Soda pans of the Pannonian steppe harbor unique bacterial communities adapted to multiple extreme conditions

Attila Szabó¹, Kristóf Korponai¹, Csaba Kerepesi², Boglárka Somogyi³, Lajos Vörös³, Dániel Bartha⁴, Károly Márialigeti¹, Tamás Felföldi¹

¹ Department of Microbiology, Eötvös Loránd University, Pázmány Péter stny. 1/C., 1117 Budapest, Hungary.
² Institute for Computer Science and Control, Hungarian Academy of Sciences (MTA SZTAKI), Kende u. 13-17., 1111 Budapest, Hungary.
³ MTA Centre for Ecological Research, Balaton Limnological Institute, Klebelsberg Kunó u. 3., 8237 Tihany, Hungary.
⁴ Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Hungária krt. 21, 1143 Budapest, Hungary.

Corresponding author:
Tamás Felföldi
Department of Microbiology, Eötvös Loránd University; Pázmány Péter stny. 1/C., 1117 Budapest, Hungary.
E-mail: tamas.felfoldi@gmail.com; Tel: +36-1-372-2500/8384; Fax: +36-1-381-2178

Acknowledgements
The authors are thankful to Emil Boros, Balázs Németh and Tamás Sápi for their assistance during sampling. We are thankful to Annamária Kéri and János Kubisch for their help in establishing the computational environment for the CoVennTree program.

Funding
This work was financially supported by the Hungarian Scientific Research Fund (grants OTKA PD105407, PD112449); the Momentum Grant of the Hungarian Academy of Sciences (LP2012-
19/2012) to CS.K, and the Bolyai János Research Grant (Hungarian Academy of Sciences) to B.S and T.F. Purchase of equipment was financed by the National Development Agency (grants KMOP-4.2.1/B-10-2011-0002, TÁMOP-4.2.2/B-10/1-2010-0030).
Abstract

Soda pans of the Pannonian steppe are unique environments regarding their physical and chemical characteristics: shallowness, high turbidity, intermittent character, alkaline pH, polyhumic organic carbon concentration, hypertrophic condition, moderately high salinity, sodium and carbonate ion dominance. The pans are highly productive environments with picophytoplankton predominance. Little is known about the planktonic bacterial communities inhabiting these aquatic habitats, therefore amplicon sequencing and shotgun metagenomics were applied to reveal their composition and functional properties. Results showed a taxonomically complex bacterial community which was distinct from other soda lakes regarding its composition, e.g. the dominance of class Alphaproteobacteria was observed within phylum Proteobacteria. The shotgun metagenomic analysis revealed several functional gene components related to the harsh and at the same time hypertrophic environmental conditions, e.g. proteins involved in stress response, transport and hydrolase systems targeting phytoplankton-derived organic matter. This is the first detailed report on the indigenous planktonic bacterial communities coping with the multiple extreme conditions present in the unique soda pans of the Pannonian steppe.

Keywords
soda pan, metagenomics, bacterial community composition, high turbidity, environmental stress, osmoadaptation
**Introduction**

Astatic soda pans are characteristic aquatic environments in the steppe of the Pannonian biogeographic region (Carpathian Basin, Central Europe). According to current knowledge, soda pans in Europe are restricted to this area (Boros et al. 2014, 2017). Compared to the deep soda lakes in North America and Africa (Anthony et al. 2013; Dimitriu et al. 2008; Grant 2004; Lanzén et al. 2013), soda pans in this region are shallow and frequently dry out completely by the end of the summer. Hypersaline soda lakes of the Kulunda Steppe have much higher salinity (Foti et al. 2007), than the Pannonian soda pans; salinities at the latter sites vary generally within the hyposaline range (Boros et al. 2014). Another special limnological characteristic of the Pannonian soda pans is the high turbidity caused by high amount of inorganic suspended solid particles and/or the high humic substance content which gives brownish color to the water (Boros et al. 2014; Felföldi et al. 2009; Pálffy et al. 2014; Somogyi et al. 2009). Under the resulted light-limited conditions, the dominance of small-sized phytoplankton (i.e. picophytoplankton, PPP, <3 µm cell size) is favored (Felföldi et al. 2009; Somogyi et al. 2009) due to their increased surface to volume ratio (Raven 1998). Since nutrient availability is high in these pans, PPP blooms arise frequently (Pálffy et al. 2014; Somogyi et al. 2009). Sometimes dual blooms of green algae and purple bacteria can be observed (Borsodi et al. 2013). Organic carbon and inorganic nitrogen and phosphorous derived from decaying plant material of the shoreline vegetation and from the excrements of aquatic birds (Boros et al. 2008, 2016) provides the nutritional basis of the growth of both phototrophic and heterotrophic microorganisms. Taken together, shallowness, intermittent character (periodic desiccation), high turbidity, alkaline pH, polyhumic organic carbon concentration, hypertrophic condition and during summer high daily water temperature fluctuation create multiple extreme environmental conditions in these soda pans (Boros et al. 2017).

