UNIVERSITEIT VAN AMSTERDAM

A Meta-Omics View on Prokaryotic Life in Hypersaline Soda Lakes

C.D. Vavourakis

Metagenomics and metatranscriptomics are cutting edge techniques used to analyze the microbial communities in an environment (**Chapter 1**). Without the explicit need for cultivation, the relative abundance, transcriptional activity, metabolic potential and possible ecological role of members within microbial communities can be investigated. Making use of next-generation sequencing strategies and rapidly evolving bioinformatic techniques, metagenomics allows the reconstruction of metagenome-assembled genomes (MAGs) from novel, uncultured microorganisms that constitute the so-called "microbial dark matter". The recovery of thousands of MAGs from different environments has transformed our notion of the existing microbial diversity considerably. Major branches of prokaryotes appeared over the last ten years on the phylogenetic tree of life that lack representative isolated strains. Some span numerous proposed phyla, for example *Bacteria* from the Candidate Phyla Radiation (CPR) and *Archaea* from the DPANN group.

Hypersaline soda lakes have a salinity above 50 g L^{-1} and are strongly alkaline, with a pH typically between 9-11 buffered by dissolved carbonate/bi-carbonate salts. Soda lakes are formed through the evaporation of groundwater in closed basins in semi-arid and arid regions of Eurasia, Africa and America. The evaporative character and continental location of most soda lakes leads to highly unstable water regimes, especially in smaller soda lakes. Salinities in the same lake can vary considerably up to salt-saturating conditions of roughly 400 g L^{-1}, causing also strong fluctuations in their microbial community composition.

Extremophilic microbes that grow optimally under both high-salt and high-pH conditions are called haloalkaliphiles and are mostly prokaryotic. Haloalkaliphiles have evolved different strategies to both cope with the osmotic stress at high salinities and with the bioenergetic challenges caused by a higher pH outside the cell than in the cytoplasm at alkaline conditions (**Chapter 2**). Microbial-mediated redox reactions that occur under haloalkaline conditions have been already thoroughly investigated prior to this thesis, with microbiological cultivation approaches and molecular surveys of functional marker genes. Active carbon, nitrogen and sulfur cycles biogeochemical cycling in soda lakes is performed by diverse functional groups that are successful at different salinity ranges. Nitrification is restricted to lower salinities and most soda lake *Cyanobacteria* bloom at low-saline conditions. Heterotrophic *Archaea* and carbon-fixing purple sulfur bacteria prefer salinities up to saturation.

Bacteria that derive energy from the oxidation and reduction of inorganic sulfur compounds are highly active in hypersaline soda lakes throughout the complete salinity range. At high pH different sulfur intermediary compounds are chemically stable compared to environments with neutral or acidic pH. Most importantly, under anoxic conditions and in the presence of sulfide, stable polysulfides are formed from the reaction with elemental sulfur. Also distinct haloalkaliphilic taxonomic groups of sulfate-reducing bacteria, bacteria capable

of the reduction or disproportionation of polysulfides, thiosulfate and sulfite, and sulfur-oxidizing bacteria are characteristic for the sulfur cycle in soda lakes. Particularly successful are haloalkaliphilic sulfur-oxidizing bacteria that belong to the highly diverse, chemolithoautotrophic genus *Thioalkalivibrio*, with ten characterized species from over one hundred cultured isolates and 75 (near-) complete genomes sequenced.

For this thesis a meta-omics approach was combined with biogeochemical profiling and 16S rRNA gene amplicon sequencing to describe the prokaryote communities found in six different hypersaline soda lakes in the Kulunda Steppe, south-western Siberia (Russia). The aim was to i) identify the taxonomic and functional groups that were previously missed using more traditional microbiological approaches, ii) reconstruct MAGs from hypersaline soda lakes to expand the known microbial genomic diversity and to iii) estimate the metabolic potential and relative abundance from novel prokaryotes that were highly abundant both in the brine and sediments from the soda lakes or iv) likely involved in dissimilatory sulfur cycling.

From the brine metagenomes it was evident that *Archaea* were dominant over *Bacteria* when the salinity exceeded 250 g L^{-1} (**Chapter 3**). This trend was also previously observed in marine, hypersaline solar salterns that have a neutral pH. Several novel families and genera were detected by amplicon sequencing that lack haloalkaliphilic isolates. MAGs from novel species within the phyla *Bacteroidetes* and *Euryarchaeaota* that may play a role in the aerobic degradation of recalcitrant organic polymers, such as chitin and cellulose, were reconstructed. The first MAGs of likely haloalkaliphilic members of the "*Candidatus* Nanohaloarchaea" were described, a phylum within the DPANN group that encoded also for an enigmatic type of rhodopsin.

