RESEARCH ARTICLE

Cyanobacterial diversity in the hot spring, pelagic and benthic habitats of a tropical soda lake

Pawan K. Dadheech[1,2,3], Gernot Glöckner[4,5,6], Peter Casper[1], Kiplagat Kotut[7,8], Camila Junqueira Mazzoni[9,10], Susan Mbedi[9,11] & Lothar Krienitz[1]

[1]Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Stechlin-Neuglobsow, Germany; [2]Department of Microbiology, Central University of Rajasthan, Bandarsindri, Kishangargh, Rajasthan, India; [3]Department of Botany, Government College, Ajmer, Rajasthan, India; [4]Medical Faculty, Institute of Biochemistry I, University of Cologne, Cologne, Germany; [5]Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany; [6]FU Berlin, Berlin, Germany; [7]Department of Plant and Microbial Sciences, Kenyatta University Nairobi, Nairobi, Kenya; [8]Embu University College, Embu, Kenya; [9]Berlin Center for Genomics in Biodiversity Research, Berlin, Germany; [10]Evolutionary Genetics, Leibniz-Institute for Zoo and Wildlife Research (IZW), Berlin, Germany; and [11]Museum für Naturkunde – Leibniz-Institute for Research on Evolution and Biodiversity, Berlin, Germany

Correspondence: Lothar Krienitz, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Alte Fischerhütte 2, D-16775 Stechlin-Neuglobsow, Germany.
Tel.: 0049 33082 69926;
fax: 0049 33082 69917;
e-mail: krie@igb-berlin.de

Received 26 October 2012; revised 30 March 2013; accepted 30 March 2013.
Final version published online 25 April 2013.

DOI: 10.1111/1574-6941.12128

Editor: Riks Laanbroek

Keywords
Cyanobacteria; hot springs; Lake Bogoria; metagenomics; pyrosequencing; soda lakes.

Abstract

Hot springs and saline-alkaline lakes of East Africa are extreme habitats regarding temperature, or salinity and pH, respectively. This study examines whether divergent habitats of Lake Bogoria, Kenya, impacts cyanobacterial community structure. Samples from the hot springs, pelagic zone and sediment were analysed by light microscopy, multilocus 454-amplicons sequencing and metagenomics to compare the cyanobacterial diversity. Most of the phylogenetic lineages of *Cyanobacteria* occurred exclusively in the Bogoria hot springs suggesting a high degree of endemism. The prevalent phylotypes were mainly members of the *Oscillatoriales* (*Leptolyngbya*, *Spirulina*, *Oscillatoria*-like and *Planktothricoides*). The *Chroococcales* were represented by different clades of *Synechococcus* but not a single phylotype clustered with any of the lineages described earlier from different continents. In contrast, we found that the pelagic zone and the sediments were inhabited by only a few taxa, dominated by *Arthrospira* and *Anabaenopsis*. *Arthrospira*, the main food base of Lesser Flamingo, was detected in all three habitats by amplicons pyrosequencing, indicating its resilience and key role as a primary producer. Despite the close connection between the three habitats studied, the cyanobacterial communities in the hot springs and lake differed considerably, suggesting that they are unable to adapt to the extreme conditions of the neighbouring habitat.

Introduction

Cyanobacteria are an ancient group of photosynthetic prokaryotic organisms (Wood *et al.*, 2008). They have existed on earth for about 2.8 billion years (Olson, 2006). Because of their high adaptability, they are an integral part of many terrestrial and aquatic ecosystems and have successfully established themselves even in extreme habitats such as saline lakes and thermal springs (Hammer, 1986; Whitton & Potts, 2000).

Lake Bogoria, a saline-alkaline lake located in the Gregory Rift Valley of Eastern Africa, provides numerous habitats for *Cyanobacteria*. The shoreline of Lake Bogoria is fringed by geysers and hot springs confirming the volcanic origin of this harsh landscape (Cioni *et al.*, 1992). Hot springs are excellent examples of habitats for taxa with 'subcosmopolitan' geographical distribution (species occurring throughout the world but only in 'appropriate' habitats) (Padisák, 2009). The geographical isolation of hot spring *Cyanobacteria* has led to evolutionary divergences on both global and local scales (Castenholz, 1996; Papke *et al.*, 2003; Klatt *et al.*, 2011). A comparison of the thermal springs in USA, Japan, New Zealand and Italy has revealed the existence of cyanobacterial lineages of restricted geographical distribution (Papke *et al.*, 2003). Other studies focusing on two very different regions, the north-polar region of Greenland (Roeselers *et al.*, 2007) and the tropical region of Australia (McGregor

& Rasmussen, 2008), have provided further evidence that new cyanobacterial clades occur in hot springs. To the best of our knowledge, though, only scant African hot springs have been investigated, mainly via microscopy (Hindák, 2001; Krienitz *et al.*, 2003).

Over the last three decades, the planktonic *Cyanobacteria* of the Lake Bogoria ecosystem have been extensively studied using conventional methods. The phytoplankton community of this saline-alkaline lake is usually dominated by *Arthrospira fusiformis* (Voronichin) Komárek *et* Lund. However, the populations of *A. fusiformis* are usually replaced at irregular intervals partly by populations of the nostocalean *Anabaenopsis* sp. or by the picoplanktonic chlorophyte *Picocystis salinarum* Lewin, amongst other *Cyanobacteria* or eukaryotic algae (Harper *et al.*, 2003; Ballot *et al.*, 2005; Oduor & Schagerl, 2007a, b; Schagerl & Oduor, 2008; Krienitz & Kotut, 2010). Interestingly, a very short food chain exists in African soda lakes, with *Cyanobacteria* directly consumed by birds (Vareschi, 1978). In Lake Bogoria, the primary consumers are the Lesser Flamingos (*Phoeniconaias minor* Geoffroy Saint Hilaire) (Brown, 1959), which generally feed on the suspended primary producers in the lake water or the benthic algae of the sediments. As long as *A. fusiformis* dominates the cyanobacterial community, they form a safe food resource for hundreds of thousands of this bird species (Fig. 1a).

Sediments are habitats for benthic organisms, where they provide resting stage and refugia. Organisms can be recruited or re-suspended from sediments to the water body (Dobson & Frid, 2009). Further, sediments provide a source of chemical compounds present in the lake and its catchment.

In this study, we focus on three main and closely connected habitats of Lake Bogoria: (1) the hot springs found along the shoreline; (2) the open water of the lake; and (3) the lake's sediment layer. Springs flow directly into the lake with transport of *Cyanobacteria* enhanced by the occasional heavy rainfall. In the lake, *Cyanobacteria* can sink from the pelagic to the sediments or can be re-suspended. When the water level of the lake rises as a result of heavy rains in the catchment area, parts of the hot springs area are flooded by lake water. *Cyanobacteria* may also move between the three habitats via the Lesser Flamingo, which feeds in the pelagic zone and on the sediment surface and comes regularly to the springs to drink water and wash its feathers. The habitats studied differ considerably by their structure and physico-chemical properties. The extreme conditions are water temperature (40–80 °C) in the hot springs, and salinity (> 50 ‰) and pH (> 10) in the lake. In this study, we compare the cyanobacterial diversity in the three different environments. To enable us to carry out an exhaustive characterization of the diversity of the complex hot spring/lake/sediment system of Lake Bogoria, we used classical light microscopy, 16S rRNA gene pyrosequencing and metagenomic analysis of environmental samples. The parallel use of these different methods made it possible to estimate limitations and/or advantages of each technique used.