There are a huge number of studies targeting the prokaryotic communities inhabiting soda lakes worldwide (e.g. Dimitriu et al. 2008; Lanzén et al. 2013; Sorokin et al. 2014; Vavourakis et al. 2016), but the composition of planktonic bacteria in the unique, PPP-dominated Pannonian soda pans is practically unknown (Borsodi et al. 2013). Therefore, our research aimed to reveal the structure and function of bacterial communities inhabiting three different soda pans of this region using recent tools of metagenomics.

**Material and methods**
Site description, sample collection, determination of physical and chemical parameters

Samples were collected on 29th of November 2012 from three pans. Büdös-szék (46°51.980’N, 19°10.153°E) and Zab-szék (46°50.190’N, 19°10.283°E) soda pans have a surface area of 70 ha and 182 ha, respectively, and they represent the ‘turbid-white’ type of soda pans dominated by large amounts of suspended clay particles (Boros et al. 2014). Sós-ér pan (46°47.341’N, 19°8.679°E) is 3 ha large, has ‘non-turbid, colored’ water with large amounts of dissolved humic substances, its shoreline vegetation is dominated by bayonet grass (*Bolboschoenus maritimus*) which is the main source of the humic material (Boros et al. 2014). The characteristic pH of these sites is between 9-10, dominant ions are sodium and hydrogen carbonate, and pans have an average depth of 30-40 cm, however, in some years their water is completely evaporated (Felföldi et al, 2009; Somogyi et al., 2009; Boros et al. 2014).

In the case of each pan, composite samples were taken from at least ten different points in the deepest parts of the open water. Determination of limnological parameters and microscopic analyses were performed as described previously (Pálffy et al. 2014).

DNA extraction

Total genomic DNA was extracted from 500 µL composite water sample using the UltraClean Soil DNA Isolation Kit (MO BIO Laboratories Inc., Carlsbad, CA, USA) according to the manufacturer’s instructions with the exception that cell disruption step was carried out by shaking at 30 Hz for 2 min in a Mixer Mill MM301 (Retsch, Haan, Germany). Extracted DNA was stored at -20°C until further processing.

16S rRNA gene sequencing

For the determination of the bacterial community composition, V3-V4 region of the 16S rRNA gene was amplified using universal bacterial primers: S-D-Bact-0341-b-S-17 forward (5’- CCT ACG GGN GGC WGC AG-3’) and S-D-Bact-0785-a-A-21 reverse (5’- GAC TAC HVG GGT ATC TAA GCC-3’) (Klindworth et al. 2013), fused with proper sequencing barcodes and adapters. To minimize the stochastic effects of the reaction, the PCR amplification was performed in triplicates in 20 µL final volume containing 1× Phusion HF Buffer (Thermo Fisher Scientific Inc., Waltham, MA, USA), 0.2 mM dNTPs (Fermentas, Vilnius, Lithuania), 0.4 µg µL⁻¹ Bovine Serum Albumin (Fermentas), 0.5 µM
of each primer, 0.4 U Phusion High-Fidelity DNA Polymerase (Thermo Fisher). The following thermal conditions were used: initial denaturation at 98 °C for 5 minutes, followed by 25 cycles of denaturation (95 °C for 40 s), annealing (55 °C for 2 minutes) and extension (72 °C for 1 minute) and a final extension step at 72 °C for 10 minutes. Amplicons were pooled before the purification step, then the resulted libraries were purified with the High Pure PCR Cleanup Micro Kit (Roche/454 Life Sciences, Branford, CT, USA). Quality control of the amplicon libraries was carried out using a model 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Emulsion PCR, amplicon library processing and pyrosequencing were performed on a GS Junior sequencing platform according to the Lib-L protocol of the manufacturer (Roche/454 Life Sciences). Initial data processing was performed using a gsRunProcessor 3.0. Raw sequence data have been submitted to the NCBI Sequence Read Archive under the accession code SAMN03284852, SAMN05804901 and SAMN05804942 within the BioProject ID PRJNA272672.

Resulting sequence reads were processed using the mothur v1.35 software (Schloss et al. 2009) based on the 454 standard operating procedure (http://www.mothur.org/wiki/454_SOP - downloaded at 04/07/2015) (Schloss et al. 2011). To minimize the amplification and pyrosequenceing bias, sequences were quality filtered and denoised, furthermore the removal of chimeric sequence reads using the uChime program (Edgar et al. 2011) and singleton sequences according to Kunin et al. (2010) were carried out. Sequence alignment was performed with the SINA v1.2.11 aligner tool (Pruesse et al. 2012) using the ARB-SILVA SSU NR 99 reference database – SILVA Release 123 (Quast et al. 2013) for alignment and classification. Sequences classified as Archaea (0.05%) and ‘Chloroplast’ (2.48% of total reads) were excluded from further analysis (no reads were classified as ‘Mitochondria’, ‘Eukaryota’ or ‘unknown’). Operational taxonomic units (OTUs) were assigned at 97 % similarity threshold levels, representing bacterial species (Tindall et al. 2010). For visualization the distribution of the most abundant 50 OTUs among samples, CoVennTree (Lott et al. 2015) was used, a tool on the Galaxy platform (Blankenberg et al. 2010; Giardine et al. 2005; Goecks et al. 2010). The resulted output was visualized in Cytoscape 2.8.3 (Shannon et al. 2003). The ratio and distribution of reads are shown at different taxonomic levels corresponds to their relative abundance in the dataset in decreasing order; taxonomic assignments were made when the bootstrap values were higher than 80 based on the ARB-SILVA SSU NR reference database. For subsequent statistical analysis, sample reads were subsampled with the read number of the smallest data set. Richness estimators and diversity indices were calculated using mothur.

