universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Adaptation of two microbial communities to low water potential

verfasst von / submitted by

Grdo. Juan Bosco Gracia Alvira

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 830
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molecular Microbiology, Microbial Ecology and Immunobiology
Betreut von / Supervisor:	Assoz. Prof. Dr. Dagmar Wöbken

Adaptation of microbial communities to low water potential

ABSTRACT

Desiccation and hyper-salinity are often treated as the same stress since both lead to the reduction of bioavailable water in the environment. Microorganisms adapted to these stresses have developed survival strategies like sporulation, metabolic dormancy, biosynthesis and accumulation of compatible solutes, expression of catalases, ROS scavengers and organisation in multi-species stratified biofilms. However, it is still not clear which of these adaptations are specific to either desiccation or hyper-salinity and which are common to both stresses. Our primary goal was to gain insights into the adaptation strategies of microorganisms adapted to arid and hyper-saline environments. We investigated two microbial communities that endure low water bioavailability: a desert biological soil crust and a hyper-saline microbial mat. We used FISH to evaluate the changes in the ribosomes content in cells from the desert biological soil crusts upon hydration. We saw increase in the FISH detection rate and the signal intensity throughout the 12 h incubation period, suggesting increase in the protein synthesis potential upon hydration. We looked at the spatial distribution of Archaea, Bacteria and Alphaproteobacteria throughout the depth profile of two hyper-saline microbial mats with different salinity levels using CARD-FISH. We compared the relative abundance of the taxa of study with metagenomic reads estimations from the same samples. Relative abundance diverged between CARD-FISH and metagenomic reads estimations, but both methods detected a decrease in the Alphaproteobacteria relative abundance in the middle layer of the mat under low salinity. Finally, we searched for potential habitat-specific adaptation strategies in closely related genomes sequenced from arid and saline environments. Although phylogeny was more important than adaptation to the environment in terms of genomic content, we found several strategies shared by closely related bacteria from both habitats, like ion transport, and strategies that were characteristic from hyper-saline bacteria, like compatible solutes transport. From our work we can conclude that desiccation and hyper-salinity are different stresses and have different impact on the genome, despite leading both to lower water activity.

Adaptation of microbial communities to low water potential

ZUSAMMENFASSUNG

Austrocknung und erhöhter Salzgehalt werden in der Wissenschaft oft als dieselbe Belastung behandelt, da beide zu einer Verringerung des bioverfügbaren Wassers in der Umwelt führen. Mikroorganismen, die an diese Belastungen angepasst sind, haben Überlebensstrategien entwickelt wie Sporulation, Stoffwechselruhe, Biosynthese und Akkumulation kompatibler gelöster Stoffe, Expression von Katalasen, ROS-Fängern und die Organisation in geschichtete Biofilme mit mehreren Spezies. Es ist jedoch immer noch nicht klar, welche dieser Anpassungen genau für die Austrocknung und den hohen Salzgehalt verantwortlich sind und zu welcher der beiden Belastungen sie beitragen. Unser Hauptziel war es, Einblicke in die Anpassungsstrategien von Mikroorganismen zu gewinnen, die an trockene und hyper-saline Umgebungen angepasst sind. Wir untersuchten zwei mikrobielle Kulturen, die einer geringen Wasserverfügbarkeit ausgesetzt sind: eine biologische Bodenkruste in der Wüste und eine mikrobielle Matte in hyper-salinem Gebiet. Wir verwendeten die Methode FISH, um die Veränderungen des Ribosomengehalts in den Zellen der biologischen Wüstenbodenkruste nach erfolgter Hydratation zu analysieren. Die FISH-Erkennungsrate und die Signalintensität nahmen während der 12-stündigen Inkubationszeit zu, was auf eine Zunahme des Proteinsynthesepotenzials nach der Hydratation schließen lässt. Des Weiteren, untersuchten wir die räumliche Verteilung von Archaea, Bakterien und Alphaproteobakterien im Tiefenprofil von zwei hyper-salinen mikrobiellen Matten mit unterschiedlichem Salzgehalt mittels CARD-FISH. Wir verglichen die relative Häufigkeit der untersuchten Taxa mit den Schätzungen der metagenomischen Daten aus denselben Proben. Die relativen Häufigkeiten von CARD-FISH und die metagenomischen Daten wichen voneinander ab, aber beide Methoden zeigten eine Abnahme der relativen Häufigkeit von Alphaproteobakterien in der mittleren Schicht der Matte bei niedrigem Salzgehalt. Schließlich suchten wir in eng verwandten Genomen, nach potenziellen habitatspezifischen Anpassungsstrategien, die zuvor in ariden und salzigen Umgebungen sequenziert wurden. Obwohl die Phylogenie mehr Einfluss auf den genomischen Inhalt hat als die Anpassung an die Umwelt, fanden wir mehrere Strategien, die eng verwandte Bakterien aus beiden Lebensräumen gemeinsam haben, wie z. B. den Ionentransport. Aber wir konnten auch spezifischere Mechanismen für hyper-saline Bakterien entdecken wie z. B. den Transport kompatibler gelöster Stoffe. Unsere Arbeit zeigt, dass Austrocknung und Übersalzung unterschiedliche Arten der Belastung sind und auch unterschiedliche Auswirkungen auf das Genom haben, aber dennoch beide zu einer geringeren Wasseraktivität führen.

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Dr. Dagmar Wöbken, for providing me with useful feedback. I would also like to thank Dr. Dimitri Meier and Dipl.-Biol. Stefanie Imminger for the the invaluable guidance and knowledge they have provided me with throughout my work.

I am also grateful to the wonderful people from the soil group and from the Division of Microbial Ecology for their interesting discussion and feedback as well as for always making me feel very welcome during my time working in the department.

Finally, I would like to express my gratitude to my family and friends for all their support during my academic education.

Index

Chapter 1

Introduction

1.1 Microbial life under low water activity

Water is essential for biological processes such as protein folding, enzymatic performance and cell wall integrity [1]. Nevertheless, approximately 35% of the Earth's surface corresponds to drylands, which are populated by microbial life [2]. Microorganisms inhabiting these harsh habitats are termed "xerotolerant", from the Greek *xēros* (dry) [3]. Microorganisms inhabiting saline and hyper-saline environments have been described as "xerotolerant" due to the fact that water is not biologically available in these environments [4]. Lack of water has physical similarities with hyper-osmolarity since both stresses lead to a reduction of the water activity [5]. Furthermore, both stresses are often not distinguished [6]; desiccation in laboratory conditions is simulated by the addition of high concentration of salt to liquid medium [7]. However, desiccation and high salt concentration have been reported to be different stresses [8, 9]. Despite the low water potential, the presence of a solute in hyper-saline medium still allows for the transport of microorganisms and the diffusion of nutrients, which is not possible in arid environments [6]. For this reason I will differentiate between xerotolerant (i.e. adapted to high aridity) and halo-tolerant (i.e. adapted to high salinity) microorganisms throughout this thesis.

1.2 Adaptation strategies to low water activity

Low water activity due to either lack of water or hyper-salinity leads to stresses such as loss of membrane integrity, disruption of the ion potential across the membrane, reduction of turgor pressure and loss of cytoplasmic volume, accumulation of reactive oxygen species (ROS), denaturation of proteins and DNA damage [3]. In many environments, low water activity is coupled to other stressors (i.e. high UV radiation, extreme temperature and nutrient limitation), which pose additional difficulties to cell survival. Microorganisms inhabiting xeric and hyper-saline environments have developed a repertoire of different strategies, that act sometimes simultaneously, to overcome these stresses. However, adaptations to low water activity have been only partially

characterised to date [10]. Although it is known that desiccation and hyper-salinity induce different responses in the cells [8], further studies are needed to understand which microbial adaptations are specific to lack of water or high salinity and which ones are common to both stresses.

Sporulation and metabolic dormancy are common responses to water deprivation [11, 12]. Arid soil microorganisms experience extended periods of desiccation. In response, some bacterial species undergo differentiation into highly resistant spores which are protected from DNA damage and can endure nutrient starvation for long periods of time [13]. Non-sporulating prokaryotes can enter a quiescent state in which cell growth and division are stopped and energetically costly biological processes, such as protein synthesis, are reduced drastically [14]. This state is called metabolic dormancy. In order to decrease protein synthesis, some dormant bacteria carry out ribosome hibernation. Ribosomes cease their activity upon association to small peptides called hibernation factors. Ribosome modulation factor (RMF) and hibernation promoting factor (HPF) induce the dimerisation of 70S ribosomes into inactive 100S dimers, whereas RaiA stabilises 70S monomers in an inactive state [15]. As a result, global protein synthesis and energy consumption decrease drastically but can be rapidly resumed upon changing conditions [15].

Low water activity, as well as high UV radiation, induce reactive oxygen species (ROS), which damage proteins, lipids and DNA [16, 17]. Microorganisms adapted to these stresses encode several proteins to overcome ROS damage such as catalases, thioredoxin, DNA-binding ferritin, and the ROS scavenger super oxide dismutase (SOD) [18, 19, 20].

Accumulation of compatible solutes is a widely used strategy to overcome low water activity [1]. Under hyper-osmolar conditions, intracellular accumulation of small organic molecules (osmolites) and ions restores the osmotic equilibrium, allowing the correct folding of proteins and inhibiting water removal and membrane disruption [21]. Additionally, under extreme desiccation conditions some compatible solutes such as trehalose and glycerol can replace water inside the cytoplasm, maintaining the stability of intracellular macromolecules in a state of "vitrification" [22]. Osmolites can be carbohydrates (trehalose or sucrose), polyols (mannitol), amino acids (proline) or trimethylammonium compounds, such as glycine betaine. They can be either synthesised *de novo* or taken from the environment via proteic transporters [23]. An alternative strategy to compatible solutes accumulation is the so called "salt-in" strategy, that consists in the accumulation of potassium chloride (KCl) in the cytoplasm to match the extracellular osmolarity [24]. The salt-in strategy consumes less energy but requires the global adaptation of the cell to high KCl concentration (e.g. proteins have a lower isoelectric point [25]), making it unable to grow under low salt conditions [24].

Changes in external water availability affect severely the cellular membrane of bacteria [26]. Some bacteria cope with high temperatures and low water activity by increasing the ration of trans to cis fatty acids and the proportion of negatively charged phospholipids [27]. These modifications help to preserve the lipidic bilayer in an ordered phase [28]. Cell wall can be modified upon desiccation and have been reported to be thicker [29].

Besides unicellular survival strategies, microorganisms adapted to high aridity and high salinity can organise in biofilms at community level [30, 31]. Biofilms are thought to be the first line of defence against xeric stress. Algae, fungi, bacteria and archaea adapted to xeric habitats co-inhabit multi-species biofilms. Most of the biofilm's biomass is composed of extracellular polysaccharides (EPS), that hold water and provide a substrate for the community [32, 33].

1.3 Microbial organisation in xeric and hyper-saline environments

Microorganisms inhabiting extreme environments organise in communities in order to trap all necessary resources of energy and matter in an inwardly oriented diffusive cycle [31]. In habitats like hot springs, salt lakes and man-made salterns, microorganisms organise forming phototrophic microbial mats: stratified, small-scale, self-sustained ecosystems that comprise the major element cycles, trophic levels and food webs [34]. Microbial mats are vertically stratified due to physicochemical gradients that are generated and maintained by microbial activity [35]. Each layer of the mat provides different micro-environments and biological niches [35]. Cyanobacteria are the dominant primary producers; they harvest light as source of energy and fix CO_2 to synthesise organic matter, thus sustaining metabolically diverse heterotrophs [36]. Additionally, cyanobacteria are responsible for the production of an EPS matrix that embeds the community and provides protection against desiccation [33]. Key nutrients such as sulphur, carbon and oxygen are efficiently recycled within the mat ecosystem [36]. Sulphate-reducing bacteria, sulphide-oxidising bacteria and anoxygenic phototrophic bacteria are vertically stratified according to microgradients of oxygen, sulphide and light [35].