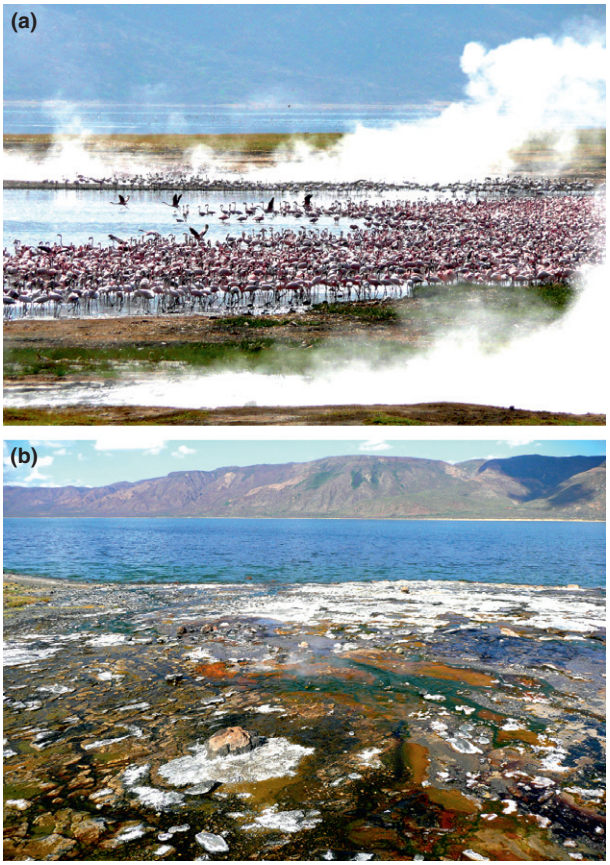

Fig. 1. Lake Bogoria, Kenya. (a) Flamingos at the lake shore, collecting food near the outlets of steaming hot springs. (b) The Chemurkeu hot spring area, colonized by cyanobacterial biofilms; the springs divert the water directly into the lake.

Materials and methods

Sampling

Lake Bogoria National Reserve, a Ramsar site, is located in an arid to semi-arid region (Mugo, 2007). Lake Bogoria is located in the centre of the reserve. Its endorheic saline-alkaline basin has a surface area of about 34 km² and a maximum depth of about 10 m. Detailed characteristics of the lake and its catchment are provided by Harper *et al.* (2003) and Jirsa *et al.* (2012). Due to erosion from the alkaline volcanic rocks of the catchment area, sodium and hydrogen carbonates are the dominant

salts in the waters of Lake Bogora National Reserve. While the lake is meso- to hypersaline (30–55 ‰), the hot springs are slightly hyposaline (3–4 ‰) (Cioni *et al.*, 1992; Owen *et al.*, 2004; Krienitz *et al.*, 2013). Our sampling site is located in the Chemurkeu area at the western shore of Lake Bogoria (geographical coordinates 0° 13′ 33″ N, 36° 05′ 41″ E) (Fig. 1b). At this site, the hot spring waters drain directly into the lake. Salinity, pH and temperature were measured using a WTW Multiline P4 meter (Wissenschaftlich Technische Werkstätten, Weilheim, Germany). Microscopic analyses of the *Cyanobacteria* in the hot springs, the lake and its sediments begun in 2001, and sampling has been carried out twice yearly. This investigation was carried out in the frame of a long-term monitoring project focusing on the interaction of *Cyanobacteria* from the three compartments and the feeding behaviour of Lesser Flamingos. In the present study, samples were collected from the water column, lake sediment and the hot springs on 18 July 2008 and on 19 January 2011.

Water samples were scooped with a 1-L plastic ladle from a few centimetres below the surface. For microscopic analyses, 1 L of the water was passed through a plankton net with a mesh size aperture of 25 μm to concentrate the sample to a volume of 10 mL and fixed with formaldehyde (final concentration 2.5%). Sample preparation for molecular analyses in the field involved the filtration of 10 fresh subsamples of 100 mL each through glass fibre filters (Whatman GF/C, Whatman International Ltd, Maidstone, UK). The subsamples from all the sites were combined and air-dried onsite using a polycarbonate filter holder SM 16510 (Sartorius AG, Göttingen, Germany) and a manual vacuum pump VacuMan (Bürkle GmbH, Bad Bellingen, Germany) and thereafter refrigerated.

Sediment samples were collected *c.* 20 m from the shoreline, using a plastic liner. The tube was forced manually into the sediment and stoppered at the top and, while still under water, a piston was pushed in from the bottom (Dadheech *et al.*, 2009). On land, the piston was pressed towards the top, hence pushing the sediment core into a smaller tube held by hand to the top of the sediment sampler. From each of three sediment cores, the top 2.5-cm-thick layer was collected. This section was air-dried on site within 24 h under dry tropical conditions. Afterwards, the predried sediments were dried in an oven at 60 °C until weight was constant and stored in the dark.

Cyanobacterial biofilms were scrapped with a spatula from the bottom of hot spring rivulets, while microbial mats were collected from sites with average temperatures of 40 and 80 °C, respectively. Subsamples of 10 mL fresh cyanobacterial material were collected from 10 different rivulets. From each sample, 5 mL was set aside and stored for further individual sample examination. The remaining 5 mL of each subsample from all the sites was combined and examined to establish a broad overview of the species composition. The pooled sample was taken as the raw material for molecular analyses. All samples were air-dried onsite and thereafter stored in a refrigerator. Fresh samples of hot spring *Cyanobacteria* suspended in 10 mL original water were fixed with formaldehyde (final concentration 2.5%) for microscopic examination.

Light microscopy

Cyanobacterial populations were studied under a Nikon Eclipse E6000 light microscope with differential interference contrast (Nikon Corporation, Tokyo, Japan). Microphotographs were taken using a Nikon digital camera DS-Fi1 and Nikon software NIS-Elements D. Two handbooks on *Chroococcales* and *Oscillatoriales* (Komárek & Anagnostidis, 1998, 2005), and the publication of Hindák (2001) were used for the determination of the cyanobacterial taxa.

Genomic DNA extraction and 454 sequencing of 16S rRNA gene amplicons

Cyanobacterial cell lysis and DNA extraction from samples were performed using Ultraclean Soil DNA Isolation Kit (MoBio Inc., Solana Beach, CA) following the manufacturer's instructions and the procedure described by Dadheech *et al.* (2009). PCR amplification was carried out using the Taq PCR Core Kit (Qiagen GmbH, Hilden, Germany) in Peltier Thermal Cycler PTC 200 (MJ Research Inc., San Francisco, CA). A nested PCR approach was used for the amplification of 16S rRNA gene fragments from the cyanobacterial genomic DNA present in samples. This technique has greater specificity than regular PCR (Zwart *et al.*, 2005). A low number of PCR cycles (20) were used to avoid biased amplification of the 16S rRNA gene fragment. For the first PCR, the primers CYA106F (Nübel *et al.*, 1997; Li *et al.*, 2001) and R4R (Li *et al.*, 2001) were used to selectively amplify long fragments of cyanobacterial 16S rRNA genes from the samples as described previously (Ballot *et al.*, 2004a). For the second PCR, tagged 454-pyrosequencing fusion primers, forward CYA359F and reverse primer CYA781R (Nübel *et al.*, 1997) were used. The following PCR protocol was employed for amplification of the V3-V4 region of the 16S rRNA gene: an initial denaturation at 94 °C for 3 min, followed by 19 cycles at 94 °C for 1 min, 55 °C for 45s and 72 °C for 1 min, followed by a final extension of 5 min at 72 °C. Similarly, tagged fusion primers, cpc_arF (forward) and cpc_arR (reverse), and an amplification protocol (Ballot *et al.*, 2004a) were used for PCR amplification of beta and alpha subunits including intergenic spacer (*cpc*BA-IGS) of *Arthrospira* inhabiting the hot

springs of Lake Bogoria. The primers employed in this study for amplification of phycocyanin operon were designed specifically for *Arthrospira* (Ballot *et al.*, 2004a) and have been used to amplify this locus of *Arthrospira* strains originated from different continents (Premanandh *et al.*, 2006; Dadheech *et al.*, 2010). The cpc primers were specific for *Arthrospira* sequences as later inspection of the amplification products showed. All amplified products were purified using Qiaquick PCR purification columns (Qiagen GmbH).

Amplicons sequencing with Roche 454 FLX Genome Sequencer

Amplicons were ligated to GS Rapid Library MID Adaptors (Roche Diagnostics GmbH, Mannheim, Germany) according to the Rapid Library Preparation Method Manual (Roche Diagnostics GmbH) starting with point 3.4. The adaptor-ligated amplicons were pooled equally and sequenced with the GS Titanium XLR70t-chemistry according to the manufacturer's protocols.