Detailed description of the pipeline and the scripts used are given in the Supplementary Material.
Shotgun metagenomic analysis

For the shotgun metagenomic analysis, three libraries were prepared from three DNA isolates from a composite sample taken from the Büdös-szék pan on 29th November 2012. The three libraries were sheared and prepared for sequencing with the Ion Xpress Plus Fragment Library Kit and the Ion PGM Template OT2 200 Kit (Life Technologies). Sequencing was performed with the Ion PGM Sequencing 200 Kit v2 on 314 chips using Ion Torrent PGM (Life Technologies). Raw sequence signals were analyzed with the Ion Torrent Suite software 3.6.2 (Life Technologies). Resulted fastq files were merged together for further processing and are available in the NCBI Sequence Read Archive under the accession code SAMN03284852 within the BioProject ID PRJNA272672.

Shotgun reads were filtered based on their average quality score (Q ≥ 24) with PRINSEQ v0.20.4 (Schmieder and Edwards 2011), also sequence duplicates were removed and bases less than phred=10 were trimmed from the end of the sequences. Filtered reads containing gene sequences were identified with the blastx command of DIAMOND (Buchfink et al. 2015) against the NCBI NR database (downloaded at 22/02/2015) in sensitive mode with 0.001 e-value cutoff (default) and set the max target sequences option to 250 (default is 25). Taxonomic and functional assignments were made with MEGAN 5.11 (Huson et al. 2007) against the NCBI and SEED classification (downloaded at 12/05/2015 NCBI and 01/11/2014 SEED) using the default parameters. Additionally, raw sequence reads (637,468 reads, 105.0 Mbp) were submitted to MG-RAST (Meyer et al. 2008), processed using the default parameters, and are available under the project ‘Budos-szek soda pan metagenome’ with the accession code mgp8260 (link: http://metagenomics.anl.gov/linkin.cgi?project=mgp8260).

Results and discussion

Physical and chemical characteristics of soda pans

Measured limnological parameters are given in Table 1. Salinity values of the pans ranged between 3.74 to 10.6 g L\(^{-1}\), therefore all lakes could be defined as hyposaline according to Hammer (1986). Converting the conductivity data measured by Somogyi et al. (2009) with the empirical equation given in Boros et al. (2014), salinity values ranged 5.4-15.2 g L\(^{-1}\) and 4.8-9.6 g L\(^{-1}\) throughout a year (between July 2006 and May 2007) in Büdös-szék and Zab-szék pans, respectively. These values are
similar to those measured in this study and denote that salinity changes significantly throughout the year, although remains within the hyposaline range. Dissolved organic carbon content was the highest in Sós-ér pan (814 mg L\(^{-1}\)) corresponding to its ‘colored’ type. The concentration of total suspended solids was a magnitude higher in Büdös-szék pan (5307 mg L\(^{-1}\)) than the other two sites, since it represents a ‘turbid’ type soda pans and at the time of sampling it was close to desiccation (water depth: 2 cm). Nutrient availability was high in the case of all three pans (TP concentration, ~4.9 mg L\(^{-1}\); TN\(_{ammonium+nitrate+urea}\) concentration, ~0.2-0.5 mg L\(^{-1}\)), as in general throughout the whole year (Boros et al. 2008).

Chlorophyll \(a\) concentration in the pans ranged between 20 and 60 µg L\(^{-1}\), and were the highest in Büdös-szék pan. These values were lower than the yearly average values (289 µg L\(^{-1}\) and 109 µg L\(^{-1}\) at Büdös-szék and Zab-szék pans, respectively, recorded in 2006-2007) reported from the sites by Somogyi et al. (2009), which clearly indicated their hypertrophic status; this was also confirmed later (2009-2010) by Boros et al. (2017). Using epifluorescence microscopy, picoeukaryotes were the dominant phytoplankters in all of the studied pans, while picocyanobacteria were detected only in Zab-szék pan (Table 1). Based on the results of laboratory and field studies, planktonic picoeukaryotic algae have competitive advantage in environments with low temperature and low light intensity (Somogyi et al. 2009; Vörös et al. 2009; Weisse 1993). Lower salinity, the potentially high amount of algal-derived organic matter (based on the PPP biomass and chlorophyll a content), and the high concentrations of nitrogen and phosphorous forms may contributed to that Büdös-szék harbored the most diverse bacterial community at the time of sampling (Supplementary Table S1).