In arid environments such as desert soils, microorganisms associate with macroscopic species (algae, microfungi, lichens, mosses...) to form biological soil crusts (BSCs) [30]. Filamentous Cyanobacteria are pioneers in the BSC formation, as they form the matrix of BSC by aggregating soil particles together and secreting EPS [32, 37]. Furthermore, they supply organic carbon to the heterotrophic community members [38]. However, the potential for CO_2 fixation has also been reported in anoxygenic phototrophic bacteria and H_2-oxidising chemolithoautotrophs [39]. BSCs are stratified due to high levels of solar radiation; pigmented species occupy the upper part of the crust and quench UV radiation to non-pigmented filamentous cyanobacteria [32]. Association into BSCs helps microorganisms to tolerate extreme dehydration thanks to the high water holding capacity of secreted EPS. Additionally, BSCs act like "dust traps", since surface roughness and EPS increase the capture of nutrient-rich particles, providing nitrogen, phosphorus and potassium to the community [32].

Despite their differences in structure and biochemical processes, hyper-saline microbial mats and BSCs share filamentous cyanobacteria as primary producers; the genera *Microcoleus* and *Coleofasciculus* (formerly classified as *Microcoleus chthonoplastes*) are common and dominant taxa in both environments [34, 30].

1.4 Methods to study xerotolerant and halotolerant microbial communities

Our knowledge on microorganisms adapted to low water activity and how they organise in communities and cope with environmental stress comes from several approaches, sometimes used simultaneously. For decades, microorganisms have been cultivated and subjected to high salinity or to desiccation [1, 7, 8]. This approach allows to test which genes are associated with the stress response and their mechanism of action. However, only a fraction of the microbial community can be cultured [40] and this approach does not provide information about the relationship between community members [41].

Several culture-independent methods (e.g. Raman, NanoSims, BONCAT and FISH) allow to study the role and distribution of the microorganisms *in situ*, as well as how they interact with each other and with the surfaces they colonise [42]. These methods have been used in the research of microorganisms, reaching single cell resolution. Raman spectroscopy [43] and high resolution secondary ion mass spectrometry (NanoSIMS) [44] are techniques that provide information about the molecular composition of the samples [45]. Both methods have been used in microbial mats to study sulphur storage at single cell resolution [46] and the functions and metabolism of Chloroflexi [47]. Protein synthesis, and therefore biological activity, can be quantified using BONCAT (bio-orthogonal non-canonical amino acid tagging). This method has been used to study microbial metabolic activity in soil at single cell resolution [48].

Fluorescence *in situ* hybridisation (FISH) has become a widely used cultivation-independent single-cell method in the study of microbial communities [41]. It allows the identification, quantification and characterisation of microbial populations in complex environments [41]. In FISH, the sample is incubated with fluorescently labelled oligonucleotides that bind to the ribosomal RNAs (rRNA) and only the cells whose rRNA matches the probe will emit a fluorescent signal [49]. This enables the design of oligonucleotide probes matching a wide range of taxonomic entities: from the general Bacteria and Archaea-specific probes [50, 49] to species-specific and even strain-specific probes [51], capturing the phylogenetic diversity of the community. The FISH signal is proportional to ribosomal amount [52, 53], which can range from a few hundred ribosomes to tens of thousands depending on the taxon and the state of the cell [41]. Therefore, the FISH signal intensity among stained cells can be highly variable within a sample. Catalysed reported deposition (CARD)-FISH is an alternative to conventional FISH that improves the sensitivity of the method [54]. In CARD-FISH, probes are linked to horseradish peroxidase (HRP), which catalyses the deposition of fluorescently labelled tyramide molecules resulting in a strong fluorescent signal even in cells with few active ribosomes [54, 55].

Although FISH-based methods do not provide information about which functions are carried out by the targeted microorganisms *per se*, they can be used simultaneously with Raman microspectroscopy, NanoSIMS and BONCAT to link the metabolic activity with phylogenetic identity of the community members [46, 47, 56]. Ribosomal RNAs also determine the protein synthesis potential of cells and thus rRNA measurements can be used as an indicator of potential biological activity (termed "protein

synthesis potential") [57]. Since FISH probes binding to a cell are proportional to the concentration of rRNA molecules, FISH signal intensity can be used to quantify ribosomal amount in the whole microbial community and in single cells [58]. The potentially active fraction of the microbial community can then be calculated as the proportion of cells that display a FISH probe over the total number of cells [12, 59, 60].

Besides cultivation-based and single-cell methods, molecular genetics can provide insights into microbial community diversity, metabolism and potential survival mechanisms in hostile ecological niches. 16S rRNA gene amplicon sequencing has become the most common method to study community composition, revealing the enormous diversity of microbial communities [61, 62]. This high throughput method allows the identification of thousands of phylotypes in a single sample, but abundance estimations might not be fully reliable due to biases introduced during DNA extraction and PCR [63], to variable number of copies of the 16S rRNA gene per genome [64] and to poliploidy (i.e. multiple genome copies per cell) [65]. Furthermore, amplicon sequencing relies on the amplification of a single marker gene, usually a region of the 16S rRNA gene. Thus, amplicon sequencing studies do not provide information on the survival strategies of identified microbial community members, in contrast to metagenomic approaches. Metagenomic surveys in hyper-arid and hyper-saline environments have shown different potential mechanisms by which bacteria could overcome stress in harsh environments [3]. By binning the metagenomes it is possible to associate these mechanisms to specific taxa [39]. Since habitat adaptations are accessory (i.e. they do not belong to the core genome) [66], tolerant bacteria must be genetically equipped with xeric stress resistance strategies that are absent in their close relatives adapted to "physiological conditions", even if they belong to the same genus [67, 68]. Studying the pangenome (i.e. the sum of all the genes present in at least one genome within a clade [67]) allows the identification of potential stress resistance strategies which are characteristic to an environment [68]. However, this approach has never been used to compare the genomic potential and survival strategies of bacteria adapted to arid and hyper-saline environments.

1.5 Aim of the project

Here we aimed to gain insights into the adaptation of microbial communities to low water activity. Desiccation and high salinity have been traditionally treated as the same stress. The main objective of the master thesis was to further characterise two microbial communities adapted to these stresses: a desert BSC and a hyper-saline microbial mat, in order to test whether microorganisms adapt in the same way to aridity and hyper-salinity. We used FISH to evaluate the changes in ribosomal content in BSC cells throughout a 12 h period upon a hydration event. Further, we studied the effect of long-time storage in BSC ribosomal dynamics. We quantified the relative abundance of Bacteria, Archaea and Alphaproteobacteria from two microbial mats undergoing different levels of salinity with CARD-FISH and compared the results to abundance estimations based on metagenomic data. Finally, we used the Anvi'o analysis and visualisation platform to study the functions encoded in closely related genomes sequenced from arid and saline habitats. We looked for stress resistance strategies characteristic of each environment using a MAG-centric strategy.

Chapter 2

Materials and Methods

2.1 Sample collection

2.1.1 Biological Soil Crusts sample collection

I was provided with the biological soil crusts (BSCs) analysed in Chapter 3. Briefly, samples were collected in September 2019 at the Avdat long-term ecological research site located in the Negev Desert, Israel (N 30°36'33", E 34°44'48"). The site has a yearly precipitation of ca. 90-100 mm and is characterised by Cyanobacteria-dominated biological soil crusts covering the soil between shrub plants [39]. Crust pieces were collected in Petri dishes, sealed with parafilm and stored in the dark at 18-20 ºC and .ca 40% humidity until further use.

2.1.2 Hyper-saline Oman mats sample collection

I was provided with the microbial mats material used in Chapter 4. Briefly, samples were collected from an intertidal area in Oman (20°45'39.6"N 58°38'52.3"E) in February 2018 [69]. The mats were sampled from two different sites: "Site 1" had more than 30% salinity, no tides and salt crust on top of the mats [69] whereas "Site 2" had 18% salinity and regular tides (Site 2 corresponds to "O4" site in Vogt *et al.* 2018) [70]. Three mats were sampled from each site: Mat II, III and V at the site with salt crust, and Mat VII, VIII and IX at the site without crust. Only one mat per site was used for CARD-FISH (Section 2.2.3), but metagenomic reads estimation was based on the average abundance of the three mats from each site.

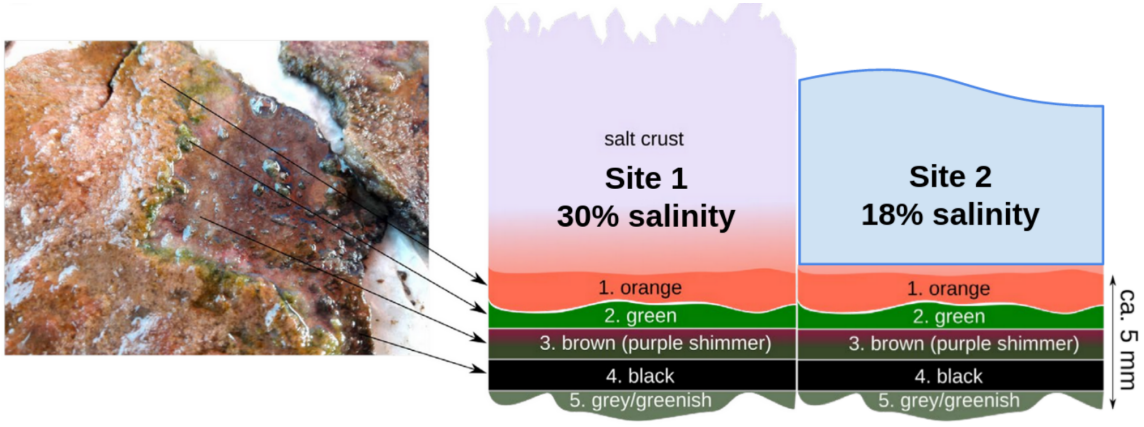

Figure 2.1: Piece of microbial mat with layers partially scraped off (left). Scheme of the layers in the high and the low salinity mats (right). Figure adapted from Meier *et al.* 2021 [69]

2.2 Sample preparation

2.2.1 Incubation and fixation of BSCs

Crust pieces with a dry weight of ca. 1 g were put in Petri dishes, hydrated up to a water content of 26% (w/w), corresponding to 75% of the maximum water holding capacity and sealed with parafilm. The pieces were incubated in a climate chamber (FitoClima 600 & 1200 Bio, Aralab, Lisbon, Portugal) mimicking "day" conditions (light, 27 ºC). Crust pieces were incubated for a determined time, one crust for each time point: 15 minutes, 30 minutes, 3 hours and 12 hours. Additionally, a crust piece was fixed without previous hydration and incubation (time point 0 h).

After incubation, the crust pieces were fixed in 4% PFA in 1x phosphate buffered saline (PBS), pH = 7.6. Each crust piece was placed in a 50 mL polyethylene tube and incubated at RT for 1 h in 6 mL of fixation solution. The polyethylene tube was centrifuged (Eppendorf centrifuge 5804R, Germany) at 10,000 g for 5 min at 4ºC and the supernatant discarded. The pellet was washed in 1x phosphate buffered saline (PBS) pH 7.6 three times (10 mL of PBS, mixing and centrifugation: 10.000 g, 5 min, 4 ºC). The pellet was re-suspended in 10 mL EtOH/PBS (1:1) and frozen at -20ºC until further use.

