686   Supplement Volumes 56, 247-345.
687   Komárek, J., Kastovsky, J., Jezberova, J., 2011. Phylogenetic
688   and taxonomic delimitation of the cyanobacterial genus
689   *Aphanothece* and description of *Anathece gen. nov.*
690   European Journal of Phycology 46(3), 315-326.
691   Kumari, N., Srivastava, A.K., Bhargava, P., Rai, L.C., 2009.
692   Molecular approaches towards assessment of
693   cyanobacterial biodiversity. African Journal of
694   Biotechnology 8(18), 4284-4298.
695   Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R.,
696   McGettigan, P.A., McWilliam, H., Valentin, F., Wallace,
697   I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J.,
698   Higgins, D.G., 2007. Clustal W and Clustal X version 2.0.
699   Bioinformatics 23(21), 2947-2948.
700   Li, Y., Horsman, M., Wu, N., Lan, C.Q., Dubois-Calero, N.,
701   2008. Biofuels from microalgae. Biotechnology Progress
702   24(4), 815-820.
703   Lindgren, A.R., Daly, M., 2007. The impact of length-variable
704   data and alignment criterion on the phylogeny of
705   Decapodiformes (Mollusca: Cephalopoda). Cladistics
706   23(5), 464-476.
707   Litvaitis, M.K., 2002. A molecular test of cyanobacterial
708   phylogeny: inferences from constraint analyses.
709   Hydrobiologia 468(1-3), 135-145.
710   Lund, M.A., Davis, J.A., 2000. Seasonal dynamics of plankton
711   communities and water chemistry in a eutrophic wetland

712   (Lake Monger, Western Australia): implications for
713   biomanipulation. Marine and Freshwater Research 51(4),
714   321-332.

715   Lyra, C., Suomalainen, S., Gugger, M., Vezie, C., Sundman, P.,
716   Paulin, L., Sivonen, K., 2001. Molecular characterization of
717   planktic cyanobacteria of *Anabaena*, *Aphanizomenon*,
718   *Microcystis* and *Planktothrix* genera. International Journal
719   of Systematic and Evolutionary Microbiology 51(2), 513-
720   526.

721   McGregor, G.B., Rasmussen, J.P., 2008. Cyanobacterial
722   composition of microbial mats from an Australian thermal
723   spring: a polyphasic evaluation. FEMS Microbiology
724   Ecology 63(1), 23-35.

725   McNeill, J., Barrie, F.R., Buck, W.R., Demoulin, V., Greuter,
726   W., Hawkworth, D.L., Herendeen, P.S., Knapp, S., Turland,
727   N.J., 2012. International Code of Botanical Nomenclature
728   for Algae, Fungi and Plants (Melbourne Code): Adopted by
729   the 18th International Botanical Congress Melbourne
730   Australia July 2011, Koeltz Scientific Books.

731   Neilan, B.A., Jacobs, D., Goodman, A.E., 1995. Genetic
732   diversity and phylogeny of toxic cyanobacteria determined
733   by DNA polymorphisms within the phycocyanin locus.
734   Applied and Environmental Microbiology 61(11), 3875-
735   3883.

736   Nübel, U., Garcia-Pichel, F., Muyzer, G., 1997. PCR primers to
737   amplify 16S rRNA genes from cyanobacteria. Applied and
738   Environmental Microbiology 63(8), 3327-3332.

739   Oren, A., 2004. A proposal for further integration of the
740   cyanobacteria under the Bacteriological Code. International
741   Journal of Systematic and Evolutionary Microbiology
742   54(5), 1895-1902.

743   Oren, A., 2011. Cyanobacterial systematics and nomenclature
744   as featured in the International Bulletin of Bacteriological
745   Nomenclature and Taxonomy/ International Journal of
746   Systematic Bacteriology/ International Journal of
747   Systematic and Evolutionary Microbiology. International
748   Journal of Systematic and Evolutionary Microbiology
749   61(1), 10-15.

750   Oren, A., Tindall, B.J., 2005. Nomenclature of the cyanophyta/
751   cyanobacteria/ cyanoprokaryotes under the International
752   Code of Nomenclature of Prokaryotes. Archiv für
753   Hydrobiologie Supplement 159, 39-52.

754   Otsuka, S., Suda, S., Shibata, S., Oyaizu, H., Matsumoto, S.,
755   Watanabe, M.M., 2001. A proposal for the unification of
756   five species of the cyanobacterial genus *Microcystis*
757   Kutzing *ex* Lemmermann 1907 under the rules of the
758   Bacteriological Code. International Journal of Systematic
759   and Evolutionary Microbiology 51, 873-879.