Taxonomic composition of bacterial communities in soda pans

The 16S rRNA gene amplicon sequencing of the samples resulted a total of 14,488 high quality reads classified within the Bacteria domain. Similarly to planktonic bacterial communities inhabiting other soda lakes (Dimitriu et al. 2008; Lanzén et al. 2013; Paul et al. 2016; Vavourakis et al. 2016), all three samples were dominated by members of the phyla Proteobacteria (61-30%) and Bacteroidetes (53-22%), while in Büdös-szék, ratio of Actinobacteria (25%) was also significant (Fig. 1a). Cytophagia and Flavobacteria were detected as the most abundant classes within phylum Bacteroidetes. Within phylum Proteobacteria the dominance of Alphaproteobacteria was observed, however in other soda lakes Gammaproteobacteria was detected as the most abundant class of this phylum. Within Alphaproteobacteria, several genera (Roseococcus, Rhodobaca and Salinarimonas) were identified which consist strains capable (or putatively capable) of photoheterotrophic growth (Brenner et al. 2005; Cai et al. 2011), those were mainly affiliated with the order Rhodobacterales (Fig. 1c). In general, other
identified genera contain mainly aerobic heterotrophs and have many halophilic and halotolerant members [Altererythrobacter, Loktanella, Seohaeicola, Pseudospirillum, Salinarimonas, Aliidiomarina, Idiomarina, Flavobacterium and Indibacter (Anil Kumar et al. 2010; Brenner et al. 2005; Cai et al. 2011; Chiu et al. 2014; Jung et al. 2014; Krieg et al. 2010; Satomi et al. 2002; Van Trappen et al. 2004; Yoon et al. 2009)]. Some members of these genera are even alkaliphilic [Mongoliicoccus and Nitriliruptor (Goodfellow et al. 2012; Liu et al. 2005)] corresponding to the relatively high pH (9.1-9.7) of these pans.

There were 31 shared OTUs among the three pans, 70 OTUs were shared between the Büdös-szék and the Sós-ér sample, 55 between the Büdös-szék and the Zab-szék, and 66 between the Sós-ér and the Zab-szék (Figure 1b). Abundant shared OTUs (with relative abundance ≥ 1%), representing a core bacterial community of the pans, were related to the taxa Flavobacteriaceae (OTU1), Rhodobacteraceae (OTU4, OTU16), BIg5 (family-level group of Cytophagia) (OTU5), Comamonadaceae (OTU6), Rhizobiales (OTU8), Microbacteriaceae (OTU21) and Verrucomicrobiaceae, and the genera Loktanella (OTU9), Luteolibacter (OTU12, OTU22), Indibacter (OTU14), Salinarimonas (OTU19) and Methylophaga (OTU20) (Figure 1c, Supplementary Table S2).

Based on the phenotypic properties deduced from species descriptions, functional groups of bacteria were represented by markedly different genera from those observed in other soda lakes worldwide (reviewed in Sorokin et al. 2015), e.g. Methylophaga was the dominant methylotrophic bacterium (Kalyuzhnaya et al. 2006) not Methylomicrobium and Methylophaga as in other soda lakes. Similarly, previous studies have shown that planktonic primary producers also had different community composition in these habitats, cyanobacteria are dominated by Synechococcus, while eukaryotic algae by Chloroparva and Choricystis (Felföldi et al. 2009, 2011; Somogyi et al. 2009, 2010, 2011, 2016) contrary to Spirulina, Arthospira and Picocystis, Dunaliella, respectively, abundant in other soda lakes (Krienitz and Kotut 2010; Schagerl et al. 2015; Sorokin et al. 2015).

Metagenomic overview of Büdös-szék soda pan

The highest species richness was found in the Büdös-szék pan sample (Supplementary Table S1), therefore this was processed for functional metagenomic analysis. In this shotgun approach, quality-filtering resulted 497,312 high-quality reads with a mean length of 170 ± 48 nt (overall 84.6 Mbp sequence data) and the following taxonomic assignment: 94.0% Bacteria, 0.2% Archaea, 2.0% Eukaryota and 3.8% viruses. Abundant bacterial orders were Cytophagales, Flavobacteriales, Bacteriodales, Sphingobacteriales (phylum Bacteroidetes), Actinomycetales (phylum Actinobacteria),
Rhodobacterales (class Alphaproteobacteria), Burkholderiales and Methylophilales (class Betaproteobacteria) as in the 16S rRNA amplicon study (Fig 1a).

Viral sequences were dominated by hits assigned to bacteriophages (mainly Caudovirales), which may control bacterial community composition and through host cell lysis affects the availability of organic carbon compounds and nutrients (Atanasova et al. 2015; Mühling et al. 2005; Wilhelm & Shuttle 1999).

A total of 165,823 functional hits were identified using the SEED classification in MEGAN and 44,083 were assigned to subsystems (Supplementary Table S3). Results showed a functionally complex community with several genes related to the harsh environmental conditions present in the studied soda pans (Table 2, Supplementary Table S4). According to all the obtained data, the following processes and mechanisms related to planktonic bacteria are presumed.