2.2.2 Cell detachment, Nycodenz density gradient separation and filtration of BSCs

Fixed soil crust samples were washed twice (10000 g, 5 min, 4 °C) with 10 mL of 1x PBS. Cell detachment was performed according to Eichorst et al. 2015 using the "combinational treatment" [45]. Afterwards, the samples were additionally sonicated ten times for 10 s at 35 kHz in a sonication bath (Sonorex, Bandelin, Berlin, Germany).

The samples were added on equal volumes of 1.42 g/mL Nycodenz (Axis-shield PoC, Oslo, Norway) and centrifuged with an Ultracentrifuge (Beckman, Indianapolis, USA) at 10,000 rpm for 90 min at 4 °C. After centrifugation, the aqueous upper layer of the gradient containing the extracted cells and the greenish interphase containing the cyanobacteria were collected and filled up to 50 mL with 1x PBS.

Aliquots of the cell suspensions were filtered onto 0.2 μm pore-size polycarbonate filters (diameter 25 mm, Millipore GTTP 02500, Eschborn, Germany). The volume of cell suspension that was filtered varied between samples and ranged between 5 mL and 10 mL. Aliquot volume for each sample was chosen empirically based on optimal cell density for visualisation on the filters (i.e. abundant, non-overlapping cells were visualised via staining with 4',6-diamidine-2'-phenylindole dihydrochloride (DAPI) as explained in Section 2.3.3). The filters were air-dried and stored at -20 ^{o}C until they were used for FISH (Section 2.3.1).

2.2.3 Fixation and filtration of hyper-saline mats

Mat samples were provided to me already fixed in PFA. Fixation and filtration are described in Meier *et al.* 2021 [69]. The salt crust was washed from the mat surface, each mat layer was separated with a scalpel according to the colour: orange, green, brown, black and grey and separated in different tubes. The material was fixed in 4% PFA in 1x PBS (pH = 7.6) for 90 minutes at RT, washed twice with 1x PBS and stored in an ethanol:PBS mixture (1:1) at -20 °C until further analysis.

Material from all the layers of the mats III (30% salinity) and VIII (18% salinity) were filtered. For the filtration, 100-200 mg of mat material from each layer were suspended in 1 mL PBS:EtOH (1:1), sonicated with a UW 2070 probe (Bandelin Electronic, Berlin, Germany) for 30 seconds (1 pulse/second, 20% intensity) and briefly span. 1 mL of supernatant was collected and replaced by 1 mL of PBS:EtOH (1:1). The procedure was repeated six more times, collecting 7 mL of cell suspension per sample. Thirty μL of each cell suspension were filtered onto 0.2 μm pore-size polycarbonate filters (diameter 25 mm, Millipore GTTP02500, Eschborn, Germany). Cell density was inspected on filter pieces stained with 4',6-diamidine-2'-phenylindole dihydrochloride (DAPI) as explained in Section 2.3.3. The filters were air-dried and stored at -20 ^{o}C until they were used for CARD-FISH (Section 2.3.2).

2.3 Microscopy

2.3.1 Fluorescence *in situ* hybridisation on BSC filters

Sections of the BSC filters were cut with scissors and hybridised with FITC mono-labelled EUB338 I-III and NON338 probes as described in Glöckner *et al.* 1996 [71]. Shortly, filter sections were placed facing up on a glass slide and each section was covered with a droplet of hybridisation solution (for each section, 3 μL of probe 50 ng DNA/μL were mixed with 27 μL of hybridisation buffer). A humidity chamber was prepared by putting a piece of blotting paper into a 50 mL polyethylene tube and soaking it with 2.5 mL of hybridisation buffer. The slide was placed inside the humidity chamber horizontally. The closed chamber was incubated for 3 h at 46 ^{o}C. Hybridised sections were transferred to a polyethylene tube with 50 mL of preheated washing buffer (48 ^{o}C) and incubated at 48 ^{o}C for 15 min. Some modifications were made to the original protocol: SDS concentration in hybridisation and washing buffers was increased to 0.02% and the hybridisation time was extended to 3 h. Sequence, target organisms, hybridisation conditions and references of the probes used for FISH are shown in Table 2.1.

2.3.2 Catalysed reporter deposition (CARD)-FISH on hypersaline mat filters

Endogenous peroxidases were inactivated by incubating mat material filters in H_2O_2 (0.15% in methanol) for 30 min at RT. Sections of the peroxidase-deactivated filters were cut with scissors and used for CARD-FISH as described in Ishii *et al.* 2004 [72] with some modifications: embedding step was skipped since it resulted in cell aggregation and cell loss due to permeabilisation was not significant (personal communication, Dr. Dimitri Meier - ETH Zürich), and an additional permeabilisation step was included for detection of Archaea. Shortly, filter sections were incubated in freshly prepared lysozyme solution (10 mg/mL in 0.05 M EDTA, pH 8.0; 0.1M Tris-HCl, pH 8.0) for 60 min at 37°C and then washed in MQ H_2O. Filter pieces hybridised with the archaeal probe ARC915 were additionally permeabilised in HCl (0.1 M in H_2O) for 1 min at RT and then washed in MQ H_2O. Hybridisation of the filter pieces was carried out by immersing them in an Eppendorf tube filled with 600 μL of hybridisation solution (2 μL of HRP probe 50 ng/μL in 598 μL hybridisation buffer) for 3 h at 46^{o}C. After the hybridisation, filter sections were transferred to a polyethylene tube with 50 mL of pre-warmed washing buffer (48 ^{o}C) and incubated at 48 ^{o}C for 15 min. Sections were incubated for another 15 min in 1xPBS pH = 7.6 at RT (in order to equilibrate the HRP). Amplification was carried out by immersing the filter pieces in an Eppendorf tube filled with amplification solution for 45 min at 46 ^{o}C in the dark. Amplification solution was prepared by mixing 1000 μL of amplification buffer with 10 μL of H_2O_2 0.15% v/v in 1xPBS pH 7.6 and 1 μL of fluorescent labelled tyramide (Alexa Fluor^{TM} 594 Tyramide Reagent, Thermo Fisher Scientific, Massachusetts, USA). Finally, amplified filter sections were washed in excess MQ H_2O, subsequently in EtOH and then air-dried. Sequence, target organisms, hybridisation conditions and references of the probes used for CARD-FISH are shown in Table 2.1.

Probe	Sequence (5' to 3')	Target	Formamide (%) at 46 °C
EUB338 I-III[ab] [73, 50]	GCTGCCTCCCGTAGGAGT (I) GCAGCCACCCGTAGGTGT (II) GCTGCCACCCGTAGGTGT (III)	Bacteria	35
ARC915[b] [49]	GTGCTCCCCCGCCAATTCCT	Archaea	20
ALF968[b] [74]	GGTAAGGTTCTGCGCGTT	Alphaproteobacteria	20
CFX1223[b] [75]	CCATTGTAGCGTGTGTGTMG	Chloroflexi	35
EHB412[b] [76]	TACGCCCCATAGGGGTGT	Rhodothermales	45
NON338[ab] [77]	ACTCCTACGGGAGGCAGC	Negative control	35

Table 2.1: Oligonucleotide probes used for FISH (a) and CARD-FISH (b)

2.3.3 DAPI staining

Filter sections were incubated in DAPI solution (1μg/mL in H_2O) for 3 minutes at RT. The filters were washed in MQ H_2O for 1 minute, dehydrated in 100% EtOH for 1 minute and air-dried.

2.3.4 Microscopic visualisation of the cells

The filters were mounted in the non-hardening and anti-bleaching mounting medium CitiFluor^{TM} AF1 (Citifluor, Ltd., London, United Kingdom) and inspected under a Zeiss Axio Imager 2 (Carl Zeiss, Jena, Germany) with a 100x oil objective and a set of filters for DAPI (405 nm), FITC (480 nm) and Alexa Fluor 594 (660 nm) dyes. For image acquisition in 660 nm (samples hybridised with CARD-FISH protocol) and the DAPI channel, exposure time was set as the maximum value at which the image did not show saturation. Exposure time was set to 300 ms in the FITC channel (samples hybridised with FISH protocol) in order to compare the signal intensity of BSC filters of different incubation time points.

2.3.5 Image analysis

FIJI [78] was used to count the DAPI, as well as FISH and CARD-FISH signals (of BSC and mat samples, respectively) in at least ten images for each filter. More than 2500 cells were counted per filter except in filters corresponding to Oman mat layers 3 and 4, in which only 500 cells could be counted. A semi-automatic approach was used for counting in order to reduce the workload and the bias. The function "Find maxima" was used with a prominence > 25 to select cell candidates in both channels, that were then inspected visually. Signals in the FISH channel were checked to have a corresponding DAPI signal. Auto-fluorescent particles (i.e. particles detected in the FISH channel but not in the DAPI channel) were deselected and adjacent cells detected as a single maximum were corrected. The final list of detected cells in the FISH and DAPI channels were saved as Regions of interest (ROIs) for tractability.

Relative abundance of FISH-targeted cells was calculated as FISH counts divided by DAPI counts.

For filters of BSC, the software *daime* v2.2 [79] was used to measure the signal intensity of all FISH-targeted cells on each picture. Cells were identified by dark-light contrast with an object size threshold of 28 and including internal dark regions into the object area. Probe clumps (visualised as high intensity particles), that were detected as cells automatically by *daime*, were removed manually.

2.3.6 Statistics

Data sets did not follow a normal distribution according to Shapiro test. For this reason, non-parametric Kruskal-Wallist test and pairwise Wilcoxon rank-sum test ($\alpha = 0.05$) were used to evaluate statistically significant differences. False discovery rate was set as p-values adjustment method in Wilcoxon test. Analysis of similarities (ANOSIM) function from the package "vegan" [80] was used with the following settings: 9999 permutations and Bray-Curtis distance metric. The statistical analysis was done with R [81] in RStudio [82].

2.4 Pangenomic analysis

2.4.1 Data availability

All the isolate genomes and metagenome-assembled genomes (MAGs) used in the pangenomic analysis (Chapter 5) were obtained either from NCBI [83], from Bay *et al.* 2021 [84] or provided by Dr. Dimitri Meier (Supplementary table S1). In all the cases they were already assembled and binned. Twenty-four out of 41 genomes were sequenced from the same biological soil crust and hyper-saline mat samples described in Section 2.1 (Supplementary table S1). Two genomes originated from isolates cultured from the BSC material (previously cultured on Avdat desert soil extract agar plates). Genomes were sequenced with the Illumina MiSeq chemistry and assembled with SPAdes v.3.11.0 [85] using standard settings at the Joint Microbiome Facility. Sequencing, assembly and binning workflows for BSC MAGs can be found in Meier *et al.* 2021 [39]. Hyper-saline mat material was sequenced with Illumina NovaSeq S4 sequencer in 2x150 bp mode. Assembly and annotation were performed by the standard automated pipeline of the JGI and loaded into the Integrated Microbial Genomes and Metagenomes (IMG/M) platform for exploration and systematic gene search. The binning was done the same way as for BSC MAGs [39]. Following method section was provided by the JGI in the assembly report:

"*Reads were corrected using bbcms v. 38.34.(1) This was run using the following command line options: bbcms.sh metadatafile=counts.metadata.json mincount=2 highcountfraction=0.6 in=out.fastq.gz out=input.corr.fastq.gz. The readset was assembled using metaSPAdes assembler with metaspades 3.13.0. [86] This*

was run using the following command line options: spades.py -m 2000 –tmp-dir scratch -o spades3 –only-assembler -k 33,55,77,99,127 –meta -t 72 -1 reads1.fasta -2 reads2.fasta. The input read set was mapped to the final assembly and coverage information generated with bbmap 38.34. [87]. This was run using the following command line options: bbmap.sh nodisk=true interleaved=true ambiguous=random in=out.fastq.gz ref=assembly.contigs.fasta out=pairedMapped.bam covstats=covstats.txt bamscript=to_bam.sh."