760   Palenik, B., Haselkorn, R., 1992. Multiple evolutionary origins
761   of prochlorophytes, the chlorophyll b-containing
762   prokaryotes. Nature 355(6357), 265-267.
763   Papineau, D., Walker, J.J., Mojzsis, S.J., Pace, N.R., 2005.
764   Composition and structure of microbial communities from
765   stromatolites of Hamelin Pool in Shark Bay, Western
766   Australia. Applied and Environmental Microbiology 71(8),
767   4822-4832.
768   Rajaniemi, P., Hrouzek, P., Klára, K., Willame, R., Rantala, A.,
769   Hoffmann, L., Komárek, J., Sivonen, K., 2005.
770   Phylogenetic and morphological evaluation of the genera
771   *Anabaena*, *Aphanizomenon*, *Trichormus* and *Nostoc*
772   (Nostocales, Cyanobacteria). International Journal of
773   Systematic and Evolutionary Microbiology 55(1), 11-26.
774   Reynolds, C.S., 2006. Ecology of phytoplankton, Cambridge
775   University Press, New York. p. 93-151
776   Robertson, B.R., Tezuka, N., Watanabe, M.M., 2001.
777   Phylogenetic analyses of *Synechococcus* strains
778   (cyanobacteria) using sequences of 16S rDNA and part of
779   the phycocyanin operon reveal multiple evolutionary lines
780   and reflect phycobilin content. International Journal of
781   Systematic and Evolutionary Microbiology 51(3), 861-871.
782   Saker, M., Moreira, C., Martins, J., Neilan, B., Vasconcelos,
783   V.M., 2009. DNA profiling of complex bacterial
784   populations: toxic cyanobacterial blooms. Applied
785   Microbiology and Biotechnology 85(2), 237-252.
786   Seo, P.S., Yokota, A., 2003. The phylogenetic relationships of
787   cyanobacteria inferred from 16S rRNA, *gyr*B, *rpo*C1 and
788   *rpo*D1 gene sequences. Journal of General and Applied
789   Microbiology 49(3), 191-203.
790   Siegesmund, M.A., Johansen, J.R., Karsten, U., Friedl, T.,
791   2008. *Coleofasciculus* Gen. Nov (Cyanobacteria):
792   Morphological and molecular criteria for the revision of the
793   genus *Microcoleus gomont*. Journal of Phycology 44(6),
794   1572-1585.
795   Stackebrandt, E., Goebel, B.M., 1994. Taxonomic Note: A
796   place for DNA-DNA reassociation and 16S rRNA sequence
797   analysis in the present species definition in bacteriology.
798   International Journal of Systematic Bacteriology 44(4),
799   846-849.
800   Strunecky, O., Elster, J., Komárek, J., 2011. Taxonomic
801   revision of the freshwater cyanobacterium "*Phormidium*"
802   *murrayi* = *Wilmottia murrayi*. Fottea 11(1), 57-71.
803   Suda, S., Watanabe, M.M., Otsuka, S., Mahakahant, A.,
804   Yongmanitchai, W., Nopartnaraporn, N., Liu, Y.D., Day,
805   J.G., 2002. Taxonomic revision of water-bloom-forming
806   species of oscillatorioid cyanobacteria. International
807   Journal of Systematic and Evolutionary Microbiology 52,
808   1577-1595.

809 Talavera, G., Castresana, J., 2007. Improvement of phylogenies
810 after removing divergent and ambiguously aligned blocks
811 from protein sequence alignments. Systematic Biology
812 56(4), 564-577.
813 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M.,
814 Kumar, S., 2011. MEGA5: Molecular evolutionary genetics
815 analysis using maximum likelihood, evolutionary distance,
816 and maximum parsimony methods. Molecular Biology and
817 Evolution 28(10), 2731-2739.
818 Tomitani, A., Knoll, A.H., Cavanaugh, C.M., Ohno, T., 2006.
819 The evolutionary diversification of cyanobacteria:
820 Molecular-phylogenetic and paleontological perspectives.
821 Proceedings of the National Academy of Sciences of the
822 United States of America 103(14), 5442-5447.
823 Valério, E., Chambel, L., Paulino, S., Faria, N., Pereira, P.,
824 Tenreiro, R., 2009. Molecular identification, typing and
825 traceability of cyanobacteria from freshwater reservoirs.
826 Microbiology 155(2), 642-656.
827 Wacklin, P., Hoffmann, L., Komárek, J., 2009. Nomenclatural
828 validation of the genetically revised cyanobacterial genus
829 *Dolichospermum* (RALFS ex BORNET et FLAHAULT)
830 comb. nova. Fottea 9(1), 59-64.
831 Welker, M., Erhard, M., 2007. Consistency between
832 chemotyping of single filaments of *Planktothrix rubescens*
833 (cyanobacteria) by MALDI-TOF and the peptide patterns of
834 strains determined by HPLC-MS. Journal of Mass
835 Spectrometry 42(8), 1062-1068.
836 Welker, M., Fastner, J., Erhard, M., von Dohren, H., 2002.
837 Applications of MALDI-TOF MS analysis in cyanotoxin
838 research. Environmental Toxicology 17(4), 367-374.
839 Whitton, B.A., Potts, M., 2000. Introduction to the
840 Cyanobacteria. In: Whitton, B.A., Potts, M. (eds), The
841 ecology of Cyanobacteria. Their diversity in time and
842 space. Springer, Berlin. p. 1-9
843 Willame, R., Boutte, C., Grubisic, S., Wilmotte, A., Komárek,
844 J., Hoffmann, L., 2006. Morphological and molecular
845 characterization of planktonic cyanobacteria from Belgium
846 and Luxembourg. Journal of Phycology 42(6), 1312-1332.
847
848

Table 1 – Collection sites and identities of the 39 cyanobacterial isolates used in this study