Residence of aquatic birds and algal blooms provide high nutrient supply (Boros et al. 2008, 2016; Somogyi et al. 2009), which results in the high abundance of heterotrophic organisms (Vörös et al., 2008), such as members of phylum Bacteroidetes. These bacteria (especially from the order Flavobacteriales) favor to attach to organic particles and have high abundances in nutrient-rich habitats (Williams et al. 2013), since they participate in the degradation of biopolymers, such as algae-derived particulate organic matter (Buchan et al. 2014; Xing et al. 2015). Genes encoding receptors of the TonB-dependent transporter (TBDT) systems, responsible for biopolymer uptake (Williams et al. 2013), were among the most abundant genes in the shotgun metagenomic dataset. Most of the TonB-dependent receptor hits were assigned to orders Cytophagales and Flavobacteriales within phylum Bacteroidetes (64.8%). In general, members of Cytophagales and Flavobacteriales are well-known degraders of high-molecular-weight organic matter, such as proteins and polysaccharides in aquatic environments (Kirchman 2002). Additionally, TBDT-related degradative enzymes (e.g. glycoside hydrolases, aminopeptidases) were identified with best matches to Bacteroidetes and Proteobacteria.

Members of Rhodobacterales are also abundant during phytoplankton blooms in marine environments using algal exudates as substrate (Buchan et al. 2014; Teeling et al. 2012; Williams et al. 2013). Based on the results of shotgun metagenomics and the community structure profile, it could be hypothesized that these bacterial groups could have similar functions in the studied soda lakes as in the oceans.

Genes involved in the serine-glyoxalate cycle and other pathways related to one-carbon metabolism were also abundant with best matches to Bacteroidetes and Proteobacteria, most probably due to methane and C1-compounds originating from the sediment, which are subsequently utilized by methylotrophic bacteria (Sorokin et al. 2015). As mentioned above, genus *Methylotenera* was a characteristic methylotrophic bacterium in the 16S rRNA gene amplicon sequencing data (Fig. 1c) and
many of the functional genes were assigned to the genus in the metagenomics dataset. Members of this

287 genus assimilate C1-compounds via the ribulose-monophosphate pathway and could use methanol,
betaine, pyruvate and fructose as sole energy and carbon source (Doronina et al. 2014).

288 Genes related to fermentative metabolism were assigned to every detected major bacterial
289 phyla, however their presence was meager (n = 464) compared to respiratory processes (n = 2421). Although most of the inhabiting microorganisms have chemoheterotrophic lifestyle, several gene
290 components related to autotrophic CO2-fixation were also found. Gene hits related to photosynthesis
291 were scarce (n = 36), those were structural components related to the photosystems of cyanobacteria
292 and green algae and related to anoxygenic photosynthesis (e.g. PufQ), the latter having best matches to
293 Rhodobacterales and Burkholderiales. Interestingly rhodopsin genes were absent from the shotgun
294 dataset, which could be explained with that the organic matter content of the pans are extremely high
295 throughout the year (Boros et al. 2017) compared to marine environments, therefore complementary
296 light energy utilization for heterotrophic bacteria (e.g. members of Flavobacteria and Actinobacteria,
297 which are abundant in these sites according to 16S rRNA gene data; Fig. 1a,c) are unnecessary.

298 Although the penetration of UV-B radiation is limited to the upper few centimeters in these
299 turbid soda pans (V.-Balogh et al. 2009), its impact also depends on mixing processes (which are rather
300 intense, since the number of windy days are >120 in this region, Boros et al. 2017) reducing the
301 shadowing effect of chromophoric dissolved organic matter, suspended solids and algae. On the other
302 hand, since Büdös-szék was close to desiccation at the time of sampling, the whole water body could
303 have been exposed to UV radiation, presumably this also contributed to the high abundance of genes
304 related to DNA repair mechanisms found in every detected major bacterial phyla. Including the hits
305 assigned to the category ‘DNA replication’ the relative abundance of the hits belongs to ‘DNA
306 metabolism’ were higher (11.3%) than in the studied marine (8.4%) and freshwater (6.4%)
307 metagenomes (Eiler et al. 2014). Many genes (e.g. thioredoxin, superoxide dismutase) involved in the
308 response to oxidative stress were also abundant, since the generation of reactive oxygen species is
309 another effect of solar irradiation in aerobic waters (Williams et al. 2013). Furthermore, organisms
310 have to adapt to the high pH and to the variable salt content of the pans, which are also a source of
311 stress (Boros et al. 2017).