OTU table from BSCs collected in 2017 and 2019 referred to in Section 3.2.2 and metagenomic reads table used for estimating hyper-saline mats taxa abundance (Chapter 4) were provided by Dr. Dimitri Meier.

2.5 Taxonomic classification of the genomes

The tool GTDB-Tk v.0.3.3 [88], from the Genome Taxonomy Database [89], was used to place the genomes into the GTDB phylogenomic tree and determine the closest related genome in the database based on concatenated alignment of 120 marker protein sequences.

2.6 Genome completion estimation

CheckM v1.1.3 [90] was used to assess the quality of the genomes. The software evaluates the quality of the genomes using lineage-specific single copy marker genes. Completion is inferred from the proportion of marker genes found in the genome, whereas contamination is calculated based on the presence of multiple copies of these genes. Strain heterogeneity refers to the proportion of the multiple copy genes (contamination) that belong to closely related strains to the query genome based on the aminoacid identity (AAI). The Lineage-specific Workflow "checkm lineage_wf" was used with default parameters. This workflow identifies the most informative lineage-specific marker set suitable for evaluating each genome.

2.7 Genomes annotation and function prediction

MAGs retrieved from NCBI [83] were downloaded already annotated in GenBank file format. Prodigal v2.6.3 [91] was used to identify the coding sequences in the rest of the genomes (Supplementary Table S1). Predicted genes were scanned against Pfam [92] and KEGG [93] databases using using hmmscan from HMMER v3.2.1 and against COG database [94] using DIAMOND alignment [95]. Additionally, EggNOG-Mapper [96] was used to allocate each encoded protein to a group of orthologs in the EggNOG database [97]. For all the databases, E-value cut-off was set to 10^{-5}.

2.8 Computing of Cyanobacteria and Alphaproteobacteria pangenomes

Anvi'o pangenomic workflow [98, 99] was used to compute and visualise the Cyanobacteria and Alphaproteobacteria pangenomes. For each pangenome, a database containing the DNA and amino acid sequences of each annotated genes from each genome, as well as their predicted functions, was generated. The pangenomes were computed using the program "anvi-pan-genome" with the flag "–use-ncbi-blast" and default parameters (minbit = 0.5, inflation = 2, minimum sequence identity = 0%). "anvi-pan-genome" calculates alignment similarity of all the amino acid sequences in the database pairwise using BLASTp [100] and removes weak hits using "minbit heuristic" [101]. Minimum sequence identity is set to 0% because the minbit parameter is a better predictor for weak matches, since it takes into account the alignment length between the two sequences. MCL (Markov Cluster algorithm) [102] is used to associate the remaining hits into groups of homologs. As a result, "anvi-pan-genome" generates clusters of homologous genes. Each gene cluster can be present or absent across the analysed genomes. By default, genomes are sorted by hierarchical clustering based on gene clusters distribution using Euclidean distance and Ward clustering. The program "anvi-display-pan" was used to visualise the pangenome in an interactive interface.

2.9 Functional enrichment

The function "anvi-get-enriched-functions-per-pan-group" was used to study functions that are predominantly found in genomes from one habitat. Briefly, each Anvi'o gene cluster is associated to the most frequent COG function among its containing genes. Gene clusters associated to the same COG number are merged into a single function. A frequency table of functions in genomes is built and used as input to compute an enrichment score and a p-value for each function. Enrichment score is calculated using the Rao test. False Detection Rate correction to p-values is done using the R package qvalue. Adjusted p-value threshold was set to 0.05. The results indicate if a function is more frequent within a defined group of genomes than on average in the whole dataset.

2.10 Average Nucleotide Identity calculation

The program pyANI v0.2.10 [103] was used to calculate the average nucleotide identity (ANI). For ANI calculation, genomes are aligned pairwise by cutting the first genome (query) into consecutive fragments, that are then matched against the second genome (reference). ANI is calculated as the average percentage nucleotide identity of the matching regions [104]. The flag "-ANIb" was set in order to use BLASTN+ [105] as alignment algorithm and 1020 nucleotides long fragments as input.

Chapter 3

Measuring rRNA content in BSC cells upon hydration

Microbial dormancy is a widespread strategy that allows microorganisms to survive harsh conditions [12] by minimising basal energy consumption in order to survive for longer time. In arid ecosystems, water deprivation leads to energy limitation due to the decreased availability of dissolved organic carbon [13]. Protein synthesis by the ribosomes is one of the most energy consuming processes in the cell; ribosomes can constitute up to 45% of the total mass of the bacteria and their biosynthesis consume up to 40% of the produced energy [106]. During starvation, some microorganisms can reduce their ribosome numbers [107] while other microorganisms maintain their ribosomes in an inactive state [108] or even increase them in the case of specific akinetes [109]; therefore, ribosomal content cannot be used as a direct proxy of biological activity. However, rRNA content directly relates to the potential to catalyse protein synthesis of the microbial community [57].

It was the goal to investigate the ribosome content of BSC cells during rehydration by performing mono-labelled FISH with the general Bacteria probe EUB338 I-III. Crust pieces were hydrated and incubated for different time points, cells were fixed, detached from the crust and filtered as explained in the materials and methods chapter (Sections 2.2.1 and 2.2.2).

3.1 Results

3.1.1 FISH detection rate on BSC cells was only 20% after 12 hours hydration

FISH signal was quantified by calculating the percentage of the total cells (visualised with DAPI) that had an overlapping FISH signal. The proportion of cells detected with FISH upon hydration increased with the incubation time (Figure 3.1). Only 0.04% of the cells had a FISH signal in the time point 0 h, in which the crust was not hydrated. The percentage increased to 0.14% 15 minutes after hydration, to 7.16% after 30 minutes and reached 18.52% after 3 hours. The value remained stable 12 hours after hydration, with 20.66% of the cells having a FISH signal. The increase in the detection rate from three to twelve hours was not significant (p-value = 0.48).

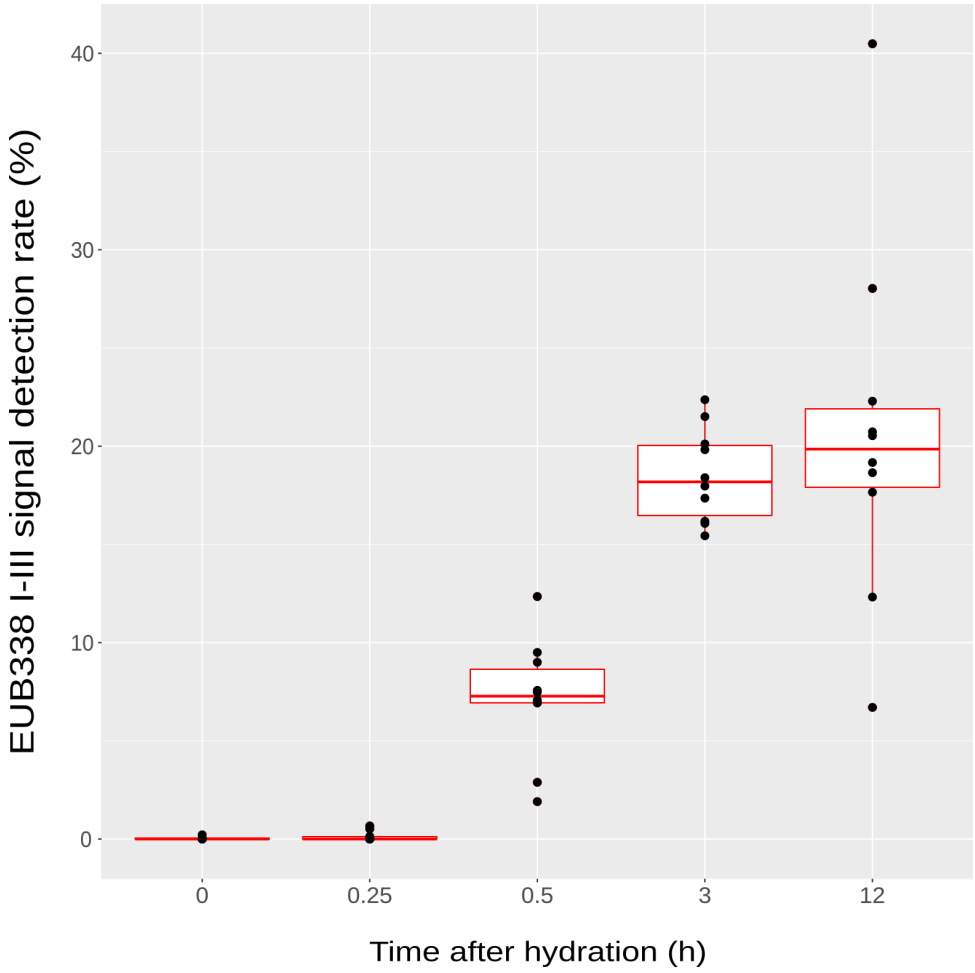

Figure 3.1: Proportion of cells detected with the EUB I-III probe mix fixed at different times after hydration. Ten images (black dots) were taken and more than 2,500 cells were counted per time point.

3.1.2 Maximum fluorescent intensity of cells was found 30 minutes after hydration

In mono-labelled FISH, signal intensity correlates with the number of ribosomes in a single cell since each probe binds to one ribosome [52, 53]. We calculated the intensity of FISH signals throughout the incubation time using the software *daime* (Section 2.3.5). Fluorescence intensity increased with incubation time and reached its maximum 30 minutes after hydration (53±12 Arbitrary Units). In time points 3 h and 12 h, FISH intensity remained stable (Figure 3.2).

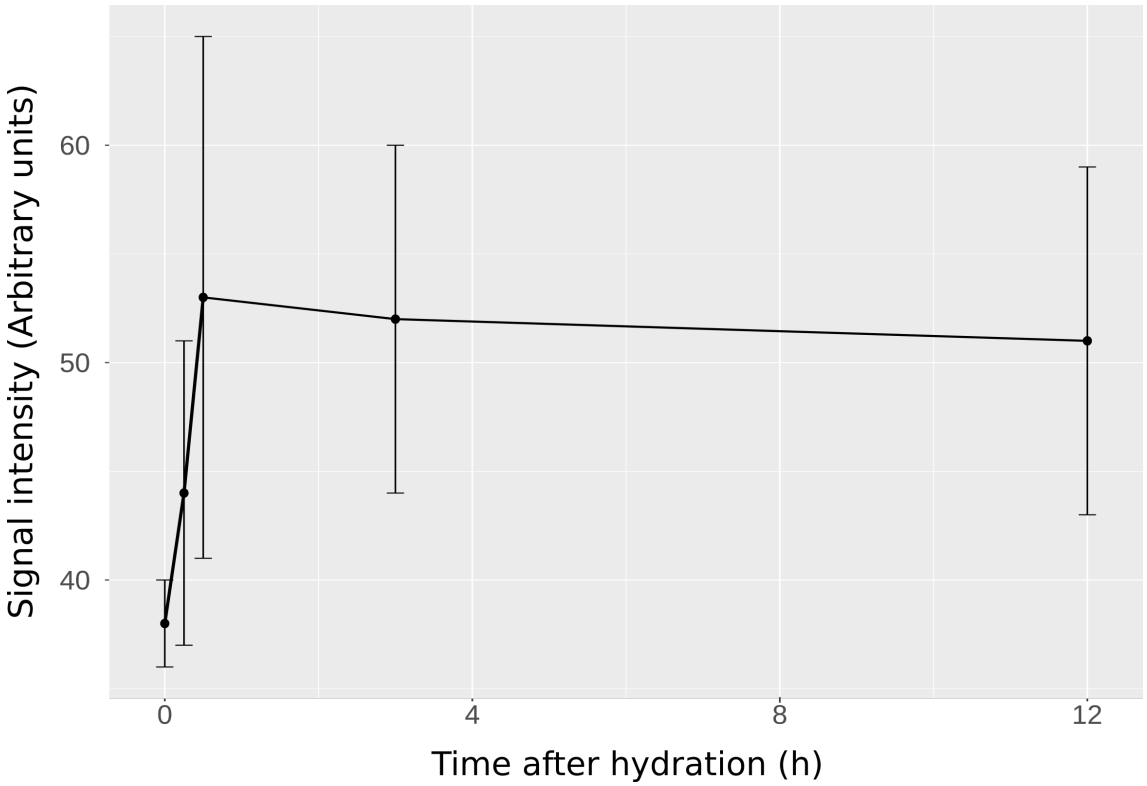

Figure 3.2: Mean intensity of FISH signal at different incubation times after hydration. At least 1000 objects were detected and measured in each time point except 30 min, in which 496 objects were detected.