Type	Sampling site	Isolates obtained	Isolate identities
Protected *freshwater reservoirs* (WA) (n=6)	Great Southern Region 1	2	GS1-1, GS1-2
	Great Southern Region 2	1	GS2-1
	Great Southern Region 3	2	GS3-1, GS3-2,
	Great Southern Region 4	2	GS4-1, GS4-2
	Great Southern Region 5	3	GS5-1, GS5-2, GS5-3
	Great Southern Region 6	2	GS6-1, GS6-2
Protected *freshwater reservoirs* (NSW) (n=4)	Chaffey Dam	1	ANA019-BR
	Burrinjuck Reservoir	1	ANA150-A
	Copeton Dam	1	ANA278-FR
	Pejar Dam	1	AWQC318
Protected *freshwater reservoir* (VIC) (n=1)	Fish Creek Farm Dam	1	ANA148-CR
Open *freshwater reservoir* (SA) (n=1)	Millbrook Reservoir	1	MIC058-B
Open urban *lentic* systems (WA) (n=3)	Chelodina Wetland Reserve	1	Chelodina wetland type 1
	Frederick Baldwin Park	3	Baldwin park types 1, 2, 3
	Hyde Park	1	Hyde Park type 1
Open rural *lentic* systems (NSW) (n=2)	Lachlan River, Booligal	1	ANA118-AR
	Willandra Creek	1	ANA196-A
Open rural *lotic* systems (WA) (n=2)	Buayanup Drain	2	Buayanup drain types 2, 4
	Vasse River	10	Vasse River types 1, 2, 3, 6, 8, 9, 12, 13, 14, 15
Open rural *lotic* systems (NSW) (n=1)	Lake Cargelligo	1	ANA131-CR
Open rural *lotic* systems (VIC) (n=1)	Murrary River, Swan Hill	1	ANA335-C
Grand-total			**39**

Table 2: Cyanobacteria identification based on 16S, *rpo*C1, Phycocyanin (Cpc) DNA sequences and isolate morphology. Percentage similarity at each locus was calculated in MEGA 5 (Tamura et al., 2011), as the pairwise evolutionary divergence, with a p-distance model (Kimura, 1980).

Sub-group	Isolate	Molecular Identification Related sequence (percentage similarity)			Morphological identification	
		16S rDNA (n=36)	*rpo*C1 (n=22)	*Cpc*BA-IGS (n=19)	Taxonomist 1	Taxonomist 2
Nostocales (n=25)	ANA019-BR	*D. affine* |FN691906 **(100%)**	–	–	*D. circinale*	N.A.
	ANA118-AR	–	*D. circinale* |AF199423 **(100%)**	–	*D. circinale*	N.A.
	ANA131-CR	–	*D. circinale* |AF199423 **(100%)**	–	*D. circinale*	N.A.
	ANA148-CR	–	*D. circinale* |AF199423 **(100%)**	–	*D. circinale*	N.A.
	ANA150-A[*]	(*D. circinale* AWQC150-A |AF247573)	*D. circinale* |AF199423 **(100%)**	–	*D. circinale*	N.A.
	ANA196-A	*Ap. gracile* |HQ157688 **(100%)**	*Ap. gracile* |EU078450 **(97%)**	–	*D. circinale*	N.A.
	ANA278-FR	*D. flos-aquae* |AB551438 **(100%)**	–	–	*D. circinale*	N.A.
	ANA335-C	*D. circinale* |AF247588 **(100%)**	–	–	*D. circinale*	N.A.
	AWQC318	*D. circinale* |AF247581 **(100%)**	*D. circinale* |AF199425 **(100%)**	–	*D. circinale*	N.A.
	Baldwin Park type 2	*A. oscillaroides* |AJ630428 (94%)	*Anabaena* sp. |AF199432 (86%) *Anabaena* sp. |AF199433 (86%)	–	N.A.	*Anabaena* sp. 1
	Baldwin Park type 3	*Nostoc* **sp. PCC8976** |**AM711525 (100%)** *Nostoc* **sp.** |**AB087403 (100%)**	*Nostoc* sp. |AY242997 (88%)	–	*Sp. aphanizomenoides*	*A. oscillaroides*
	Buayanup drain type 2	*A. oscillaroides* |AJ630428 **(99%)**	*Sp. aphanizomenoides* |FJ234841 (86%)	*A. sphaerica* |**DQ439645 (92%)**	*A. torulosa*	*A. torulosa*

			Sp. aphanzomenoides |FJ234848 (86%)		
Chelodina wetlands type 1	*Ap. gracile* |**EU078532 (99%)**	–	*Ap. gracile* |**FN552318 (97%)** *D. compacta* |**AY702239 (97%)**	*Anabaena* sp.	*Anabaena* sp.
GS1-2	*N. punctiforme* |**GQ287652 (97%)**	*A. variabilis* |CP000117 (89%)	–	*Anabaena* sp.	*Anabaena* sp.
GS2-1	*N. commune* |**DQ185223 (99%)** *N. commune* |**AB251863 (99%)**	*Pseudanabaena* sp. PCC7367 |CP003592 (76%)	*Pseudanabaena* sp. |EF680776 (80%)	*Nostocales*	*Nostoc* sp.
GS4-2	*A.bergii* |**FR822617 (100%)**	–	*A. bergii* |**FJ234863 (100%)**	*Sp. aphanizomenoides*	*Nostoc* sp.
GS5-1	Nostocaceae cyanobacterium |GQ389643 (100%)	*C. stagnale* PCC7417 |CP003642 (87%)	–	*Cylindrospermum* sp.	*Anabaena* sp.
GS5-2	*T. variabilis* |**AJ630456 (100%)** *T. variabilis* |**JQ390607 (100%)**	*A. cylindrica* PCC7144 |CP003659 (84%)	*Anabaena* sp. |**GU935369 (97%)**	N.A.	*Anabaena* sp.
GS5-3	*Nostoc* sp. |**FJ948088 (99%)**	–	*Nostoc* sp. PCC6720 |**JF740673 (94%)**	*Anabaena* sp.	*Anabaena* sp.
Hyde Park type 1	*An. circularis* |**GQ859629 (100%)**	*An. circularis* |EU078479 (89%)	*An. elenkinii* |**FN552383 (96%)**	*An. elenkinii*	*An. elenkinii*
Vasse River type 1	*D. flos-aquae* |**AY701573 (100%)**	*Anabaena* sp. |**AF199432 (98%)**	*Sp. aphanizomenoides* |**GU197719 (99%)**	*Anabaena* sp.	*D. flos-aquae*
Vasse River type 2	*Anabaena* sp. PCC9109 |**AY768408 (99%)**	*Anabaena* sp. |EU078475 (87%)	*Anabaena* sp. PCC 9109 |**AY768473 (96%)**	N.A.	*Anabaena* sp.
Vasse River type 3	*A. sphaerica* |**GQ466513 (100%)**	*A. variabilis* |**AB074795 (100%)**	*N. linckia* |**AY466120 (99%)**	N.A.	*Anabaena* sp.