Phylogenetic analyses

Prior to analysis of sequences obtained from 454 pyrosequencing of amplicons, barcodes and primer sites were removed using cutadapt (Martin, 2011). The sequences that were shorter than 200 bp in length, reads containing any unresolved nucleotides and chimeric sequences checked by Bellerophon (Huber *et al.*, 2004) were removed from the pyrosequencing-derived data sets. Only sequences with a length of > 235 bp were used for further analysis. The online program SeqMatch (available at website of Ribosomal Database Project 10; RDP, http://rdp.cme.msu.edu) was used taking into consideration the following options for bacteria (type and nontype, uncultured and cultured, less or more than 1200 bp good quality sequence and KNN 1: no of matches displayed per sequence, also number used to classify queries by unanimous vote). This approach was used to identify major lineages present in a particular habitat. Further, percentage similarity of each sequence to available sequences in GenBank was determined using the Blast 2Go program (Conesa *et al.*, 2005). We calculated the phylogenetic relationship within phylotypes of a particular genus, relatedness and divergence from nearest sequences available in GenBank. To this end, 454 sequences belonging to a cyanobacterial genus according to blastn searches and the most closely related sequences (measured in percentage identity) retrieved from NCBI database were initially aligned online using nast (DeSantis *et al.*, 2006). The alignment was later refined using the software muscle (Edgar, 2004) in-built in mega 5 (Tamura *et al.*, 2011). Ambiguous regions were checked with program Gblocks (Talavera & Castresana, 2007) with subsequent manual refinement of the alignment. By removing unaligned and ambiguous portions of the sequences, most of the sequences showed 93–99% similarity in blastn results. As we obtained good quality fragments of variable length, only the first 235–250 bp mainly representing the V3 region of 16S rRNA gene were used for the phylogenetic analysis. The presented trees were calculated using the maximum-likelihood (ML) algorithm in the program mega 5. The selection of best-fit model for a phylogenetic analysis was determined using jModelTest 2 tool (Darriba *et al.*, 2012) considering Bayesian information criteria (BIC). The number of nucleotide positions and the model employed for phylogenetic analysis are mentioned in legend of a particular tree. Bootstrap analysis was performed to access stability/support of nodes with 1000 re-sampled data sets for ML analysis.

Metagenomics

The DNA extracted from each sample as described above was converted into a 454-sequencing library according to the manufacturers protocols, and each sample individually tagged with a MID sequence for later identification. Sequencing was performed on a 454/Roche Titanium sequencing machine (Roche Biosciences, Germany) using 1/8 of a run for all three samples. The individual tags were separated using the implemented software and read sequences extracted.

The raw sequences were assembled with the Newbler software (Roche Biosciences) using standard settings. blast on contigs and raw reads was carried out locally against the microbial RefSeq library of GenBank. The results were loaded in the megan software tool (Huson *et al.*, 2007) and analysed distribution of different taxa in the habitats. Sequences with no database hits were excluded from further analysis.

Results

The three distinct sampling sites investigated exhibited contrasting physico-chemical characteristics with a higher salinity and pH in the lake as compared to the hot spring site (Table 1).

Microscopic examinations revealed higher cyanobacterial diversity in the hot springs (Fig. 2c–g) as compared to the lake water and the sediment surface with an absolute dominance by *A. fusiformis* and *Anabaenopsis* sp. (Fig. 2a and b). However, the species delineation in hot springs was very difficult due to a limited number of phenotypic characters and absence of clear cut differences between the various phenotypes. Hence, only the genus

could be defined and the dominant taxa were listed (Supporting Information, Table S1).

Assessment of cyanobacterial diversity using 16S rRNA gene pyrosequencing

Overall, we recovered 460 sequences of cyanobacterial origin in our environmental samples (Table S2). We observed the highest diversity of cyanobacterial phylotypes in the hot springs with 244 sequences. The order *Oscillatoriales* dominated with a contribution of 81% of total sequences from the library. *Leptolyngbya* was by far the dominant taxon, comprising 82% of the recovered Oscillatoriales sequences, followed by *Spirulina* (8%), *Oscillatoria*-like (6%) and *Planktothrikoides* (4%). The order Chroococcales was represented by the genus *Synechococcus* with only 4% of the sequences. In BLASTN searches, some sequences registered some similarity to *Xenococcus* (three sequences) and *Anabaena* (two sequences; Table S2). A further 13% of the sequences could only be assigned to uncultured *Cyanobacteria*/uncultured bacteria/uncultured organism with < 92% sequence similarity in BLASTN hits. As taxonomic placement could not be determined, those sequences were designated as unidentified. We did not

Table 1. Physico-chemistry of the study areas

Sampling site	pH	Conductivity ($\mu S\ cm^{-1}$)	Salinity (‰)
Lake Bogoria	9.7–10.7	58 300–71 100	39.2–43.0
Hot Springs Bogoria	8.9–9.5	6 500–7100	3.5–3.9

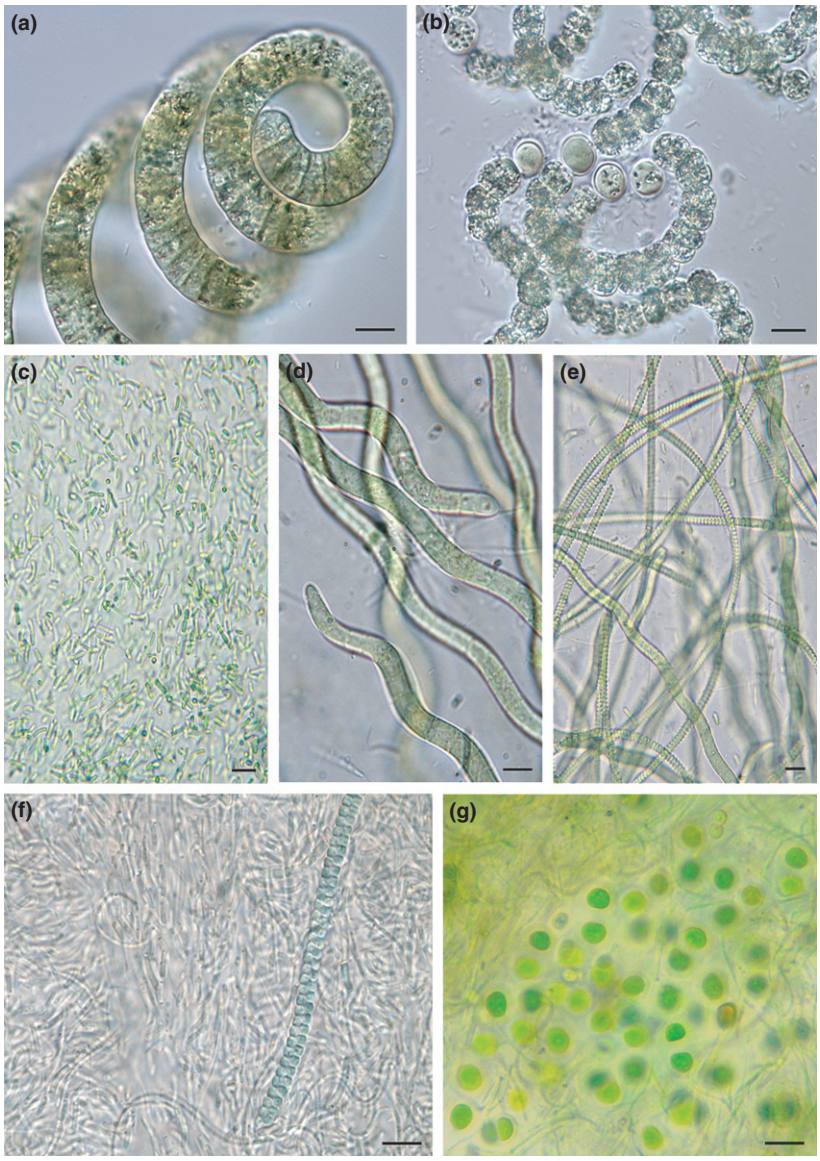

Fig. 2. Dominant morphotypes of *Cyanobacteria* from connected habitats of Lake Bogoria. (a, b) From pelagic habitat. (c–g) From hot spring habitats. (a) *Arthrospira fusiformis*. (b) *Anabaenopsis*. (c) *Synechococcus*. (d) *Arthrospira*-like undulating filaments. (e) *Spirulina* and *Oscillatoria*-like. (f) *Leptolyngbya*, *Spirulina*. (g) *Xenococcus* intermingled in *Leptolyngbya* filaments. Scale bar: 10 μm.

detect 16S rRNA gene sequences of *Arthrospira* species in the hot spring sample. However, as metagenomics data (see below) strongly indicated the presence of this genus, we sequenced amplicons of the phycocyanin locus (*cpc*BA-IGS) using primers of Ballot *et al.* (2004a). The raw sequences had 95–97% similarity in BLASTN searches to different strains of *Arthrospira,* while similarity was increased to 97–100% after refining (by removing ambiguous sections) the sequences.