3.2 Discussion

3.2.1 Hydration of the BSC lead to an increase in protein synthesis potential of the microbial community

Hydration experiment of a biological soil crust showed an increase of the FISH detection rate with the incubation time, and stabilisation after 3 hours incubation. The proportion of cells that were hybridised with the FISH probe EUB338 I-III in the community did not vary significantly between the 3 hours and 12 hours incubation time points, comprising a maximum of about 20% of the bacteria that were detected with DAPI (Figure 3.1). After 12 hours incubation, the majority of the cells (80%) had no FISH signal. This could be due to different reasons. First, that the number of ribosomes were below the FISH detection limit. It has been calculated that 1400±170 rRNA copies per cell are needed to display FISH signal in *E. coli* from activated sludge hybridised to the EUB338-I probe [55]. However, an alternative explanation could be that the probe could not enter a fraction of the cells due to rigid cell walls, regardless the ribosomal content. Adaptations to desiccation stress include modifications in the bacteria cell wall [29].

The mean intensity of the FISH signals increased with the incubation time too, reaching its maximum 30 minutes after hydration (Figure 3.2). Nevertheless, the size of the error bars in all the time points except for time point 0 hours indicates that there is a lot of heterogeneity in the fluorescence signal, and therefore the ribosomal load, of the BSC bacteria that could be hybridised.

Ribosomal content and growth rate are not always correlated [57]. Furthermore, bacterial response to starvation in terms of ribosomal content is variable between cells [107, 108], and ribosomal content can even be higher in dormant cells than in vegetative ones [109]. However, we saw that the fraction of cells in the community displaying a FISH signal and the mean signal intensity increased with incubation time. Our results suggest that protein synthesis potential increases in the BSC microbial community upon the hydration event.

3.2.2 Post-sampling storage time had no influence in ribosomal content of the BSC

The crust pieces used in this experiment belonged to a crust that had been sampled in 2019 and was stored for two years under stable dry and dark conditions (Section 2.1). Pieces from the same biological crust were previously used in reactivation experiments by Lisa Boden [58], in which the FISH patterns of a two years old crust (sampled in 2017 and hydrated in 2019) and a crust sampled and hydrated in 2019 were compared [58] (Figure 3.3). FISH detection rate increased rapidly in the freshly sampled 2019 crust and reached its maximum after 3 h hydration, whereas in the 2017 crust detection rate increased gradually throughout the 12 h incubation period (Figure 3.3a). In 2019 samples, the intensity of the FISH signals did not differ between time points, whereas

in 2017 it increased with the incubation time (Figure 3.3b). From these results it was hypothesised that rapid reactivation of the fresh 2019 crust was due to hibernating ribosomes reactivation, whereas ribosomes had to be synthesised *de novo* after two years storage, leading to a more gradual reactivation of the 2017 crust [58].

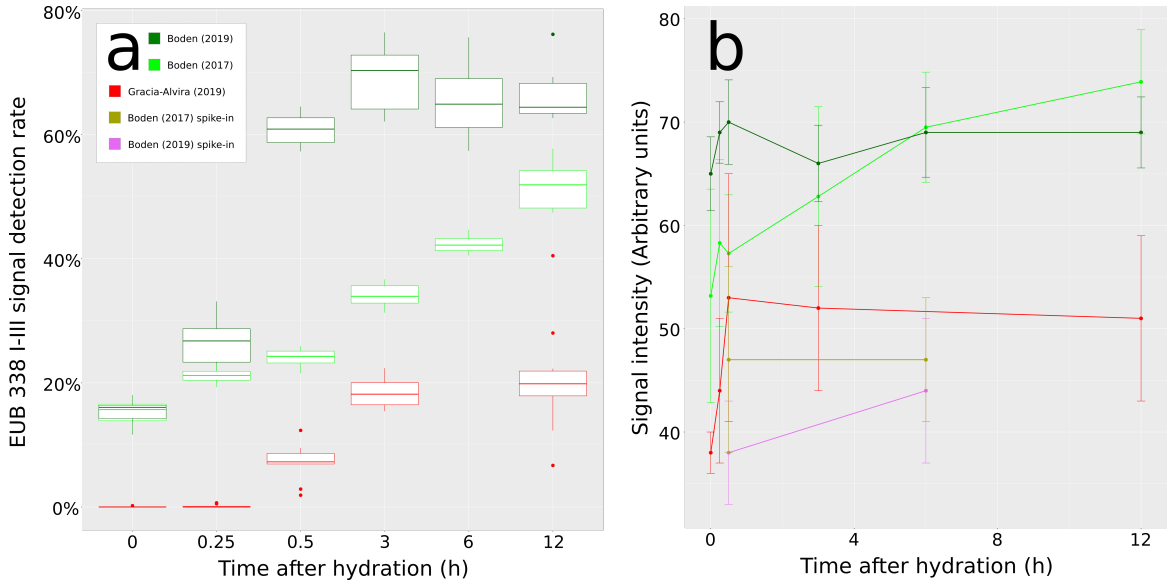

Figure 3.3: a: Proportion of cells detected with EUB I-III probe mix at different incubation times. b: Mean intensity of FISH signal at different incubation times. Results shown in figures 3.1 and 3.2 (in red) were compared to Boden 2020 results of BSC sampled in 2017 (in light green) and in 2019 (dark green) [58] and to spike-ins (i.e. the same filters used in Boden 2020 were hybridised *de novo*) of 2017 (yellow) and 2019 (pink) filters. In Boden 2020 an additional 6h time point was used.

Our results on the 2019 biological soil crust after two years storage show the same reactivation trend as the 2019 crust hybridised shortly after sampling: a rapid increase in the FISH detection rate and the signal intensity followed by stabilisation (Figure 3.3). However, both the proportion of hybridised cells in the community and the intensity of the FISH signals were much lower. In order to test if this discrepancy was due to lower ribosomal amounts after two years storage or due to variations in FISH protocol, the same filters used in Boden 2020 (time points 30 min and 6 h, fixed in 2019) were hybridised *de novo*, resulting in at least 10 intensity units below what it was observed in the same filters in Boden 2020 (Figure 3.3). Thus, the decrease in the detection rate and signal intensity are not due to ribosomal degradation after 2 years storage but either due to the fact that the FISH hybridisation of the same samples was carried out by two different people, which could lead to different FISH hybridisation efficiencies.

Alternatively, different reactivation trends of 2017 and 2019 BSCs could be explained by different starting community composition in these two years. It has been reported that microbial community composition is altered during wetting events and subsequently recovers to pre-wetting levels as the soil desiccates [110]. ANOSIM test of amplicon sequencing data from BSC collected in 2017 and 2019 resulted in small but significant dissimilarity between the microbial communities (R = 0.36, p-value =

0.0001). SIMPER test indicated that 207 out of 1362 OTUs contribute significantly to the differences between 2017 and 2019 crusts (personal communication, Dr. Dimitri Meier - ETH Zürich). In conclusion, ribosomal content of BSC cells increased throughout a 12 h period upon hydration. However, we did not find support for the hypothesis that post-sampling storage time had an influence in ribosomal content of the BSC, as the detection rate and signal intensity trends throughout a 12 h incubation did not change after 2 years storage of the 2019 BSC.

3.2.3 Perspectives: using conventional FISH for quantifying ribosomal content in cells

Comparison of FISH signal intensities of the same BSC showed a consistent trend over 12 hours, but there was disparity in the intensity values (Figure 3.3) despite following the same protocol and setting the same exposure time as in Boden 2020. This could be due to other sources of variation that could not be controlled, including the human factor, the state of the probe stock solution, the age of the microscope lamp or the state of the fluorescence filter cubes, that could reduce the intensity of the FISH signal between assays. From this we can conclude that the method is semi-quantitative; absolute intensity values should not be compared between assays but only the trends throughout the incubation time.

Autofluorescent particles were seen in microscopy images and detected automatically by the software *daime* as signals, leading to high standard deviation in the measurements caused by artifacts. Even though it was possible to remove manually signals corresponding to probe clumps (visualised as high intensity particles), other artifacts such as debris particles, that are more difficult to detect and remove due to their low intensity and small size, could be interfering with the analysis. I would suggest using double-labelled (DOPE) instead of mono-labelled probes to enhance the signal to noise ratio. DOPE probes are proportional to the ribosome content as well and result in more intense FISH signals with lower exposure time, reducing the autofluorescence and the noise in the intensity measurement.

FISH is frequently used to identify microorganisms present in environmental samples [57]. Here we used this technique in a quantitative way, to evaluate the changes in the ribosomal amounts of cells from a BSC throughout a 12 hours period upon hydration. The quantification of rRNA molecules is not a robust indicator of biological activity since some microorganisms accumulate ribosomes as they enter dormant state [109]. Furthermore, FISH probes do not seem to distinguish between hibernating and active ribosomes based on previous studies in dormant communities [111, 112]. However, the ribosome levels of the cells can inform us about the protein synthesis capacity of the community and how this changes with the time. By applying FISH in a quantitative way to BSC cells we concluded that the protein synthesis potential increased during the 12 h incubation and we did not find support to the hypothesis that long time storage of the BSCs leads to ribosome degradation that alters the detection rate and signal intensity patterns.

Chapter 4

Characterisation of Oman hyper-saline mats' microbial community

The microbial communities in hyper-saline mats are highly stratified, with phylogenetically and metabolically different taxa in each layer due to strong gradients in redox and light intensity across the mat depth [56]. Primary production in the mat is controlled by light and temperature, and the flow of carbon and nutrients follows internal cycles [113]. In microbial mats from a salt flat at the coast of Oman, five layers were distinguished and separated in the mats based on colour and texture [69]. We used catalysed reporter deposition fluorescence *in situ* hybridisation (CARD-FISH) to compare the spatial distribution of Archaea, all Bacteria, and Alphaproteobacteria specifically in two microbial mats with different salinity levels: 18% and 30%.

4.1 Results

4.1.1 Cell density does not vary with salinity

We aimed to calculate the cell abundance in each layer of the mats (Figure 2.1). Cell density and relative abundance of bacteria and archaea (see section 4.1.3) was calculated simultaneously in the same filter pieces. Material from each layer was hybridised and amplified with CARD-FISH probes EUB338 and ARC915 and subsequently stained with DAPI. Ten images were taken from each layer and counted. Unspecific binding of DAPI was seen in all the filters belonging to the high salinity mat upon CARD-FISH amplification. This incompatibility of DAPI staining with CARD-FISH amplification did not occur in the low salinity mat, in which we could clearly quantify the cells. Cell density of the high salinity mat had been calculated by Dimitri Meier in filters which were not previously amplified with CARD-FISH [69], so the cell density at both salinities could be compared.