	Vasse River type 6	*Sp. aphanizomenoides* **|FM177473 (100%)**	*Sp. aphanizomenoides* **|FJ830555 (96%)**	*Sp. aphanizomenoides* **|GU197719 (97%)**	*Sp. aphanizomenoides*	*Sp. aphanizomenoides*
	Vasse River type 8	*Calothrix* **sp. |GQ859627 (98%)**	–	–	*Gloeotrichia* sp.	*Rivularia* sp.
Oscillatoriales (n=6)	Baldwin Park type 1	*Planktothrix* **sp. |AF212922 (100%)** *L. redekei* **|EU078512 (100%)** *Limnothrix* **sp. |EF088338 (100%)**	–	*G. amphibium* **|FJ545644 (100%)**	*Oscillatoriales*	*Planktolyngbya* sp. 1
	Buayanup drain type 4	*Leptolyngbya* **sp. |EU729062 (98%)**	–	Cyanobacterium txid129981 |AJ401183 (75%)	N.A.	*Trichocoleus* sp
	GS1-1	*Leptolyngbya* **sp. |HM217044 (100%)**	–	*Spirulina laxissima* |DQ393286 (84%)	N.A.	*Planktolyngbya* sp.
	GS3-2	*Pseudanabaena* **sp. |GU935355 (100%)**	–	*Pseudanabaena* sp. PCC 7409 |M99426 (82%)	*Pseudanabaena* sp.	*Ps. galeata*
	GS4-1	*Planktothrix* **sp. |AF212922 (100%)** *L. redekei* **|EU078512 (100%)** *Limnothrix* **sp. |EF088338 (100%)**	–	*G. amphibium* **|FJ545644 (100%)**	*Geitlerinema* sp.	*Planktolyngbya* sp. 1
	GS6-2	*Ps. mucicola* **|GQ859642 (99%)**	–	*Pseudanabaena* sp. PCC7409 |M99426 (83%)	*Pseudanabaena* sp.	*Ps. galeata*
Chroococcales (n=8)	GS3-1	*M. flos-aquae* **|AF139328 (98%)**	–	–	N.A.	*Microcystis* sp.
	MIC058-B	*M. flos-aquae* **|AF139328 (100%)**	*M. aeruginosa* **|AP009552 (98%)**	–	*M. flos-aquae*	N.A.
	GS6-1	*Synechococcus* **sp. |HE975005 (100%)**	*Synechococcus* **sp. PCC7920 |AF245158 (96%)**	–	N.A.	*Aphanothece* sp.

Vasse River type 9	*Synechococcus* sp. |HE975005 (100%)	*Synechococcus* sp. PCC7920 |AF245158 (96%)	–	*Aphanothece* sp.	*Aphanothece* sp.
Vasse River type 12	*Synechococcus* sp. |HE975005 (100%)	–	*Synechococcus* sp. PCC7918 |AF223462 (97%)	N.A.	*Aphanothece* sp. 1
Vasse River type 13	*Synechococcus* sp. |HE975005 (100%)	*Synechococcus* sp. PCC7920 |AF245158 (96%)	*Synechococcus* sp. PCC7918 |AF223462 (96%)	N.A.	*Aphanothece* sp. 1
Vasse River type 14	*Synechococcus* sp. |HE975005 (100%)	–	–	N.A.	*Aphanothece* sp.
Vasse River type 15	*Synechococcus* sp. |HE975005 (100%)	–	–	N.A.	*Aphanothece* sp. 1

[*]16S rDNA sequence previously available (GenBank Acc. No AF247573); – = No sequence data obtained; N.A.= Not examined; *A.= Anabaena; An.=Anabaenopsis; Ap.=Aphanizomenon; C.=Cylindrospermum; D.=Dolichospermum; G.=Geitlerinema; L.=Limnothrix; M.=Microcystis; N.=Nostoc; O.=Oscillatoria; P.=Planktothrix; Ps.=Pseudanabaena; Sp.=Sphaerospermopsis; T.=Trichormus.* Sequences with percentage molecular similarity above the threshold used for identification are in bold. For species identification, the threshold was 98% for the 16S rDNA, and 95% for the *rpo*C1 and *cpc*BA-IGS loci. For genus, the threshold was 95% for the 16S rDNA, and 90% for *rpo*C1 and *cpc*BA-IGS.