The 16S rRNA gene sequences derived from hot springs of Lake Bogoria and outgroups from GenBank were used to reconstruct phylogenies. This approach yielded phylogenies with mostly unambiguously resolved branches for each tree. Several sequences in each tree clustered together to form Lake Bogoria hot spring-specific clades (Fig. 3a–c and Figs S2–S8), while some sequences clustered more with phylotypes from other hot springs (Fig. 3d and Figs S4, S5, S8, S9). Thus, endemic and cosmopolitan species seem to coexist in the Lake Bogoria hot springs.

The sequence library of the pelagic zone was made up of 148 cyanobacterial sequences. The library of 16S rRNA gene amplicons mostly comprised of phylotypes related to members of *Oscillatoriales* and *Nostocales*. The nostocalean sequences accounted for 36% of the total sequences and matched 95–99% to *Anabaenopsis abijatae*, *Anabaenopsis* cf *abijatae* and *Cyanospira rippkae*. All corrected sequences allocated to *Oscillatoriales* shared 96–100% similarity to *Arthrospira*. The sequences belonging to *Arthrospira* covered 60% of the sequences from the pelagic region. There were also some undefined sequences (4%) that could not be assigned to a cyanobacterial taxon in BLASTN analysis.

The 16S rRNA gene amplicons from the sediment of Lake Bogoria yielded 68 cyanobacterial sequences. BLASTN analysis revealed that the sediment sequences were from the Oscillatoriales (66%) and Nostocales (34%). The phylogenetic analysis revealed that only two genera, *Arthrospira* and *Anabaenopsis*, were recovered in the sediment, while in the plankton, some additional cyanobacterial phylotypes whose identity could not be delineated were present in low numbers (Fig. 4).

A ML phylogenetic tree yielded 15 well-defined clades together with some singular sequences representing possibly minor clades (Fig. 4). Interestingly, we often found sequences generated from different sampling points within some clades, while others (e.g. *Synechococcus*, *Spirulina*, *Planktothricoides* and *Xenococcus*) were restricted to the hot springs only.

Metagenomic analysis

Extracted DNA from each site was submitted to a metagenomics analysis (Table S3). An assembly of the raw sequences mainly yielded contigs presumably belonging to only the most abundant taxa in each habitat. Most of the contigs were derived from taxa belonging to *Oscillatoriales* followed by *Nostocales* (data not shown). Interestingly, frequently sequences derived from different habitats assembled in the same contigs, indicating very close relationships in the different habitats. Thus, as indicated by the microscopic observations and the 16S rRNA gene sequencing, only members of these two orders appeared to dominate in all three habitats.

Each raw sequence read was then submitted to a BLAST analysis against the full RefSeq protein library of NCBI (version Dec 2011). While the bulk of our sequences were cyanobacterial, the hot spring habitat also yielded a considerable number of other extremophilic prokaryotes (Fig. S10). It was only in this habitat that we found sequences similar to *Chloroflexi,* underlining the partly extreme conditions in this specific habitat. We also observed that the cyanobacterial taxa were not evenly distributed among the habitats investigated. The hot springs seemed to hold a considerable number of *Chroococcales* (Fig. 5a). The *Oscillatoriales* was mainly represented by the genus *Arthrospira*. The hot spring and pelagic habitats exhibited additional specific hits to other *Oscillatoriales* (e.g. *Microcoleus*, *Lyngbya*, *Oscillatoria*, and *Trichodesmium,* Fig. 5b). Bacteria of the phylum *Planctomycetes* were only found in the pelagic habitat (RL2).

Discussion

Microscopic examination identified numerous *Cyanobacteria* (Table S1), while our 16S rRNA gene 454-sequencing data mainly confirmed the microscopic observations. However, these sequences can easily be assigned to phylotypes and thus yield a differentiated picture of each habitat.

The hot spring habitat is mainly dominated by *Leptolyngbya* as evidenced by microscopy and sequencing (Fig. 4), although this taxon as currently described is polyphyletic and in need of revision (Albertano & Kovácik, 1994; Nelissen *et al.*, 1994; Turner, 1997; Ishida *et al.*, 2001; Casamatta *et al.*, 2005; Komárek & Anagnostidis, 2005; Taton *et al.*, 2006; Johansen *et al.*, 2008; Bohunická *et al.*, 2011). Sequences from hot springs around the world have been intensively sampled and studied using molecular methods such as DGGE and cloning (Papke *et al.*, 2003; Lau *et al.*, 2005; Ward *et al.*, 2006; McGregor & Rasmussen, 2008; Sompong *et al.*, 2008; Oren *et al.*, 2009; Ionescu *et al.*, 2010). Hence, it is possible to compare the cyanobacterial diversity of the Bogoria hot springs with those from other regions using 16S rRNA gene amplicons pyrosequencing. Several authors have hypothesized that geographical isolation of