Cell density per gram of material in the low salinity mat almost doubled from the top layer (layer 1) ($1.25 \cdot 10^{10}$ cells/g) to layer 3 ($2.29 \cdot 10^{10}$ cells/g). In layers 4 and 5 the number of cells decreased 10-fold, to $1.93 \cdot 10^{9}$ cells/g and $1.45 \cdot 10^{9}$ cells/g respectively (Figure 4.1). Cell density in the high salinity mat remained stable in the three upper layers and dropped in the last two, showing the same trend than low salinity mat (Figure 4.1). Differences in cell abundance were significant between the layers (p-value < 0.0005), but not between mats (p-value $= 0.607$) according to Kruskal-Wallist test. A subsequent pairwise Wilcoxon rank-sum test was performed to evaluate the differences in abundance between layers. In the low salinity mat differences in abundance were significant between all the layers (p-value < 0.015). In the high salinity mat, abundance in layers 1-3 (p-value > 0.15) and 4-5 (p-value $= 0.18$) was not significantly different.

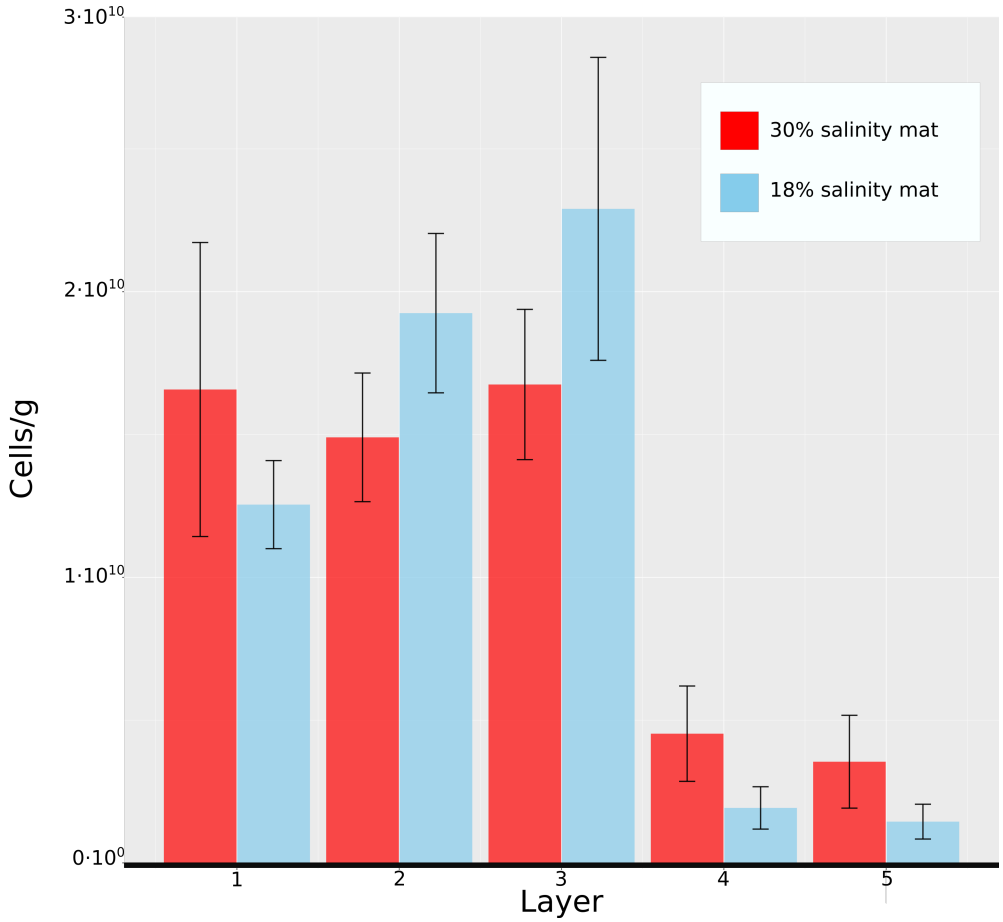

Figure 4.1: Barplot depicting cell abundance per gram of mat material in each layer of the high salinity mat (red) and the low salinity mat (blue). Cell numbers were obtained by counting DAPI-stained cells by epifluorescence microscopy. Data of the high salinity mat were taken from Meier *et al.* 2021 [69]. Please see Figure 2.1 for information on layer numbers.

4.1.2 Visualisation of Archaea and Bacteria in hypersaline mats

Relative abundances of Archaea and Bacteria in the microbial mats had been previously estimated by Dr. Dimitri Meier by mapping metagenomic reads to the SILVA SSU132 database using PhyloFlash [114]. PhyloFlash estimated high Archaeal abundance in the upper layers, reaching 80% of the total microbial reads percentage (Figure 4.3). Sequencing-based diversity estimation methods carry a potential source of bias introduced during DNA extraction. Furthermore, they provide indirect abundance information since they detect DNA, not whole cells. Thus, relative abundance could be inaccurate due to poliploidy and variable 16S rRNA gene copy number [65, 64]. I used CARD-FISH microscopy to verify the abundance estimation using a single-cell approach. Filters from each layer of a low salinity mat were hybridised with probes ARC915, specific for Archaea and EUB338 I-III, specific for Bacteria (Table 2.1). NON338 probe, which does not hybridise to any 16S rRNA sequence, was used as a negative control.

CARD-FISH was chosen over conventional FISH because mat material presented high autofluorescence, especially in 480 nm (green) and 550 nm (orange) channels (Supplementary figure S2). Lower autofluorescence levels were seen in the 660 nm (far red) channel (Supplementary figure S2). For this reason, Alexa Fluor 594 Tyramide was used in CARD-FISH amplification. Bacteria inhabiting the mats had heterogeneous morphologies in the three upper layers, being the most abundant: bundles of filamentous bacteria, cyanobacteria (identified as thick filaments), cocci, bacilli and diplobacilli. Bacteria from the two bottom layers were mainly bacilli and filamentous bacteria. Regarding Archaea, they had mainly coccoid shape and were arranged in diplococci and tetrads in all the layers (Figure 4.2). The two bottom layers, in which cell density decreased drastically, were abundant in sediment particles and dead cells.

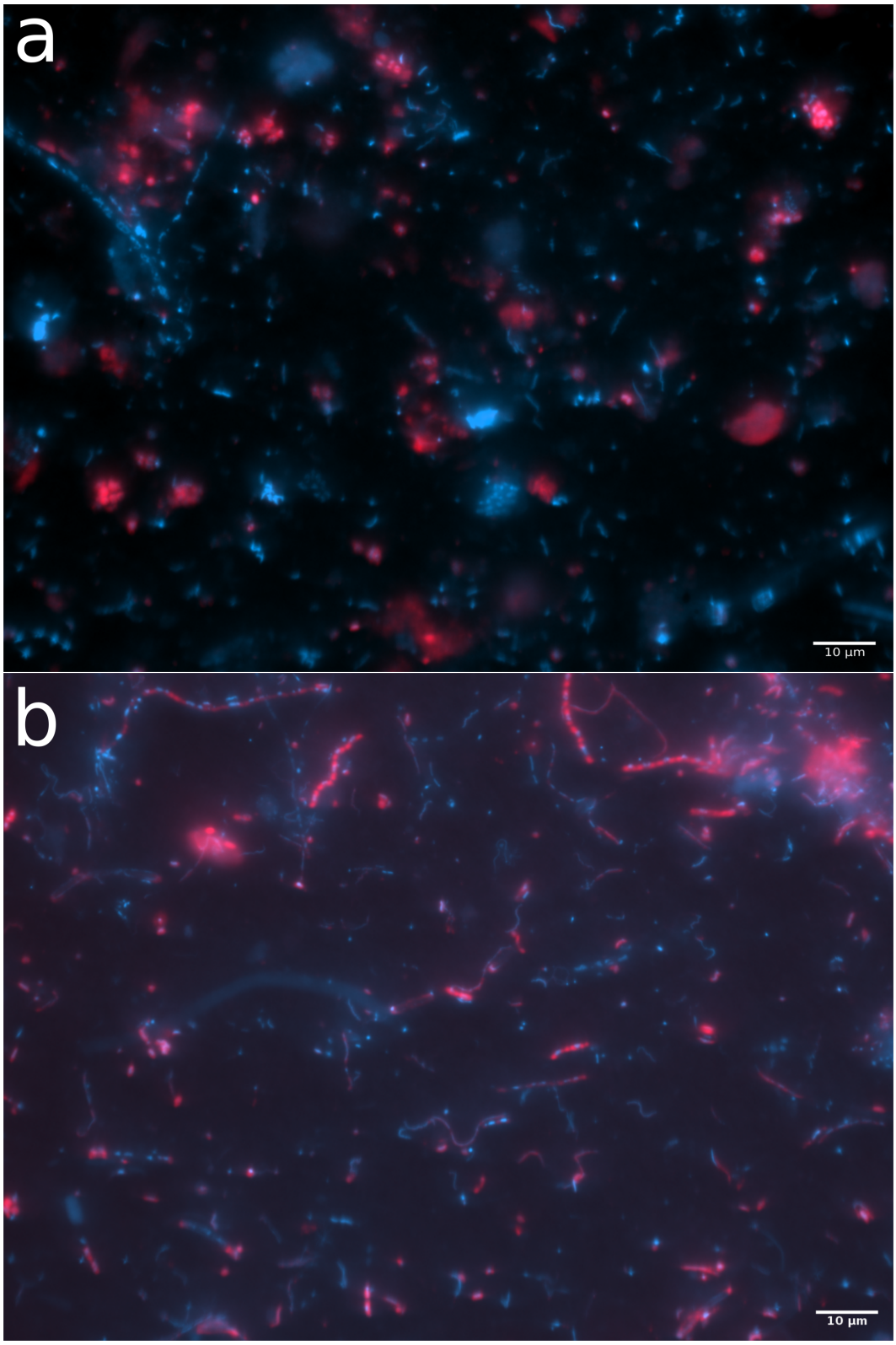

Figure 4.2: Epifluorescence microscopy images of hyper-saline mat material from layer 1 hybridised with ARC915 (a) and from layer 3 hybridised with EUB338 I-III (b). CARD-FISH probe (red) and DAPI (blue) channels are overlapped in the same image. Scale bar of 10 μm.

4.1.3 Relative abundance diverged between quantification methods

In order to verify the abundance estimations based on metagenomic reads (obtained from three replicate mat cores, each separated in five layers) throughout the depth profile of the low salinity mats, CARD-FISH was performed on the five layers of one mat core. CARD-FISH and DAPI signals were counted in ten images for each layer and probe, and the CARD-FISH/DAPI ratio was calculated. The trend of Archaea relative abundance was the same as seen in PhyloFlash estimation; an "U" pattern with highest abundance in the top and bottom layers of the mat. However, the maximum abundance of Archaea identified by CARD-FISH in layer 1 was 20% in comparison to an average 60% abundance estimated based on the metagenomic reads (Figure 4.3a). The layers with highest relative abundance of Bacteria differed between analysing metagenomic reads and CARD-FISH. Based on metagenomics, highest bacterial abundance was found in layers 3 and 4 with 92% and 91%, whereas with CARD-FISH the highest abundance was detected in layer 2, reaching 82%.

The sum of Archaea and Bacteria relative abundance detected with CARD-FISH was close to 100% in layers one and two. In layer three, detection rate fell to 70%, and reached the lowest values in layers four and five with 40.19% and 55.96% respectively (Figure 4.3b). Filters from layers four and five were hard to quantify due to high abundance of disintegrating cells, in which the cell borders are diffused due to loss of membrane integrity, and autofluorescent particles. However, lower detection was also seen in layer three, which had the highest total cell abundance (Figure 4.1) and in which most cells kept their integrity.