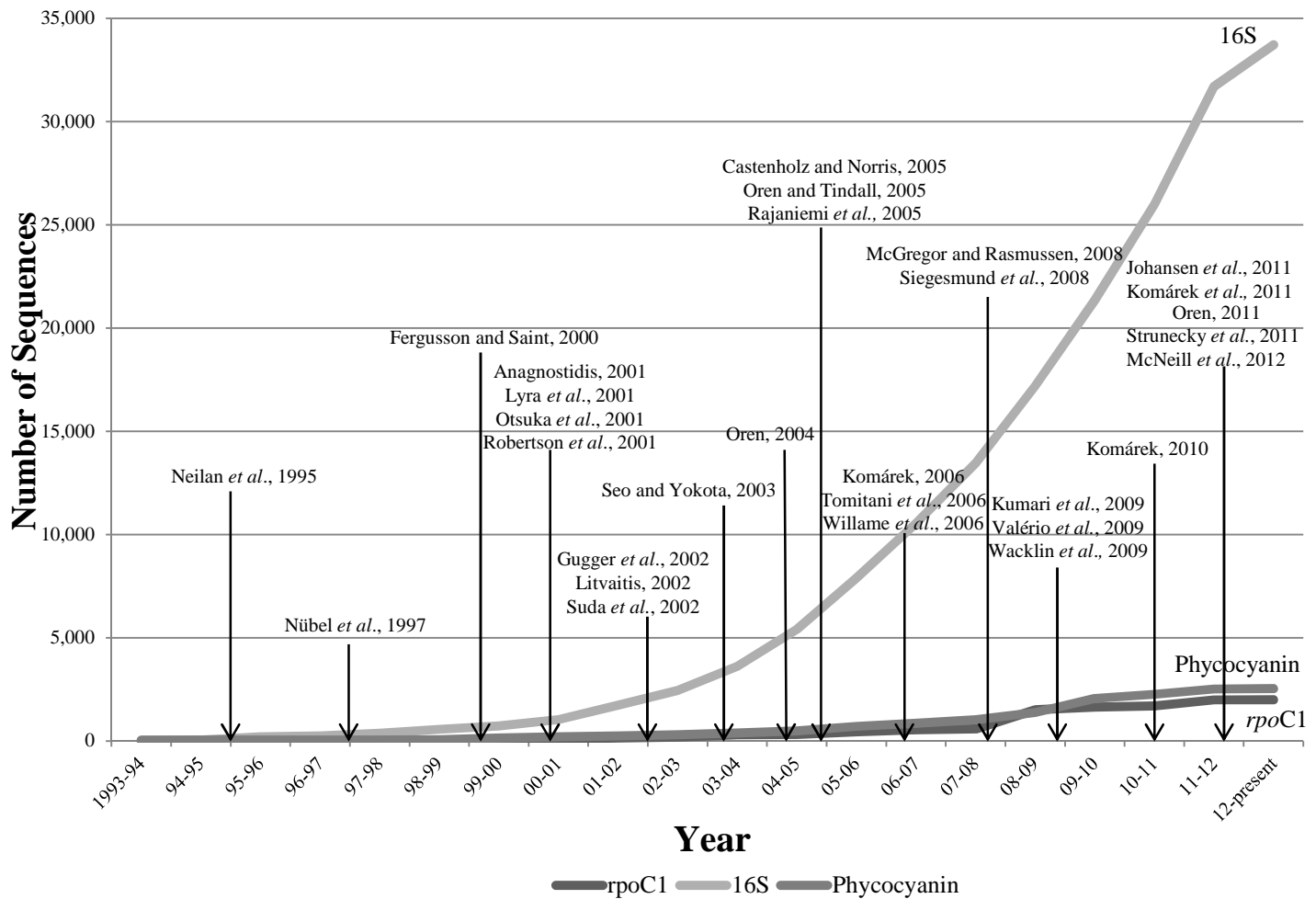**Fig. 1** – Number of sequences within GenBank for the 16S, *rpo*C1 and phycocyanin loci over time. Data obtained from NCBI Nucleotide Database (http://www.ncbi.nlm.nih.gov/nuccore). References indicate papers characterising or reporting changes to cyanobacteria taxonomy and nomenclature.
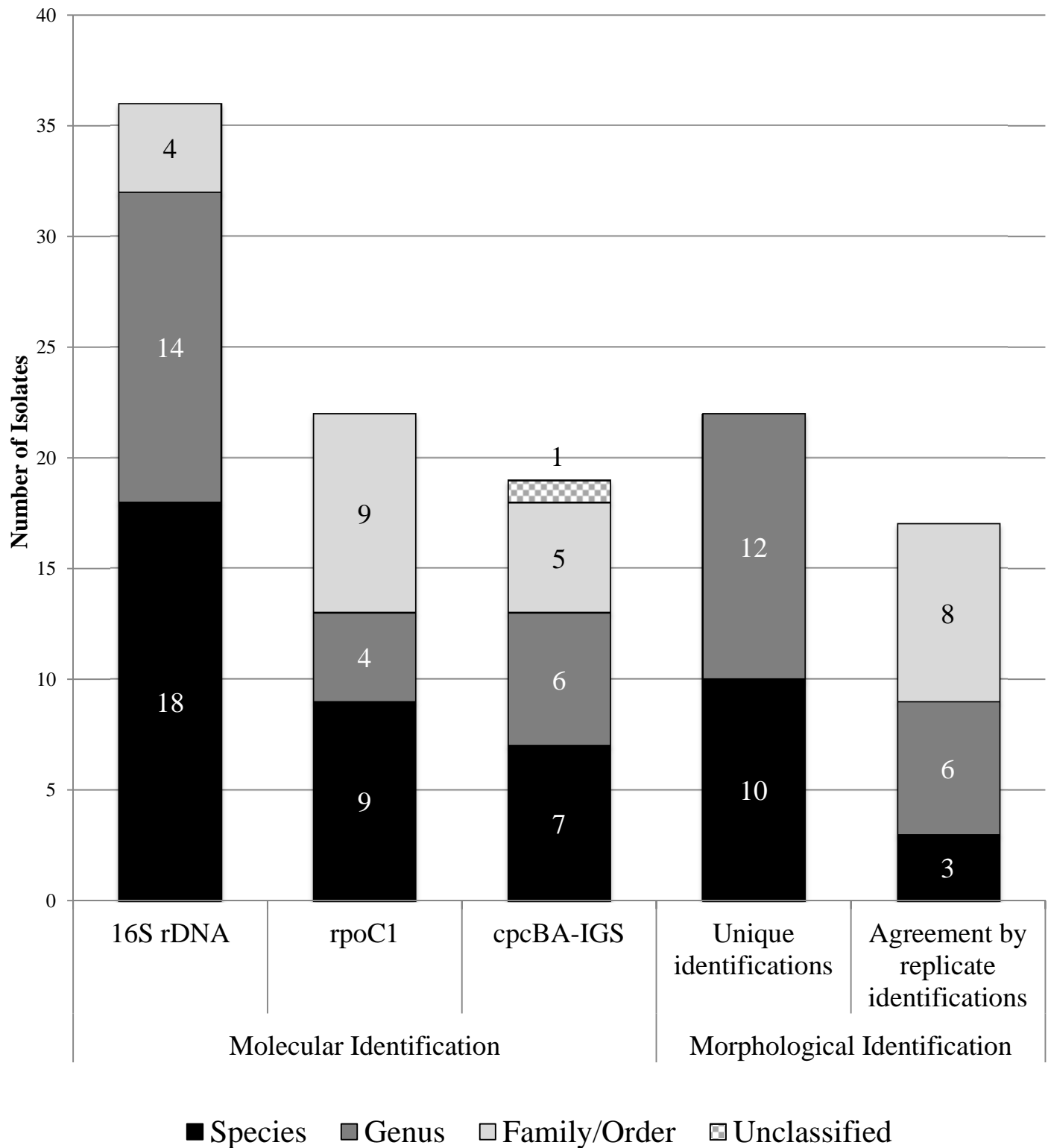
Fig. 2 – Number of cyanobacterial isolates identified at order, genus or species-level, using molecular and morphological methods. Depending on the isolate, morphological identifications were carried out by either one (unique identifications), or two independent taxonomists (replicate identifications). Where isolates were examined in duplicate, the number of isolates providing agreement between the replicate identifications is shown.
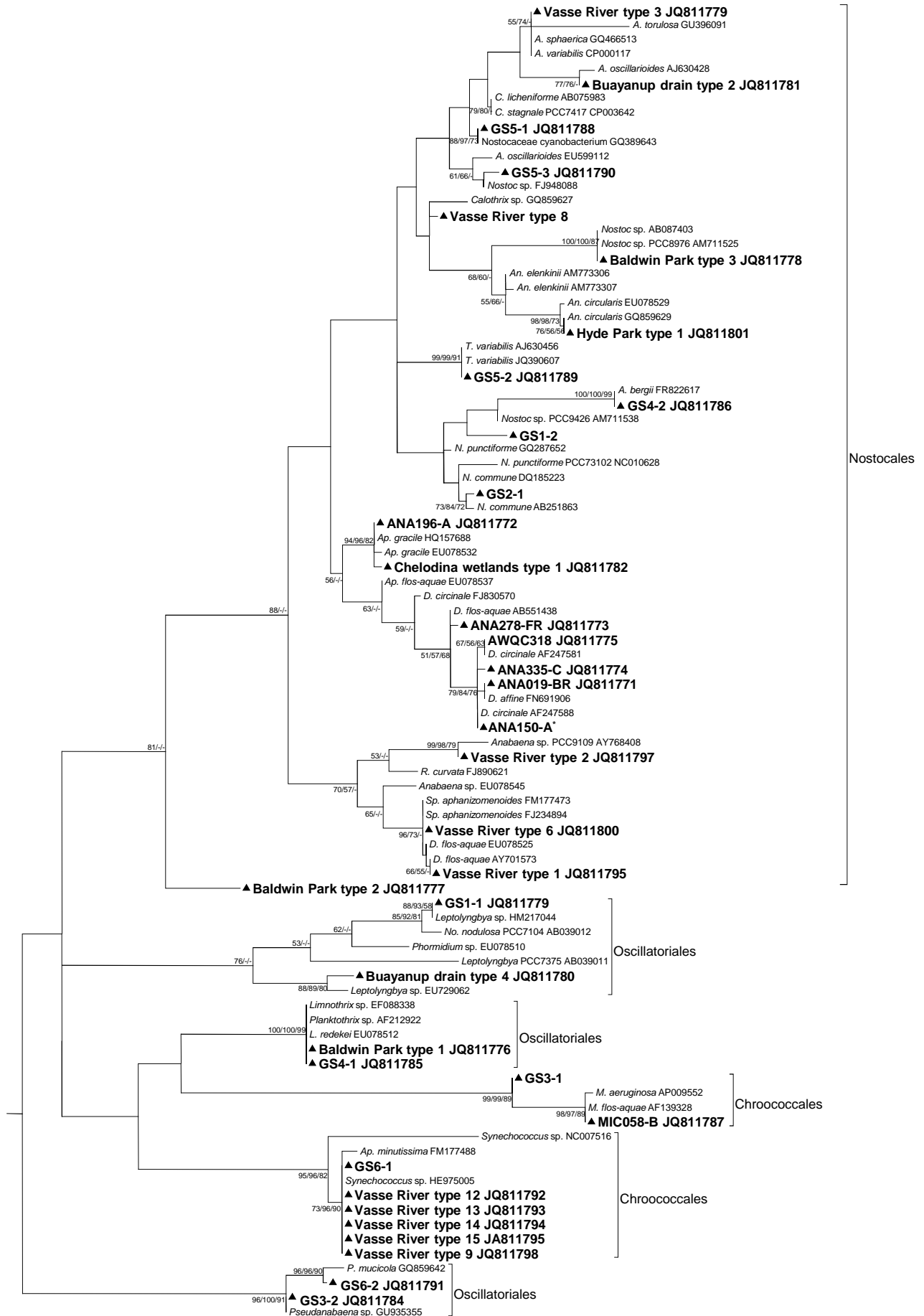
Fig. 3 – Maximum Likelihood tree based on the 16S rDNA sequences (313 bp) showing the clustering of isolates obtained. Branch support values greater than 50% for Maximum Likelihood, Maximum Parsimony, and Distance analyses respectively are indicated left of the nodes. Bar, 0.05 substitutions per site. The outgroup was removed to