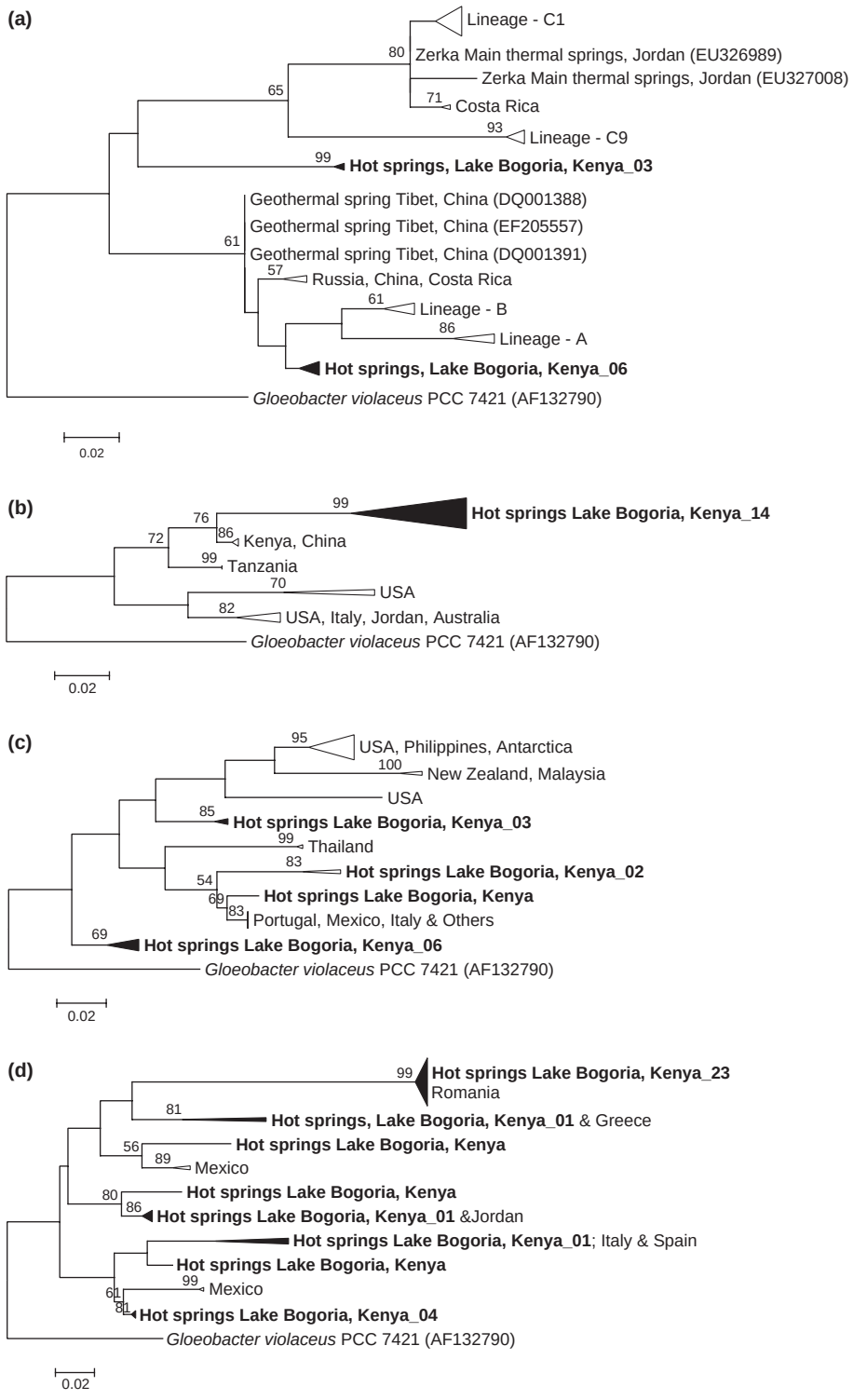

Fig. 3. Maximum-likelihood (ML) phylogenetic trees (condensed) based on 16S rRNA gene sequences (~240 bp) indicating obtained sequences in the present study from hot springs, together with a number of sequences closest in BLASTN analysis from different geographical regions. The numbers of sequences obtained in this study are indicated by a suffix. The tree was reconstructed using maximum-likelihood analysis and determined by Kimura 2-parameter + G model. The topology of the tree was tested with bootstrap method using 1000 replicates. Bootstrap values > 50% are given at nodes. (a) *Synechococcus* related sequences. (b) *Spirulina* related sequences. (c) *Oscillatoria*-like sequences that do not match to *Spirulina*, *Leptolyngbya* and *Planktothrikoides* sequences. (d) Unidentified sequences.

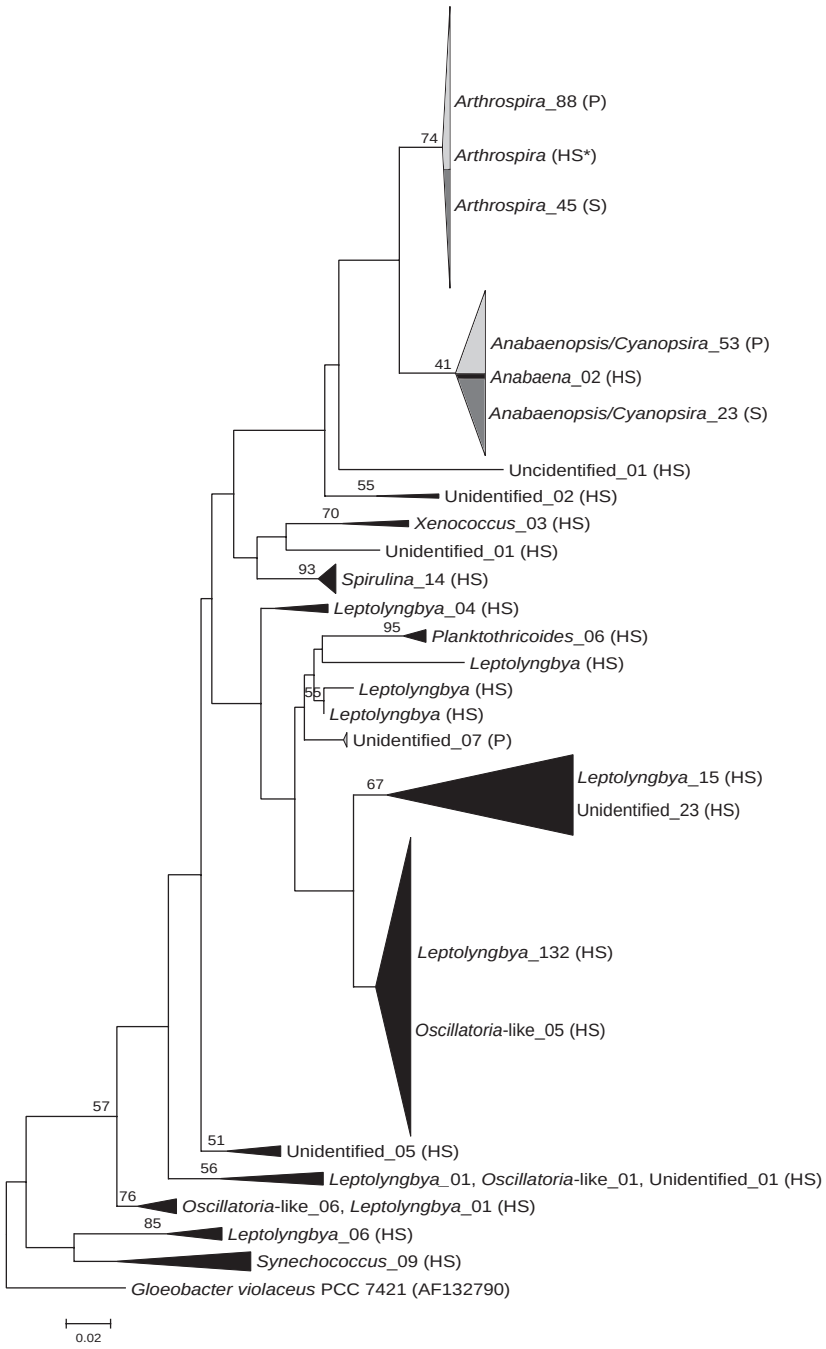

Fig. 4. Maximum-likelihood (ML) tree of partial 16S rRNA gene (~240 bp) showing phylogenetic relationship among all obtained cyanobacterial sequences in this study from different habitats (hot springs, pelagic zone and sediment) of Lake Bogoria. The numbers of sequences are given after the designation of the best matching cyanobacterial taxon in BLASTN analysis. The tree was determined by Tamura 3-parameter model, and topology was tested with bootstrap method using 1000 replicates. Bootstrap values > 50% are given at nodes. HS, hot springs; P, pelagic zone; S, sediment; HS*, detected with *cpc*BA-IGS primers; Colour code: black, hot springs; light grey, pelagic; dark grey, sediment.

hot springs leads to occurrence of indigenous thermophilic *Cyanobacteria* (Castenholz, 1996; Papke *et al.*, 2003; Miller *et al.*, 2007; Finsinger *et al.*, 2008). Our phylogenetic analysis showed that most of cyanobacterial communities in Bogoria hot springs were phylogenetically distinct from those of the thermal/hot springs of other continents. This is most pronounced in *Synechococcus* sp. where C1, C9 and A/B lineages have been described

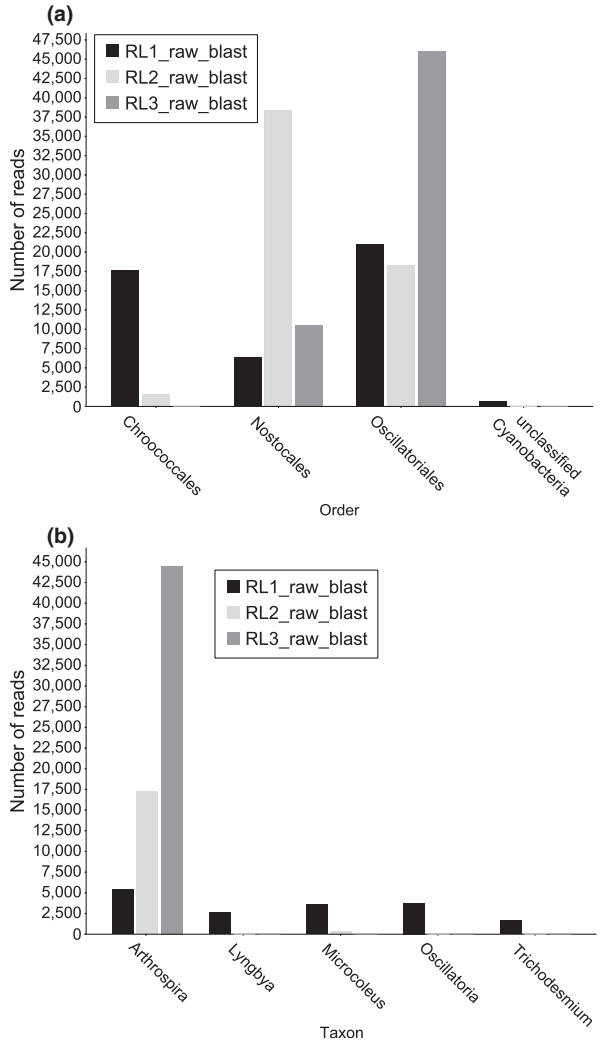

Fig. 5. Cyanobacterial clades. Data were obtained by performing BLAST analysis with all raw reads against the NCBI microbial RefSeq database. To adjust for different sampling amount, data were normalized to 100 000. The MEGAN LCA values were set to 50 for the minimum number of reads to support a clade, to 100 for minimum BLAST score, and only hits within 3% of the best hit were allowed to be counted. (a) Main cyanobacterial clades observed. (b) Distribution of genera in the *Oscillatoriales*.