312 The survival and growth for an aerobic microorganism in highly alkaline conditions is quite
313 challenging: the organism generally use the proton motive force for energy conversion, however the
314 proton concentration of the environment is lower than the intracellular, therefore retaining H+ in the
315 periplasm and importing H+ into the cytoplasm are crucial for cellular homeostasis. Huge variety of
316 possible adaptation mechanisms was detected to maintain the optimal intracellular pH in these alkaline
environments. Several genes of Na\(^+\)/H\(^+\) (n = 196) and K\(^+\)/H\(^+\) antiporters (n = 75) were identified in the shotgun dataset mostly related to Bacteroidetes and Proteobacteria. Their role is to import protons to the cytoplasm while pumping out a counterbalancing monovalent cation to the periplasm. Additionally, several other transporters, which generally have K\(^+\)/H\(^+\) symporter function were identified (n = 102) along with other H\(^+\)/solute symporters. Another way to translocate protons into the cytoplasm is the higher expression of V-type (n = 10) and F-type ATP synthases (n = 297) (Krulwich et al. 2011).

Catabolic activities producing organic acids such as the identified genes of deaminases could also increase the intracellular proton concentration (Krulwich et al. 2011).

Alkaline environments like the studied soda pans contain high amounts of sodium (91.2 - 97.0 e\% in the cation pool, Boros et al. 2014). Prokaryotic cells could maintain a Na\(^+\) cycle in which sodium pumps and sodium motive force consumers like Na\(^+\)-dependent membrane transporters, ATP synthases and flagellar motors operate in concert (Mulkidjanian et al. 2008). Na\(^+\)/solute symporters could use the sodium motive force and import Na\(^+\) to the cytoplasm to support the Na\(^+\)/H\(^+\) antiporter activity. Excess of sodium could be expelled from the cell via Na\(^+\)-pumping NADH-CoQ reductase (NQR) (n = 88), assigned mostly to Bacteroidetes (60%) and Na\(^+\)-pumping NADH: ferredoxin dehydrogenase (RNF) (n = 29), assigned mostly to Proteobacteria (83%). The latter could transport Na\(^+\) and H\(^+\) in both direction (Banciu and Muntyan 2015; Reyes-Prieto et al. 2014). Using our approach, we were not able to identify genes of Na\(^+\) channel proteins, voltage gated sodium channels, Na\(^+\) dependent flagellar motors, and the distinction between H\(^+\) or Na\(^+\) translocating ATPases and H\(^+\) or Na\(^+\)-motive cytochrome c oxidases (Banciu and Muntyan 2015; Muntyan et al. 2015; Sorokin et al. 2014) was not possible.

Based on metagenomic analyses, the preferred usage of potassium instead of sodium for osmoregulation is a previously described feature of freshwater communities compared to marine habitats (Eiler et al. 2014; Oh et al. 2011). Based on our findings probably both cation transporters are important in the community, however the concentration of sodium (97.0 e\%) is much higher than potassium (0.5 e\%) in the studied environment (Boros et al. 2014).

The salt content of soda lakes causes osmotic stress to the inhabiting microorganisms. For the maintenance of osmotic balance, organisms can accumulate inorganic osmolytes such as KCl (‘salt-in’ strategy) or organic compatible solutes e.g. ectoine, glycine betaine (‘salt-out’ strategy). Based on previous studies, the ‘salt-out’ strategy is the main osmoadaptive mechanism of the vast majority of aerobic soda lake bacteria (Banciu and Muntyan 2015; Oren 1999). Several functional components of the uptake and synthesis of compatible solutes were identified in the Büdös-szék metagenome. However, ectoine was described as a dominant organic osmoprotectant of halotolerant organisms favoring low to moderate salt concentration values, while glycine betaine represents the typical organic
osmolyte for extreme salt-tolerant haloalkaliphiles (Banciu and Muntyan 2015). Contrary to this, in our
sample (a habitat with moderate salinity), hits assigned to choline and betaine uptake and biosynthesis
were five times more abundant (related to Alphaproteobacteria, Cytophagia, Flavobacteria and
Actinobacteria) compared to ectoine biosynthesis (related to Alpha- and Gammaproteobacteria and
Planctomycetes). The presence of numerous $\mathrm{K}^+$/H$^+$ symporter and $\mathrm{K}^+$ channel protein genes may
contribute to the accumulation of inorganic potassium salts within the cytoplasm (‘salt-in’ strategy) and
this was described as a characteristic feature for archaeal taxa and anaerobic halophilic bacteria
(Banciu and Muntyan 2015; Oren 1999), however the relative abundance of these prokaryotes in the
studied soda pans was low.

In contrast to other alkaline lakes worldwide, Archaea have surprisingly low abundance and
presumably have only a minor role in the planktonic communities based on the shotgun metagenomic
data. This could be due to the hyposaline milieu and the permanently high amounts of nutrients.
Seemingly bacterial taxa have a broad range of adaptation mechanisms and under these conditions
outcompete Archaea in the pans.