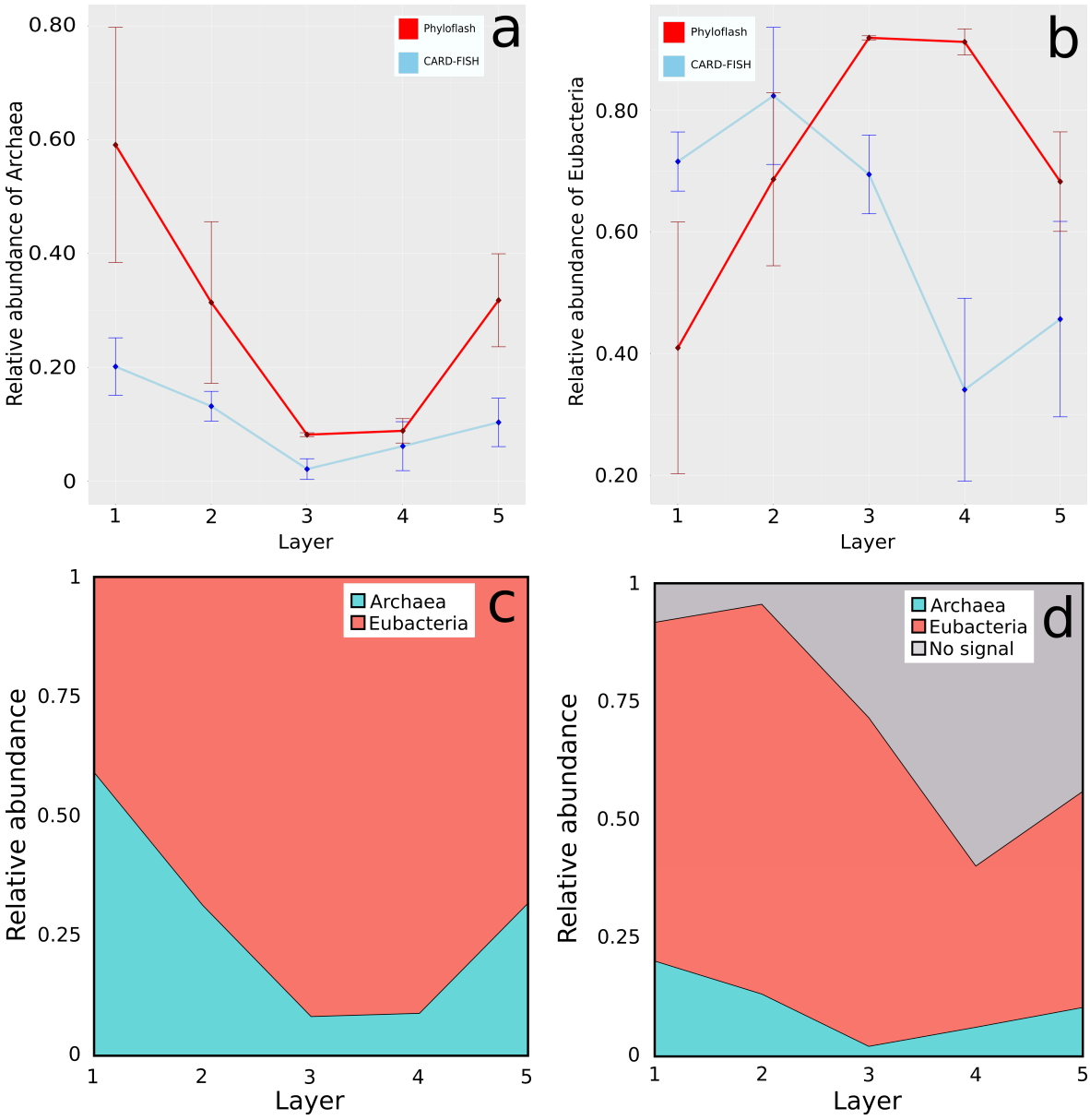

Figure 4.3: Relative abundance of Archaea and Bacteria across the mat layers according to metagenomic reads abundance (PhyloFlash) and fluorescence microscopy.

Top panel: mean relative abundance of Archaea (a) and Bacteria (b) according to Phyloflash (red) and CARD-FISH (blue). Metagenomes were produced from three low salinity mat cores (Section 2.1) and the Phyloflash relative abundance was calculated as the average of the three replicates in each layer. CARD-FISH was only done in one mat core, and ten images and at least 500 cells were counted for each layer (Section 2.3.5). (obtained from three replicate mat cores, each separated in five layers).

Bottom panel: stacked relative abundance of Archaea (blue) and Bacteria (red) in each layer of the low salinity mat. Quantification with Phyloflash (c) and CARD-FISH (d). "No signal" (grey) refers to the proportion of the cells detected with DAPI that were not covered by either EUB338 or ARC915 probes.

4.1.4 Alphaproteobacteria, Rhodothermaceae and Chloroflexi change in abundance with salinity

PhyloFlash abundance estimation highlighted differential presence of three taxa between the metagenomes from low salinity and high salinity mats and across the depth profile: Alphaproteobacteria, Bacteroidetes and Chloroflexi.

Alphaproteobacteria populations were biologically interesting due to their potential role in the microbial mat and their distribution across the depth profile. The predominant Alphaproteobacteria taxa estimated by PhyloFlash were the genus *Rhodovibrio* and the family Rhodobacteraceae, being respectively the most abundant Alphaproteobacteria in the metagenomes from the high salinity and the low salinity mats (Supplementary figure S1). Both clades could occupy the same niche as phototrophic sulphide-oxidisers, since the MAGs binned from the metagenome harbour the *sox* genes. These genes encode for the Sox enzyme system (SoxABCDXYZ), which confers the ability to oxidise sulphur compounds like hydrogen sulphide or thiosulphate [115].

The familiy Rhodothermaceae, including the genus *Salinibacter*, were the most abundant Bacteroidetes, especially in the metagenomes of the top two layers of the high salinity mat. *Salinibacter* was more abundant in high salinity than in low salinity mats (Supplementary figure S1). A clade belonging to the family Marinilabiliaceae increased in abundance with depth in both salinities.

Two Chloroflexi taxa were predominant in the metagenomes at different layers. *Candidatus* Chlorothrix was the most abundant Chloroflexi taxon in metagenomes of the three upper layers of both mats, whereas SBR1031 was dominant in the bottom ones (Supplementary figure S1).

4.1.5 Only Alphaproteobacteria could be quantified in microbial mats

I used CARD-FISH to verify the relative abundance of the aforementioned taxa of interest in the community. I used probes specific to Rhodothermaceae, Chloroflexi and Alphaproteobacteria, as well as the NON probe as negative control (Table 2.1). Prior to performing CARD-FISH, I tested *in silico* if the probes matched any of the microbial taxa from the mats. PhyloFlash was used to assemble 16S rRNA sequences from the metagenomic raw reads. From the 319 sequences that were reconstructed, 281 were larger than 1000 bp. The three probes matched 16S rRNA sequences with no mismatches. ALF968 probe matched nineteen 16S rRNA sequences. Although most of ALF968 matches did not belong to Alphaproteobacteria (i.e. Desulfobacterota, Myxococcota, Planctomycetota, Zixibacteria, FW113 and "Unclassified taxa"), the summed outgroup hits supposed less than 1% of the metagenomic reads relative abundance in all the layer. CFX1223 matched twenty three Chloroflexi sequences, covering the taxa *Candidatus* Chlorothrix and SBR1031. EHB412 matched six sequences belonging to the order Rhodothermales, including the genus *Salinibacter*.

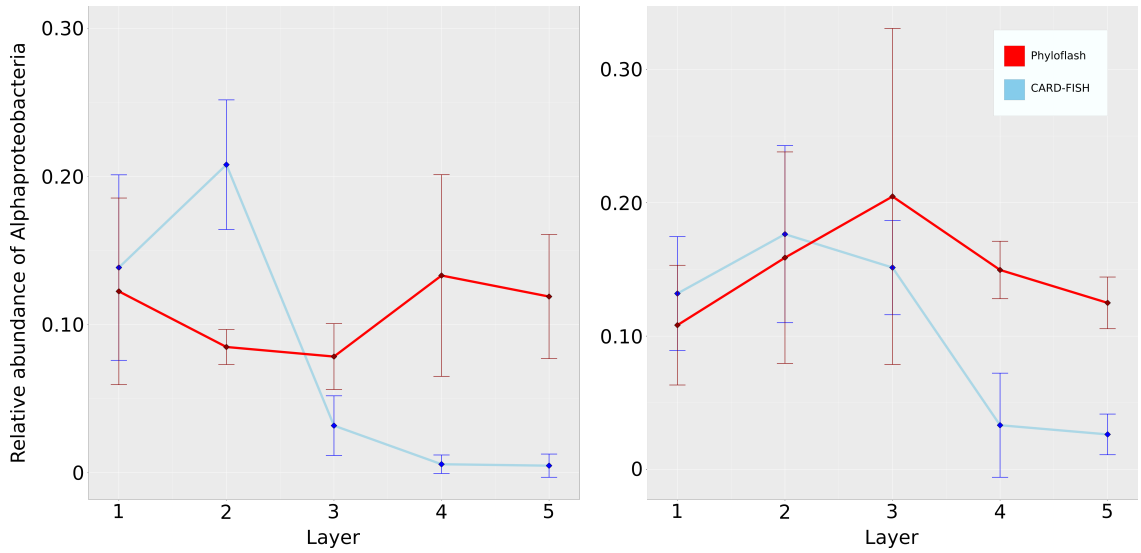

Figure 4.4: Relative abundance of Alphaproteobacteria in the low salinity (left) and high salinity (right) mats according to CARD-FISH (blue) and PhyloFlash (red).

CARD-FISH signal could not be detected upon filter hybridisation with CFX1223 or EHB412 probes. Thus, only Alphaproteobacteria abundance could be quantified at both salinities and compared to metagenomic estimations. The Alphaproteobacteria distribution pattern obtained by CARD-FISH diverged from the one estimated from the metagenomic reads. According to CARD-FISH, Alphaproteobacteria exhibited similar dynamics throughout mat depths in both salinities (Figure 4.4). The maximum relative abundance was found in the top two layers of both mats and decreased significantly in the two bottom layers. However, the middle layer had a different abundance in each mat. In the high salinity mat, relative abundance in layer three was 15%, resembling the top mats' abundance (13% and 17%), whereas in the low salinity mat it was 3%, following the decreasing trend and leading to 0.5% and 0.4% in layers four and five.

4.2 Discussion

4.2.1 Archaea abundance was overestimated in the metagenomic reads

In Meier *et al.* 2021, 16S rRNA amplicon sequencing of different mat layers resulted in Archaea relative abundance of 56% in the mat layer with the highest relative abundance (layer 4) [69]. This matches the observed relative abundance based on the proportion of metagenomic reads mapped to archaeal 16S rRNA sequences, which amounts to on average 60%, albeit in a different mat layer (layer 1) (Figure 4.3a). This is in contrast to the relative abundance of Archaea of 21% based on CARD-FISH quantification. However, CARD-FISH quantification was in concordance with previous reports in microbial mats, in which Archaea make up between 1% and 20% of the community in hyper-saline microbial mats [116, 117].

Poor correspondence between 16S rRNA amplicon sequencing and CARD-FISH quantification has been reported in the literature [63, 118, 119]. In Ushio *et al.* 2014 the same soil samples were quantified with amplicon sequencing and CARD-FISH. They reported an Archaea abundance of 1% with amplicon sequencing and more than 30% with CARD-FISH. Piwosz and colleagues compared the accuracy of amplicon sequencing and CARD-FISH in mock and environmental communities and concluded that CARD-FISH was more accurate estimating the true taxa abundance, but both methods were reliable when studying the influence of environmental parameters in the community composition [118, 119]. Abundance estimation based on the mapping of metagenomic reads avoids part of the biases and artifacts associated with amplicon sequencing (e.g. primer bias and sequence chimerism [120]) [114], but bias is still introduced during DNA extraction step. Furthermore, poliploidy, which is common in bacteria and archaea [65], and variable number of 16S rRNA gene copies per genome [64] can lead to the underrepresentation of taxa with fewer genome and 16S rRNA gene copies in the metagenomic reads pool. Here we report differences in abundance between PCR-independent metagenomic sequencing and CARD-FISH. A previous study has shown similar proportion of Bacteroidetes in CARD-FISH and 16S rDNA sequences recovered from marine water metagenomes [121]. However, studies using mock communities are needed to assess the true accuracy of metagenomic profiling, amplicon sequencing and FISH and CARD-FISH quantification.