facilitate the visualisation of the isolates. Sequence previously submitted to GenBank with accession number AF247573
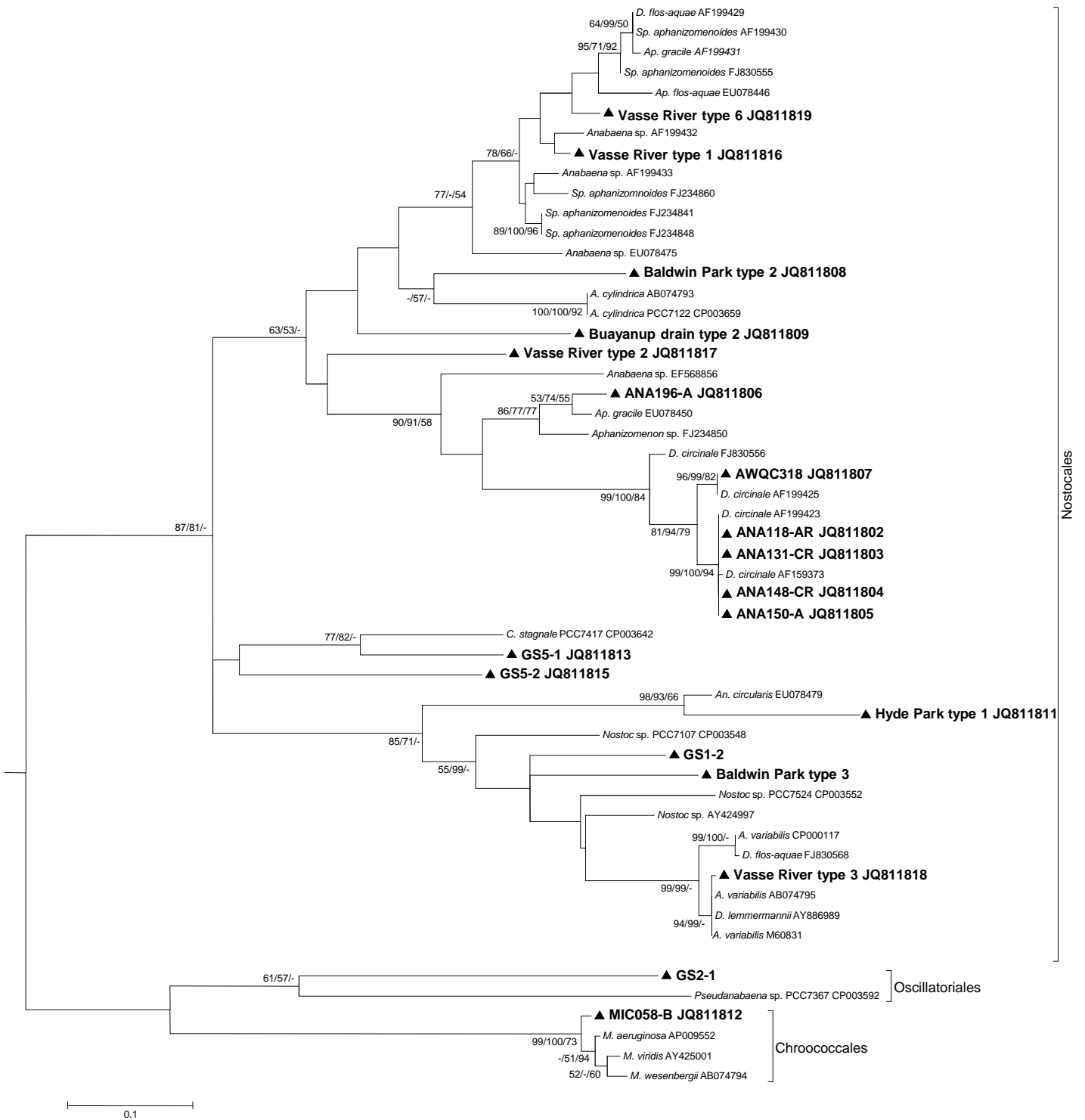
Fig. 4 – Maximum Likelihood tree based on the *rpo*C1 sequences (409 bp) showing the clustering of isolates obtained. Branch support values greater than 50% for Maximum Likelihood, Maximum Parsimony, and Distance analyses respectively are indicated left of the nodes. Bar, 0.1 substitutions per site. The *Synechococcus* cluster containing GS6-1, Vasse River types 9 and 13 were removed to facilitate visualisation of the other isolates