The sediment metagenomes were far more diverse than those obtained from the brines and even in soda lakes with salt-saturated brines, *Bacteria* were overall dominant over *Archaea* (**Chapter 4**). Among the detected dominant groups that lack cultured isolates were ML635J-40 (phylum *Bacteroidetes*), the families *Syntrophomonadaceae* and *Halobacteroidiaceae (Firmicutes)* and HOC36 (class *Gammaproteobacteria*). A total of 871 MAGs were reconstructed, several belonging to candidate phyla and groups within the CPR that appeared to be abundant in the sediments, yet were detected here for the first time in soda lakes. Several MAGs from uncultured groups within the *Actinobacteria*, *Chlamydiae*, "*Ca.* Handelsmanbacteria" and "*Ca.* Atribacteria" encoded for key-enzymes of the Wood-Ljungdahl pathway, which was never reported before for these phyla in other environments. This pathway was previously only characterized as a route for anaerobic carbon fixation or dissimilation in acetogens, some syntrophic acetate oxidizers, sulfate reducers and methanogenic *Archaea*. Possibly, the discovered organisms here belong to one of the first two functional groups, as they did not encode the capacity for methanogenesis or sulfate reduction.

The genes and pathways involved in syntrophic acetate oxidation were further investigated by metagenomic sequencing of a highly enriched consortium dominated by a hydrogenotrophic methanogen and the syntrophic acetate oxidizer "*Ca*. Syntrophonatronum acetioxidans" (family *Syntrophomonadaceae*) that was previously obtained from a hypersaline soda lake in the Kulunda Steppe (**Chapter 5**). Acetate can be oxidized through the reversal of the Wood-Ljungdahl pathway and activity measurements additionally showed that formate or H_2 was the interspecies electron carrier transferred to the methanogenic partner. From the abundance of the family *Syntrophomonadaceae* and the near-absence of other acetate oxidizers in the 16S rRNA gene sequence and metagenomic datasets obtained from the Kulunda soda lake sediments in Chapter 4, it was deduced that syntrophic acetate oxidation might be the major route for anaerobic acetate transformation in the hypersaline soda lake sediments.

Under moderately hypersaline conditions, a complete sulfur cycling between the most oxidized and reduced sulfur species, sulfate and sulfide, can occur in soda lakes, while at higher salinities the cycle might be shortened between polysulfides/elemental sulfur and sulfide due to the partial inhibition of sulfate reduction. To target all possible functional groups, the sulfur cycle was investigated in more detail in the moderately hypersaline Cock Soda Lake (**Chapter 6**). The majority of the prokaryotes were found to be present and transcriptionally active at the brine-sediment interface. *Bacteria* oxidizing or reducing inorganic sulfur compounds co-existed and were transcriptionally active in the top 2-5 cm sediment layer. Unexpectedly, thiosulfate oxidation in sediment slurries was partially inhibited at high substrate concentrations under light compared to dark conditions. Members of *Thioalkalivibrio* were relatively abundant throughout the investigated 25 cm deep sediment layer, while the chemolithoautotrophic sulfur-oxidizers from the genus *Thioalkalimicrobium* were abundant mostly in the brine. Distinct genera involved in the reductive half of the sulfur cycle were relatively more abundant at different sediment depths. Surprisingly, functional genes of a member of *Desulfonatronobacter*, the only known genus of haloalkaliphilic sulfate reducing bacteria capable of acetate oxidation, were highly transcribed in the sediment top layer. Novel MAGs were obtained, including from putative photoheterotrophic SOB from the family *Rhodobacteraceae*, colorless SOB with the capacity for N_2 fixation from the family *Thiohalomonadaceae*, thiosulfate oxidizing heterotrophic *Bacteroidetes* and a new member within the "*Ca.* Woesearchaeota" (DPANN group) that might be involved in sulfur or sulfite reduction. Several transcripts of tetrathionate reductases and polysulfide/thiosulfate reductases in the sediment top layer were highly abundant. Therefore, tetrathionate might be, in addition to polysulfide and thiosulfate, an important intermediate at haloalkaline conditions, despite its chemical instability at alkaline pH.

As only a handful of genomes from the more than two thousand reconstructed MAGs were analyzed in this thesis, future studies may further tap into this

wealth of genomic information to get more insights into the possible metabolic features encoded by these soda lake prokaryotes (**Chapter 7**). Genomes provide non-conclusive evidence for possible metabolic activity and the functional role of a specific organism in its environment. It is therefore important to remain critical about the hereby obtained results and to further invest in the verification of the proposed hypotheses. The information held by the novel soda lake MAGs could be used for the targeted enrichment and isolation of novel consortia and organisms, to ultimately prove their activity under specific conditions. Absolute abundances of uncultivated organisms and functional genes can be more accurately quantified by means of whole-cell analysis such as fluorescence *in situ* hybridization (FISH) and quantitative PCR (qPCR). Finally, the datasets provided here can be mined for novel genes with biotechnological potential when heterologously expressed.