(Papke *et al.*, 2003). The C9 lineage has been recorded in most sampling sites, with the exception of northern Italy, while the A/B cluster only occurs in North America (Papke *et al.*, 2003; Ionescu *et al.*, 2010). It is noteworthy that not a single phylotype of *Synechococcus* from Bogoria hot springs clusters with any other lineages described earlier. Our observations therefore confirm a considerable degree of endemism/speciation of most of the species inhabiting the hot springs at Lake Bogoria.

On the other hand, a number of phylotypes belonging to *Leptolyngbya*, *Oscillatoria*-like and unidentified/uncultured *Cyanobacteria* showed some phylogenetic relationship to phylotypes found in other biogeographical regions. This confirms the presence of cosmopolitan species in the Bogoria springs. Endemic taxa can evolve when the rate of evolution is faster than the rate of dispersal, and in contrast, cosmopolitan taxa evolve when the rate of evolution is slower than the dispersal rate (Padisák, 2009). Local speciation may occur depending on the specific biogeography of a habitat, such as some clusters of *Fischerella* due to selection of stable, ecologically restricted genotypes (Finsinger *et al.*, 2008). Many explanations have been put forward to explain the selection of ecologically restricted genotypes. The nature of the substrate, climatic and geographical conditions influences microbial community structure (Ragon *et al.*, 2012). The theory of Baas-Becking, later clarified by Wit & Bouvier (2006), stated that 'everything is everywhere but nature selects' emphasizing the global distribution of phylotypes. Another theory based on concepts of genetic speciation suggested that strong chemical/physical boundaries prevent gene outflow and resulted in genetic lineages which diverge from their relatives (Wright, 1931, 1943). Ionescu *et al.* (2010) argued that the nonuniform biogeographical patterns in a complex cyanobacterial community, in which taxa similar to other thermal environments in the world and endemic species occurring in a single sampling site of a hot spring in Jordan (the Zerka Ma'in) can only be explained by a combination of the two, principally contradicting theories mentioned above. The isolated and complex habitats of the hot springs at Lake Bogoria system is now a further example of a habitat harbouring endemic and cosmopolitan species alike. Variation in the rates of dispersal as compared to the rates of evolution depends on the geographical isolation of a given area and the dynamics of the local cyanobacterial communities (Ionescu *et al.*, 2010).

Sometimes amplicon sequencing may fail for unknown reasons. In our case, 16S rRNA gene sequencing failed to register the presence of *Arthrospira* in the hot springs, contrary to the findings of the metagenomics data. Amplicon sequences from the phycocyanin operon (*cpc*BA-IGS) supported the metagenomic analysis and provided the final proof for the presence of *Arthrospira* in hot springs. Therefore, we suggest that amplicons sequencing should be performed on different targets for each habitat for the detection of hidden taxa.

The genera *Arthrospira* and *Anabaenopsis* are the dominant taxa in the pelagic zone and sediments. The lack of the additional unidentified/uncultured *Cyanobacteria* from the pelagic zone in sediments might be due to a rapid decay of those taxa at the lake bottom. The pelagic habitat of Lake Bogoria is a hypersaline environment with high pH (Table 1) and poor in cyanobacterial diversity. It has been demonstrated that salinity reduces productivity (Srivastava *et al.*, 2005) and phylogenetic diversity (Wang *et al.*, 2011) of some *Cyanobacteria*.

We were interested in elucidating whether the *Cyanobacteria* can survive in different environments of the connected aquatic habitats of Lake Bogoria ecosystem by employing 16S rRNA gene sequences and metagenomic data of cyanobacterial phylotypes. We were able to demonstrate the presence of *Arthrospira* (evidenced by phycocyanin operon sequencing and metagenomic analysis) and *Anabaenopsis* (as per 16S rRNA gene sequencing) in hot springs.

Interestingly, hot spring and pelagic habitats provided contrasting extreme conditions for *Cyanobacteria*. Whereas the hot springs are characterized by temperature extremes, the pelagic habitat has extremes in its chemical composition (salinity and pH). Hot springs are species-rich habitats compared with the pelagic and sediment systems. The existence of a given cyanobacterial entity depends on its survival capability in a particular extreme condition. For example, the newly described cyanobacterium *Desertifilum tharense* (*Oscillatoriales*) from hot desert can tolerate high temperature but shows sensitivity to salinity (Dadheech *et al.*, 2012). Most of cyanobacterial taxa found in hot springs of Bogoria appear to be able to tolerate high temperatures but are susceptible to high salinity and pH. However, our findings demonstrate that some cyanobacterial taxa such as *Arthrospira* and *Anabaenopsis* have the ability to tolerate high temperatures, high salinity and pH. It has previously been shown that different species of *Arthrospira* and *Anabaenopsis* can be found in fresh and alkaline-saline waters of tropical environments, yet they form a compact cluster in 16S rRNA gene phylogeny, thus indicating a monophyletic origin (Ballot *et al.*, 2004a, 2008). Variable genetic loci (16S-23S ITS, *cpc*BA-IGS) of other *Cyanobacteria* were found to be highly conserved in isolates of *Arthrospira* from different habitats (Baurain *et al.*, 2002; Dadheech *et al.*, 2010). Our metagenomics data show that different phylotypes can belong to the same taxon even if sampled from very diverse environments. This finding points to a high plasticity of *Arthrospira* and *Anabaenopsis* taxa towards environmental conditions without genetic changes and thus perhaps exchange of these taxa between the connected habitats. Despite the very high abundance of *Synechococcus*-phenotypes in the hot springs sample, we were not able to detect its genotypes in the lake water. These coccoid *Cyanobacteria* were previously identified in the plankton of this lake (Ballot *et al.*, 2004b; Schagerl & Oduor, 2008), indicating that *Synechococcus* also has the ability to survive under both extreme conditions. Hence, their absence from our plankton samples is possibly due to a difference in sampling times.

Metagenomics can detect more taxa than genomic amplification with specific primers. In this study, we found the exclusive presence of the thermophilic bacteria *Chloroflexi* and *Deinococcus* in hot springs and Planctomycetes in the pelagic habitat. Other bacteria besides the targeted *Cyanobacteria* are found among the habitats. We found additional sequences of a taxon within *Oscillatoriales* in hot springs and pelagic habitats, which according to the metagenomics data, have a close relationship to *Microcoleus* but currently have no genomic sequence available in databases. The metagenomics approach also has its limitations. Raw reads can be affected by sequencing errors which would make proper taxon assignments difficult (Rho *et al.*, 2010). Furthermore, a taxon can only be assigned properly by metagenomic analysis if its genome is present in a database. Overall, all three methods employed yielded a similar picture for the cyanobacterial taxa distribution in the habitats investigated. However, our approach reveals the strengths and weaknesses associated with each method: Microscopy has a limited resolution, amplicons sequencing may be unreliable, and metagenomics works best only if closely related taxa are already present in the databases.

Data accessibility

The alignments used to build each phylogenetic tree as well as the phylogenetic trees obtained are available at TreeBASE (A Database of Phylogenetic Knowledge) under the study accession number 13183. The raw sequences generated and analysed for in this study have been deposited in the Short Read Archive (SRA) under accession numbers SRP014550 and SRP014552.

Acknowledgements

We thank the Government of Kenya for permission to carry out this research (No. MOEST 13/001/31 C 90, and NCST/RRI/12/1/BS/232). We thank William Kimosop, the Senior Warden of Lake Bogoria National Reserve for collaboration and providing valuable information on the flamingos. We thank Dale Casamatta, USA, for fruitful discussion and linguistic improvement. PKD was supported by a fellowship of the Leibniz-Institute of Freshwater Ecology and Inland Fisheries. PKD is grateful to Department of Higher Education, Government of Rajasthan, India, for granting him permission to carry out the research work at IGB, Germany.