Fig. 5 – Maximum Likelihood tree based on the *cpc*BA-IGS sequences (423 bp) showing the clustering of isolates obtained. Branch support values greater than 50% for Maximum Likelihood, Maximum Parsimony, and Distance analyses respectively are indicated left of the nodes. Bar, 0.2 substitutions per site.
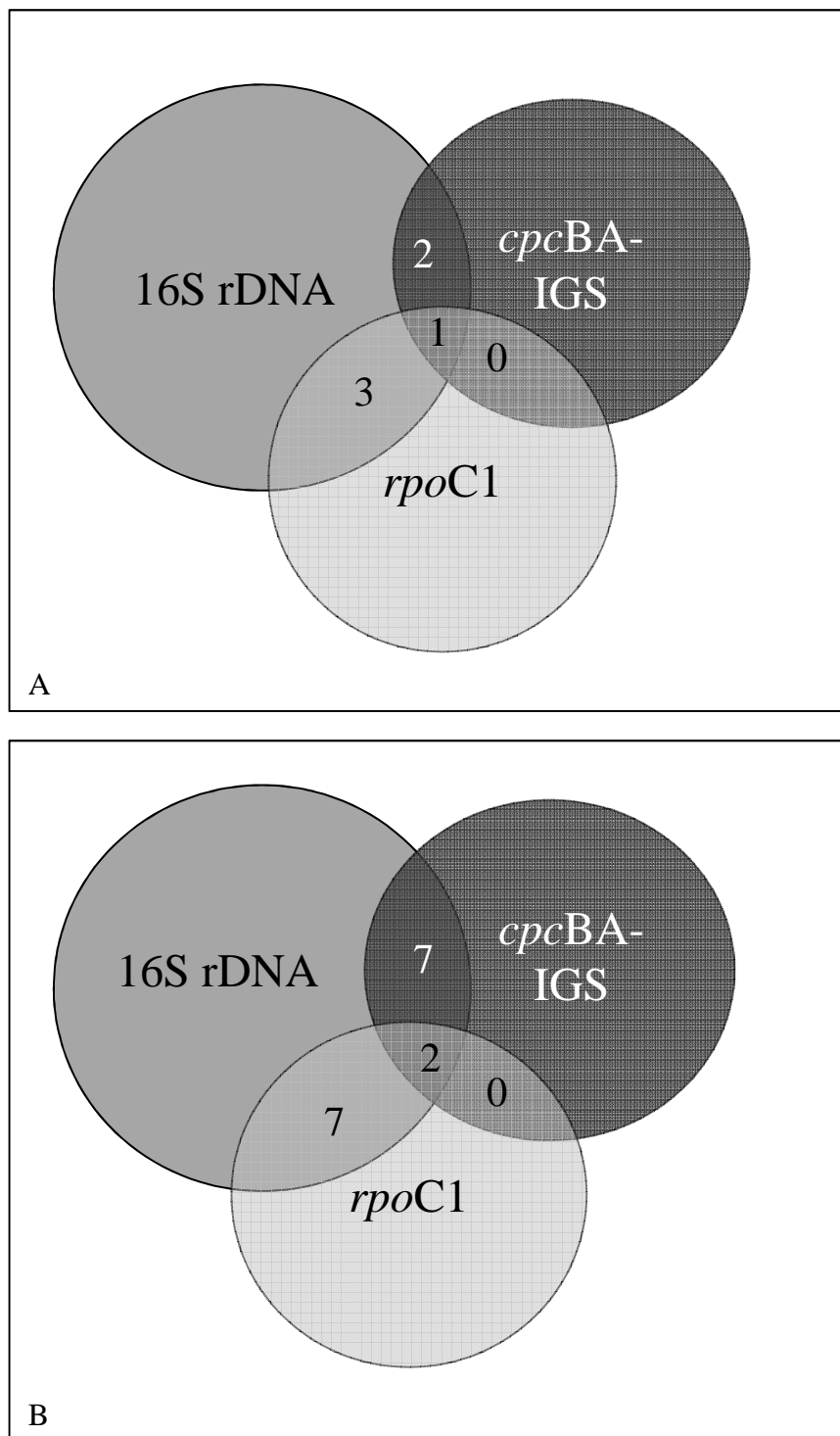
Fig. 6 – Extent of agreement between the molecular identifications, of various cyanobacterial isolates, obtained by the analysis of three loci: 16S rDNA, *rpo*C1, and *cpc*BA-IGS. The numbers indicate the isolates for which agreement between the loci was found at species level (panel A), or at least genus level (i.e. agreement at either species or genus level) (panel B).

- Molecular and morphological agreement was obtained for only 26% of isolates
- Molecular agreement at multiple loci does not necessarily occur
- Polyphasic characterization does not necessarily lead to conclusive identification
- Despite recent advances, cyanobacteria identification remains problematic
- Two potentially new species of *Nostocales* from Western Australia identified