Concluding remarks

In conclusion, the nutrient-rich, alkaline pans in the Pannonian steppe with the dominance of
sodium and hydrogen carbonate ions provide a unique environment for microorganisms. This first
snapshot on the taxonomic and functional diversity revealed bacterial communities different from those
present in soda lakes worldwide, and special metabolic and physiological characteristics associated
with these extreme conditions.
References


Fig. 1 Bacterial taxonomic composition of soda pan samples (29th November 2012). **a** Phylum-level distribution of reads among major lineages expressed as a percentage of total sequences (in the case of Bacteroidetes and Proteobacteria most relevant classes are also shown; phyla having <5% relative abundance are combined in the category ‘other taxa’). **b** Distribution of OTUs (defined at 97% similarity level) among normalized sample datasets. **c** Distribution of the most abundant 50 OTUs among the samples visualized with CoVennTree [numbers in brackets assigned to a parent node are the VDS values (‘Venn decomposition similarity’, see details in Lott et al. 2015) representing similarity among children; color-coding is the same as in Fig. 1b; unc., unclassified; OTU numbers correspond to...
relative abundance in decreasing order; node size correlate with the number of sequences within a sample].
Table 1 Environmental and biological parameters of the studied pans (29th November 2012)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Depth (cm)</th>
<th>Temperature (°C)</th>
<th>pH</th>
<th>Salinity* (g L⁻¹)</th>
<th>Chl (µg L⁻¹)</th>
<th>NH₄⁺-N (µg L⁻¹)</th>
<th>NO₃⁻-N (µg L⁻¹)</th>
<th>urea-N (µg L⁻¹)</th>
<th>TP (mg L⁻¹)</th>
<th>DOC (mg L⁻¹)</th>
<th>TSS (mg L⁻¹)</th>
<th>CyPPP biomass** (µg L⁻¹)</th>
<th>EuPPP biomass** (µg L⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Büdös-szék</td>
<td>2</td>
<td>14.7</td>
<td>9.16</td>
<td>3.7</td>
<td>59.6</td>
<td>159</td>
<td>97</td>
<td>294</td>
<td>9.30</td>
<td>48</td>
<td>5307</td>
<td>&lt;0.1</td>
<td>425</td>
</tr>
<tr>
<td>Zab-szék</td>
<td>7</td>
<td>14.1</td>
<td>9.67</td>
<td>8.8</td>
<td>32.4</td>
<td>2.2</td>
<td>40</td>
<td>141</td>
<td>7.94</td>
<td>60</td>
<td>296</td>
<td>1.37</td>
<td>962</td>
</tr>
<tr>
<td>Sós-ér</td>
<td>25</td>
<td>13.3</td>
<td>9.15</td>
<td>10.6</td>
<td>20.8</td>
<td>179</td>
<td>75</td>
<td>247</td>
<td>4.07</td>
<td>814</td>
<td>341</td>
<td>&lt;0.1</td>
<td>20</td>
</tr>
</tbody>
</table>

Abbreviations: Chl – chlorophyll a concentration, TP – total phosphorous concentration, DOC – dissolved organic carbon concentration, TSS – total suspended solids concentration, CyPPP – picocyanobacteria, EuPPP – photoautotrophic picoeukaryotes

* Calculated from conductivity values according to the empirical formula of Boros et al. 2014