Authors' contribution

P.K.D. and G.G. contributed equally to this work.

References

Albertano P & Kováčik L (1994) Is the genus *Leptolyngbya* (Cyanophyte) a homogeneous taxon? *Algol Stud* **75**: 37–51.

Ballot A, Dadheech PK & Krienitz L (2004a) Phylogenetic relationship of *Arthrospira*, *Phormidium*, and *Spirulina* strains from Kenyan and Indian waterbodies. *Algol Stud* **113**: 37–56.

Ballot A, Krienitz L, Kotut K, Wiegand C, Metcalf JS, Codd GA & Pflugmacher S (2004b) Cyanobacteria and cyanobacterial toxins in three alkaline rift valley lakes of Kenya – Lakes Bogoria, Nakuru and Elmenteita. *J Plankton Res* **26**: 925–935.

Ballot A, Krienitz L, Kotut K, Wiegand C & Pflugmacher S (2005) Cyanobacteria and cyanobacterial toxins in the alkaline crater Lakes Sonachi and Simbi, Kenya. *Harmful Algae* **4**: 139–150.

Ballot A, Dadheech PK, Haande S & Krienitz L (2008) Morphological and phylogenetic analysis of *Anabaenopsis abijatae* and *Anabaenopsis elenkinii* (Nostocales, Cyanobacteria) from tropical inland water bodies. *Microb Ecol* **55**: 608–618.

Baurain D, Renquin L, Grubisic S, Scheldeman P, Belay A & Wilmotte A (2002) Remarkable conservation of internally transcribed spacer sequences of *Arthrospira* ("*Spirulina*") (Cyanophyceae, Cyanobacteria) strains from four continents and of recent and 30-year-old dried samples from Africa. *J Phycol* **38**: 384–393.

Bohunická M, Johansen JR & Fučíková K (2011) *Tapinothrix clintonii* sp nov (Pseudanabaenaceae, Cyanobacteria), a new species at the nexus of five genera. *Fottea* **11**: 127–140.

Brown L (1959) *The Mystery of the Flamingos*. Country Life Ltd, London.

Casamatta DA, Johansen JR, Vis ML & Broadwater ST (2005) Molecular and morphological characterization of ten polar and near-polar strains within the Oscillatoriales (Cyanobacteria). *J Phycol* **41**: 421–438.

Castenholz RW (1996) Endemism and biodiversity of thermophilic cyanobacteria. *Nova Hedwigia* **112**: 33–47.

Cioni R, Fanelli G, Guidi M, Kinyariro JK & Marini L (1992) Lake Bogoria Hot-Springs (Kenya) – geochemical features and geothermal implications. *J Volcanol Geoth Res* **50**: 231–246.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M & Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.

Dadheech PK, Krienitz L, Kotut K, Ballot A & Casper P (2009) Molecular detection of uncultured cyanobacteria and aminotransferase domains for cyanotoxin production in sediments of different Kenyan lakes. *FEMS Microbiol Ecol* **68**: 340–350.

Dadheech PK, Ballot A, Casper P, Kotut K, Novelo E, Lemma B, Pröschold T & Krienitz L (2010) Phylogenetic relationship and divergence among planktonic strains of *Arthrospira* (Oscillatoriales, Cyanobacteria) of African, Asian and American origin deduced by 16S–23S ITS and phycocyanin operon sequences. *Phycologia* **49**: 361–372.

Dadheech PK, Mahmoud H, Kotut K & Krienitz L (2012) *Haloleptolyngbya alcalis* gen. et sp. nov., a new filamentous cyanobacterium from the soda lake Nakuru, Kenya. *Hydrobiologia* **691**: 269–283.

Darriba D, Taboada GL, Doallo R & Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**: 772.

DeSantis TZ, Hugenholtz P, Keller K, Brodie EL, Larsen N, Piceno YM, Phan R & Andersen GL (2006) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* **34**: W394–W399.

Dobson M & Frid C (2009) *Ecology of Aquatic Systems*. Oxford University Press, Inc., New York.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Finsinger K, Scholz I, Serrano A, Morales S, Uribe-Lorio L, Mora M, Sittenfeld A, Weckesser J & Hess WR (2008) Characterization of true-branching cyanobacteria from geothermal sites and hot springs of Costa Rica. *Environ Microbiol* **10**: 460–473.

Hammer UT (1986) *Saline Lake Ecosystems of the World*. Dr. W. Junk Publ, Dordrecht, Boston, Lancaster, 616 pp.

Harper DM, Childress RB, Harper MM *et al.* (2003) Aquatic biodiversity and saline lakes: Lake Bogoria National Reserve, Kenya. *Hydrobiologia* **500**: 259–276.

Hindák F (2001) Thermal microorganisms from a hot spring on the coast of Lake Bogoria, Kenya. *Nova Hedwigia* **123**: 77–93.

Huber T, Faulkner G & Hugenholtz P (2004) Bellerophon; a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**: 2317–2319.

Huson DH, Auch AF, Qi J & Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Res* **17**: 377–386.

Ionescu D, Hindiyeh M, Malkawi H & Oren A (2010) Biogeography of thermophilic cyanobacteria: insights from the Zerka Ma'in hot springs (Jordan). *FEMS Microbiol Ecol* **72**: 103–113.

Ishida T, Watanabe MM, Sugiyama J & Yokota A (2001) Evidence for polyphyletic origin of the members of the orders of Oscillatoriales and Pleurocapsales as determined by 16S rDNA analysis. *FEMS Microbiol Lett* **201**: 79–82.

Jirsa F, Gruber M, Stojanovic A, Omondi SO, Mader D, Körner W & Schagerl M (2012) Major and trace element geochemistry of Lake Bogoria and Lake Nakuru, during extreme draught. *Chemie Erde*, in press, http://dx.doi.org/10.1016/j.chemer.2012.09.001.

Johansen JR, Olsen CE, Lowe RL, Fučíková K & Casamatta DA (2008) *Leptolyngbya* species from selected seep walls in the Great Smoky Mountains National Park. *Algol Stud* **126**: 21–36.

Klatt CG, Wood JM, Rusch DB *et al.* (2011) Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. *ISME J* **5**: 1262–1278.

Komárek J & Anagnostidis K (1998) Cyanoprokaryota 1. *Chroococcales*, Vol. **19/1** (Ettl H, Gärtner G, Heynig H &

Mollenhauer D, eds), pp. 1–548. Spectrum Academischer Verlag, Elsevier, München.

Komárek J & Anagnostidis K (2005) Cyanoprokaryota 2. *Oscillatoriales*, Vol. **19/2** (Büdel B, Gärtner G, Krienitz L & Schagerl M, eds), pp. 1–757. Spectrum Academischer Verlag, Elsevier, München.

Krienitz L & Kotut K (2010) Fluctuating algal food populations and the occurrence of Lesser Flamingos (*Phoeniconaias minor*) in three Kenyan Rift Valley lakes. *J Phycol* **46**: 1088–1096.

Krienitz L, Ballot A, Kotut K, Wiegand C, Pütz S, Metcalf JS, Codd GA & Pflugmacher S (2003) Contribution of hot spring cyanobacteria to the mysterious deaths of Lesser Flamingos at Lake Bogoria, Kenya. *FEMS Microbiol Ecol* **43**: 141–148.

Krienitz L, Dadheech PK & Kotut K (2013) Mass developments of the cyanobacteria *Anabaenopsis* and *Cyanospira* (Nostocales) in the soda lakes of Kenya: ecological and systematic implications. *Hydrobiologia* **703**: 79–93.

Lau E, Nash CZ, Vogler DR & Cullings KW (2005) Molecular diversity of cyanobacteria inhabiting coniform structures and surrounding mat in a yellowstone hot spring. *Astrobiology* **5**: 83–92.

Li RH, Debella HJ & Carmichael WW (2001) Isolates identifiable as *Arthrospira maxima* and *Arthrospira fusiformis* (Oscillatoriales, Cyanobacteria) appear identical on the basis of a morphological study in culture and 16S rRNA gene sequences. *Phycologia* **40**: 367–371.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**: 10–13.

McGregor GB & Rasmussen JP (2008) Cyanobacterial composition of microbial mats from an Australian thermal spring: a polyphasic evaluation. *FEMS Microbiol Ecol* **63**: 23–35.

Miller SR, Castenholz RW & Pedersen D (2007) Phylogeography of the thermophilic cyanobacterium *Mastigocladus laminosus*. *Appl Environ Microbiol* **73**: 4751–4759.

Mugo K (ed.) (2007) *Lake Bogoria National Reserve, World Ramsar Site (No. 1057). Integrated Management Plan 2007–2012*. WWF-EARPO, County Council of Baringo, County Council of Koibatek, pp. 1–48.

Nelissen B, Wilmotte A, Neefs J-M & De Wachter R (1994) Phylogenetic relationships among filamentous helical cyanobacteria investigated on the basis of 16S ribosomal RNA gene sequence analysis. *Syst Appl Microbiol* **17**: 206–210.

Nübel U, Garcia-Pichel F & Muyzer G (1997) PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl Environ Microbiol* **63**: 3327–3332.

Oduor SO & Schagerl M (2007a) Temporal trends of ion contents and nutrients in three Kenyan Rift Valley saline-alkaline lakes and their influence on phytoplankton biomass. *Hydrobiologia* **584**: 59–68.

Oduor SO & Schagerl M (2007b) Phytoplankton primary productivity characteristics in response to photosynthetically active radiation in three Kenyan Rift Valley saline-alkaline lakes. *J Plankton Res* **29**: 1041–1050.

Olson JM (2006) Photosynthesis in the Archean era. *Photosyn Res* **88**: 109–117.

Oren A, Ionescu D, Hindiyeh M & Malkawi H (2009) Morphological, phylogenetic and physiological diversity of cyanobacteria in the hot springs of Zerka Ma. *BioRisk* **3**: 69–82.

Owen RB, Renaut RW, Hover VC, Ashley GM & Muasya AM (2004) Swamps, springs and diatoms: wetlands of the semi-arid Bogoria-Baringo Rift, Kenya. *Hydrobiologia* **518**: 59–78.

Padisák J (2009) The phycogeography of freshwater algae. *Encyclopedia of Inland Waters*, Vol. **I** (Likens G, ed.), pp. 219–223. Elsevier, Inc., Oxford.

Papke RT, Ramsing NB, Bateson MM & Ward DM (2003) Geographical isolation in hot spring cyanobacteria. *Environ Microbiol* **5**: 650–659.

Premanandh J, Priya B, Teneva I, Dzhambazov B, Prabaharan D & Uma L (2006) Molecular characterization of marine cyanobacteria from the Indian subcontinent deduced from sequence analysis of the phycocyanin operon (cpcB-IGS-cpcA) and 16S–23S ITS region. *J Microbiol* **44**: 607–616.

Ragon M, Fontaine MC, Moreira D & López-García P (2012) Different biogeographic patterns of prokaryotes and microbial eukaryotes in epilithic biofilms. *Mol Ecol* **21**: 3852–3868.

Rho M, Tang H & Ye Y (2010) FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* **38**: e191.

Roeselers G, Norris TB, Castenholz RW, Rysgaard S, Glud RN, Kuhl M & Muyzer G (2007) Diversity of phototrophic bacteria in microbial mats from Arctic hot springs (Greenland). *Environ Microbiol* **9**: 26–38.

Schagerl M & Oduor SO (2008) Phytoplankton community relationship to environmental variables in three Kenyan Rift Valley saline-alkaline lakes. *Mar Freshw Res* **59**: 125–136.

Sompong U, Anuntalabhochai S, Cutler RW, Castenholz RW & Peerapornpisal Y (2008) Morphological and phylogenic diversity of cyanobacterial populations in six hot springs of Thailand. *Sci Asia* **34**: 153–162.

Srivastava AK, Bhargava P & Rai LC (2005) Salinity and copper-induced oxidative damage and changes in antioxidative defense system of *Anabaena doliolum*. *World J Microbiol Biotechnol* **22**: 1291–1298.

Talavera G & Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**: 564–577.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Taton A, Grubisic S & Ertz D (2006) Polyphasic study of Antarctic cyanobacterial strains. *J Phycol* **42**: 1257–1270.

Turner S (1997) Molecular systematics of oxygenic photosynthetic bacteria. *Plant Syst Evol* **11**: 13–52.

Vareschi E (1978) Ecology of Lake Nakuru (Kenya).1. Abundance and feeding of Lesser Flamingo. *Oecologia* **32**: 11–35.

Wang J, Yang D, Zhang Y, Shen J, van der Gast C, Hahn MW & Wu Q (2011) Do patterns of bacterial diversity along

salinity gradients differ from those observed for microorganisms? *PLoS ONE* **6**: e27597.
Ward DM, Bateson MM, Ferris MJ, Kuhl M, Wieland A, Koeppel A & Cohan FM (2006) Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function. *Philos Trans R Soc B* **361**: 1997–2008.
Whitton BA & Potts M (2000) *Ecology of Cyanobacteria: Their Diversity in Time and Space*. Kluwer Academic Publishers, Dordrecht, Boston.
Wit R & Bouvier T (2006) 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environ Microbiol* **8**: 755–758.
Wood SA, Rueckert A, Cowman DA & Cary SC (2008) Source of edaphic cyanobacterial diversity in the dry valley of Eastern Antarctica. *ISME J* **2**: 308–320.
Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**: 97–159.
Wright S (1943) Isolation by distance. *Genetics* **28**: 114–138.
Zwart G, Kamst-van Agterveld MP, van der Werff-Staverman I, Hagen F, Hoogveld HL & Gons HJ (2005) Molecular characterization of cyanobacterial diversity in a shallow eutrophic lake. *Environ Microbiol* **7**: 365–377.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Phylogenetic tree based on 16S rRNA gene sequences (~240 bp) indicating all obtained sequences from environmental samples of hot springs of Lake Bogoria.
Fig. S2. Phylogenetic tree (expanded) based on 16S rRNA gene sequences (~240 bp) indicating all obtained *Synechococcus* related sequences in the present study from hot springs, together with a number of sequences closest in BLASTN analysis from different geographical regions.
Fig. S3. Phylogenetic tree (expanded) based on 16S rRNA gene sequences (~240 bp) indicating all obtained *Spirulina* related sequences in the present study from hot springs, together with a number of sequences closest in BLASTN analysis from different geographical regions.
Fig. S4. Phylogenetic tree based on 16S rRNA gene sequences (~240 bp) indicating all obtained *Leptolyngbya* related sequences in the present study from hot springs, together with a number of sequences closest in BLASTN analysis from different geographical regions.
Fig. S5. Phylogenetic tree (expanded) based on 16S rRNA gene sequences (~240 bp) indicating all obtained *Oscillatoria*-like that do not match *Spirulina*, *Leptolyngbya* and *Planktothricoides* sequences in the present study from hot springs, together with a number of sequences closest in BLASTN analysis from different geographical regions.
Fig. S6. Phylogenetic tree based on 16S rRNA gene sequences (~240 bp) indicating all obtained *Planktothricoides* related sequences in the present study from hot springs, together with a number of sequences closest in BLASTN analysis from different geographical regions.
Fig. S7. Phylogenetic tree based on 16S rRNA gene sequences (~240 bp) indicating all obtained *Xenoccoccus* related sequences in the present study from hot springs, together with a number of sequences closest in BLASTN analysis from different geographical regions.
Fig. S8. Phylogenetic tree based on 16S rRNA gene sequences (~240 bp) indicating all obtained *Anabaena* related sequences in the present study from hot springs, together with a number of sequences closest in BLASTN analysis from different geographical regions.
Fig. S9. Phylogenetic tree (expanded) based on 16S rRNA gene sequences (~240 bp) indicating all obtained unidentified sequences in the present study from hot springs, together with a number of sequences closest in BLASTN analysis from different geographical regions.
Fig. S10. Bacterial lineages distribution according to MEGAN analysis of metagenomic reads.
Table S1. Dominant species and phenotypes of *Cyanobacteria* in hot springs at the Lake Bogoria observed under light microscope.
Table S2. Number of 16S rRNA gene sequences phylogenetically related to cyanobacterial order/taxon and employed in the present study.
Table S3. Metagenomic sequencing results with DNA extracted from three sampling sites.