** Wet weight
Table 2 Highlighted functional traits detected in the Büdös-szék soda pan metagenome

<table>
<thead>
<tr>
<th>Metabolic pathways/Adaptation mechanisms</th>
<th>Count</th>
<th>%</th>
<th>Examples</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Algae-derived organic matter uptake (TBDT system and related enzymes)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TonB-dependent transporter system</td>
<td>793</td>
<td>1.80</td>
<td>COG2152 predicted glycoside hydrolase</td>
<td>45</td>
</tr>
<tr>
<td>Glycoside hydrolases</td>
<td>151</td>
<td>0.34</td>
<td>COG1649 predicted glycoside hydrolase</td>
<td>34</td>
</tr>
<tr>
<td>Aminopeptidases</td>
<td>37</td>
<td>0.08</td>
<td>Proline iminopeptidase (EC 3.4.11.5)</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Asp-X dipeptidase</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Deblocing aminopeptidase (EC 3.4.11.-)</td>
<td>7</td>
</tr>
<tr>
<td><strong>One-carbon metabolism</strong></td>
<td>1253</td>
<td>2.84</td>
<td>Formate-tetrahydrofolate ligase (EC 6.3.4.3)</td>
<td>44</td>
</tr>
<tr>
<td>One-carbon metabolism by tetrahydropterines</td>
<td>93</td>
<td>0.21</td>
<td>Methanol dehydrogenase large subunit protein (EC 1.1.99.8)</td>
<td>40</td>
</tr>
<tr>
<td>Serine-glyoxylate cycle</td>
<td>1135</td>
<td>2.57</td>
<td>Serine hydroxymethyltransferase (EC 2.1.2.1)</td>
<td>61</td>
</tr>
<tr>
<td>Ribulose-monophosphate pathway</td>
<td>25</td>
<td>0.06</td>
<td>Formaldehyde activating enzyme</td>
<td>16</td>
</tr>
<tr>
<td><strong>Autotropic CO₂-fixation</strong></td>
<td>928</td>
<td>2.11</td>
<td>Carboxyosome NADH dehydrogenase (EC 1.6.99.3)</td>
<td>21</td>
</tr>
<tr>
<td><strong>DNA repair mechanism - UV stress</strong></td>
<td>1959</td>
<td>4.44</td>
<td>Excinuclease ABC subunits A, B and C</td>
<td>421</td>
</tr>
<tr>
<td>DNA polymerase I (EC 2.7.7.7)</td>
<td></td>
<td></td>
<td>ATP-dependent DNA helicase RecQ</td>
<td>96</td>
</tr>
<tr>
<td>DNA mismatch repair protein MutS</td>
<td></td>
<td></td>
<td>ATP-dependent DNA helicase UvrD/PcrA</td>
<td>94</td>
</tr>
<tr>
<td>DNA repair mechanism                                      <strong>Oxidative stress – reactive oxigen species</strong></td>
<td>757</td>
<td>1.72</td>
<td>Thiolredoxin</td>
<td>87</td>
</tr>
<tr>
<td>Thioredoxin                                               <strong>Adaptation to alkalinity and salinity</strong></td>
<td></td>
<td></td>
<td>Gamma-glutamyltranspeptidase (EC 2.3.2.2)</td>
<td>64</td>
</tr>
<tr>
<td>Gamma-glutamyltranspeptidase</td>
<td></td>
<td></td>
<td>Hydrogen peroxide-inducible genes activator</td>
<td>30</td>
</tr>
<tr>
<td>Manganese superoxide dismutase (EC 1.15.1.1)</td>
<td></td>
<td></td>
<td>Superoxide dismutase [Fe] (EC 1.15.1.1)</td>
<td>25</td>
</tr>
<tr>
<td>Superoxide dismutase                                      <strong>Na⁺/H⁺ antiporters</strong></td>
<td>544</td>
<td>1.23</td>
<td>Na⁺/H⁺ antiporter</td>
<td>99</td>
</tr>
<tr>
<td>Na⁺/H⁺ antiporter subunit A</td>
<td>196</td>
<td>0.44</td>
<td>Glutathione-regulated potassium-efflux system ATP-binding protein</td>
<td>30</td>
</tr>
<tr>
<td>K⁺/H⁺ antiporters</td>
<td>75</td>
<td>0.17</td>
<td>Glutathione-regulated potassium-efflux system protein KefC</td>
<td>25</td>
</tr>
<tr>
<td>K⁺/H⁺ transporters, generally symporters</td>
<td>102</td>
<td>0.23</td>
<td>Potassium uptake protein, integral membrane component, KtrB</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Trk system potassium uptake protein TrkA</td>
<td>26</td>
</tr>
<tr>
<td>Metabolic pathways/Adaptation mechanisms</td>
<td>Count</td>
<td>%</td>
<td>Examples</td>
<td>Count</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-------</td>
<td>-----</td>
<td>-----------------------------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Solute/H⁺ symporters</td>
<td>22</td>
<td>0.05</td>
<td>Potassium uptake protein TrkH</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D-xylose proton-symporter XylE</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L-rhamnose-proton symporter</td>
<td>4</td>
</tr>
<tr>
<td>Smf-driven mechanisms</td>
<td>95</td>
<td>0.22</td>
<td>Acetate permease ActP (cation/acetate symporter)</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Na⁺/pantothenate symporter (TC 2.A.21.1.1)</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Na⁺/glycine symporter GlyP</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Na⁺/malate symporter</td>
<td>6</td>
</tr>
<tr>
<td>Na⁺ translocating NADH-quinone reductase (NQR)</td>
<td>88</td>
<td>0.20</td>
<td>Na⁺-translocating NADH-quinone reductase subunit B (EC 1.6.5.-)</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Na⁺-translocating NADH-quinone reductase subunit F (EC 1.6.5.-)</td>
<td>22</td>
</tr>
<tr>
<td>Na⁺-pumping NADH: ferredoxin dehydrogenase (RNF)</td>
<td>29</td>
<td>0.07</td>
<td>Electron transport complex protein RnfC</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Electron transport complex protein RnfD</td>
<td>6</td>
</tr>
<tr>
<td>K⁺ channels</td>
<td>84</td>
<td>0.19</td>
<td>Potassium voltage-gated channel subfamily KQT</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Osmosensitive K⁺ channel histidine kinase KdpD (EC 2.7.3.-)</td>
<td>15</td>
</tr>
<tr>
<td>Choline and betaine uptake and betaine biosynthesis ('salt-out' osmoadaptive strategy)</td>
<td>83</td>
<td>0.19</td>
<td>Choline-sulfatase (EC 3.1.6.6)</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sarcosine oxidase (EC 1.5.3.1)</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Choline dehydrogenase (EC 1.1.99.1)</td>
<td>10</td>
</tr>
<tr>
<td>Ectoaine biosynthesis and regulation ('salt-out' osmoadaptive strategy)</td>
<td>13</td>
<td>0.03</td>
<td>Diaminobutyrate-pyruvate aminotransferase (EC 2.6.1.46)</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Aspartokinase (EC 2.7.2.4) associated with ectoine biosynthesis</td>
<td>2</td>
</tr>
</tbody>
</table>

Abbreviations: Smf – Na⁺-motive force