4.2.2 Lack of CARD-FISH coverage could be due to low ribosome content or cell wall impermeability

In the upper two layers, CARD-FISH had a detection rate of nearly 100% (91.64% and 95.46%, respectively, summing up relative abundances of Bacteria and Archaea in layer 1 and 2). Whereas the three lower layers had a summed Archaea and Bacteria relative abundance of 70%, 40.19% and 55.96%, respectively (Figure 4.3b). Similar relative abundance patterns of Archaea were determined with both, DNA sequencing based and CARD-FISH based method. In contrast, the relative abundance pattern of

Bacteria differed between both methods. Therefore, it is likely that the missing CARD-FISH signals in the lower layers correspond to Bacteria. Decrease in the detection rate in the lower layers of the mat could be explained by a higher relative abundance of cells that are either dead, or that maintain their cell wall integrity and DNA, but have ribosome numbers below the detection limit of CARD-FISH. The minimum number of rRNA molecules required to display a fluorescent signal in CARD-FISH is less than a hundred per cell in comparison with more than a thousand needed in conventional FISH [55]. However, we cannot discard ribosome degradation below CARD-FISH detection limit in a fraction of the community, especially in layers 4 and 5. In these layers cell abundance decreased 10-fold (Figure 4.1), pointing to decreased viability. Besides, less viable cells could be eventually buried in the bottom of the mat due to sedimentation. In Pernthaler *et al.* 2002, significant decrease in the detection rate of the community throughout the depth profile of sea water sediment was also reported [54]. A mean detection rate of 93% was reported in the depth 0-6 cm, that fell to 60% in the depth 6-14 cm. However, since only EUB338-HRP was used, this decrease could be partially explained by the presence of Archaea in the microbial community of the lower layers. Additionally, higher abundance of dead cells was hypothesised as a reason for this observation [54].

The probes EHB412 and CFX1223, that target the abundant taxa Rhodothermales and Chloroflexi (according to the metagenomes), had no signal in any of the layers of both mats. The main reasons for failure in FISH are low accessibility of the probe to the rRNA, low ribosome content and cell wall impermeability [122, 55, 123]. FISH probes design is based on the matching specificity of the sequence, but must fulfill other prerequisites such as optimal length, hybridisation behaviour and ribosomes accessibility [124]. According to *E. coli* 16S rRNA accessibility map [122], EHB412 binds to a class II region whereas CFX1223 binds to class II and IV regions. The 3D structure of the 16S rRNA molecule could vary in other organisms, especially in extremophiles, but the complete lack of signal would require the binding of both probes to class VI (lowest accessibility) regions. Furthermore, both probes have been successfully used for FISH in the case of EHB412 [76] and CARD-FISH in the case of CFX1223 [47]. Therefore, it is unlikely that unsuccessful hybridisation with probes EHB412 and CFX1223 is due to low ribosome accessibility.

Cell wall impermeability could explain the lack of signals corresponding to the abundant taxa Rhodothermaceae and Chloroflexi. Despite being a more sensible method, CARD-FISH is less efficient than FISH penetrating the cell walls because HRP-conjugated probes have a higher molecular weight than conventional FISH fluorescent probes [123]. It has been reported that Gram-positive cell walls are harder to permeabilise than Gram-negative [125]. In the same way, modifications in the bacterial cell wall and cell membrane composition, which are a well known mechanism to cope with salt stress and desiccation [3, 126, 27], could limit the permeability. Therefore, CARD-FISH efficiency could be limited only in specific taxa from the community, i.e., those that use cell wall modification strategies to survive in hyper-saline conditions.

Rhodothermales and Chloroflexi account for 32% of the total metagenomic reads in layer 2. In this layer, 95.46% of the cells had a CARD-FISH signal upon hybridisation with EUB338 and ARC915 probes. Therefore, following the hypothesis that the hybridisations with probes EHB412 and CFX1223 were unsuccessful due to impermeable

cell walls of the target organisms, then the abundance of these taxa would need to be less than 5% provided that they are the only taxa that cannot be targeted using CARD-FISH, which is not realistic. Therefore, it remains open to us why these probes did not work in our sample material. This issue points out the need for optimised CARD-FISH protocols that allow to cover all the diversity in complex samples.

4.2.3 Taxonomic profiling is needed to verify the absence of false positives in CARD-FISH samples

The probe ALF968 matched nineteen 16S rRNA assembled sequences, from which only six were Alphaproteobacteria. The thirteen matched outgroup sequences included the phyla Desulfobacterota, Myxococcota, Planctomycetota, Zixibacteria, as well as unclassified sequences. Nevertheless, these outgroups supposed less than 1% of the hits in all the layers. Phyloflash performs taxonomic profiling by mapping the metagenomic reads against the SILVA SSU database and calculating taxa relative abundance as the percentage of 16S rRNA reads assigned to each taxon over the total number of reads assigned to the 16S rRNA gene [114]. Additionally, full-length 16S rRNA sequences are assembled from the reads [114]. Targeted assembly of the 16S rRNA gene from metagenomic reads is challenging due to the presence of highly conserved regions in the sequence, that lead to loops in the assembly graph [114]. For this reason, the list of 16S rRNA assembled sequences recovered from the metagenome does not provide a reliable overview of the community composition due to under-detection of certain taxa. Assembled 16S rRNA sequences can be used to design probes and study the phylogeny, but do not provide relative abundance information.

Additionally, we used the software TestProbe from the SILVA database to evaluate the outgroups that are targeted with no mismatches by the widely used FISH probe ALF968. Only 74% (23,962/32,242) of the hits against the database belonged to Alphaproteobacteria. Furthermore, 22% of the Desulfobacterota sequences, 58% of the Mycococca and 22% of the Zixibacteria sequences collected in the SILVA database matched the ALF968 probe. Previous evaluations of the probe ALF968 revealed many outgroup hits (false positives) throughout the bacterial domain and 20% of untargeted Alphaproteobacteria (false negatives) in the SILVA database [41]. This addresses the importance of performing taxonomic profiling of the community before carrying out CARD-FISH, despite the potential bias discussed in Section 4.2.1. We could use the ALF968 probe since mapping-based taxonomic profiling resulted in a minimal relative abundance of outgroup taxa in the hyper-saline mat metagenomes even if the majority of the hits in the 16S rRNA assembled sequences were not Alphaproteobacteria. Even though the two main Alphaproteobacteria clades, *Rhodovibrio* and Rhodobacteraceae, are targeted by the ALF968 probe, we should be aware that we could be missing a fraction of the Alphaproteobacteria present in the community.

4.2.4 Alphaproteobacteria distribution could be influenced by sulphur and oxygen availability

Abundance of Alphaproteobacteria throughout the mat depth differed between metagenomic reads estimation and CARD-FISH microscopy quantification. However, with both methods Alphaproteobacteria in the layer 3 was higher in the high salinity mat than in the low salinity mat (Figure 4.4).

Rhodobacteraceae, the most prominent Alphaproteobacteria clade under low salinity conditions, are known to be versatile photoheterotrophic organisms that can adopt different trophic strategies and therefore occupy various niches in the mats [34]. Although Rhodobacteraceae are typically classified as purple non-sulphur bacteria, the MAGs binned from Oman mats metagenome harbour the *sox* genes, so they could use reduced sulphur compounds like hydrogen sulphide or thiosulphate. The other major Alphaproteobacteria clade, *Rhodovibrio*, harbours the *sox* genes too.

Changing salinity alters the biochemical fluxes in the mat [69]. Under high salinity conditions, oxygenic photosynthesis is inhibited leading to an increase in sulphide concentration since O_2 cannot be used as electron acceptor. Differences in O_2 and sulphide concentrations could define the majoritarian Alphaproteobacteria clade (Rhodobacteraceae under low salinity and *Rhodovibrio* under high salinity S1) and its depth profile. Vertical migration of bacteria in response to changing sulphur and oxygen levels has been hypothesised previously [127, 128]. Both, sulphate-reducing bacteria and phototrophic anoxygenic bacteria were differentially found during the day in the mid layer of a hyper-saline microbial mat, that undergoes diel fluctuation between oxic and anoxic zone [127, 128].

4.2.5 Perspectives: combination of metagenomic and microscopy data in microbial diversity assessment

In order to characterise the microbial communities inhabiting two hyper-saline microbial mats, we used a combined approach in which high throughput metagenomic data was used to target taxa of interest that were estimated to change with the salinity and the depth profile, and low throughput CARD-FISH was used to quantify them. From our results and previous works [119, 129, 63], we can conclude that both methods present biases and challenges for the characterisation of halophilic microorganisms in microbial mats. Their physiologies or survival strategies can interfere with the molecular assays: high amounts of EPS and autofluorescent pigments can interfere with fluorescence microscopy evaluation, whereas cell wall modifications could affect CARD-FISH efficiency. Microscopy seems to be particularly informative in the deepest mat layers, as it shows that cells disintegrate, while their DNA could still be detected. Taking this into account, a combined quantification strategy seems the best way to obtain reliable results.

Chapter 5

Pangenomic analysis of Cyanobacteria and Alphaproteobacteria inhabiting arid and saline environments

High salinity and lack of water have as a common consequence: reduced bioavailable water [3]. Nevertheless, it has been reported that the cellular response to desiccation and hyper-osmolarity is different [8, 9]. I used a MAG-centric comparative genomics approach to study the differences in putative resistance genes and strategies found in bacteria from arid and saline habitats. I compared the genes encoded by closely related microorganisms inhabiting arid soils and aquatic saline environments ranging from brackish to hyper-saline conditions. Additionally, I used available freshwater genomes as a reference for organisms putatively unadapted to low water potential. For this analysis I mainly used data from two metagenomes that were produced by Meier *et al.* in the following habitats: hyper-saline microbial mats in Oman [69] and biological soil crusts in the Negev desert [39] (Section 2.4.1).

Microbial mats material was sampled from an intertidal area that reached salinity levels of up to 40% (saturation). According to preliminary 16S rRNA data, mat community was stratified by depth. The mat community was characterised by a high proportion of archaeal sequences, that could reach 56% of the relative abundance in some layers [69]. Cyanobacteria constituted a minor proportion of the reads, whereas anoxygenic phototrophs such as Chloroflexi were much more abundant. In total, 393 MAGs were assembled and binned from the sequenced mat material.

BSC metagenomes were produced from an arid site in the Negev Desert, Israel, with less than 100 mm of annual rainfall [39]. Based on both, 16S rRNA amplicon sequencing and metagenomic reads analysis, the community was composed of large proportions of Cyanobacteria and Actinobacteria and smaller fractions of Alphaproteobacteria, Bacteroidota, Chloroflexota, Gemmatimonadota and Acidobacteriota [39]. Ninety-six MAGs were binned from the metagenomes and were dominated by Actinobacteria (54/96 MAGs), whereas Cyanobacteria abundance was lower in the MAGs and in 16S rRNA amplicon sequencing data (4/96 MAGs and 29/922 OTUs) [39].

In order to study habitat-specific adaptations, I looked for the functions that were conserved exclusively in one of the environments (i.e. arid or saline). I compared the genomes from two taxa: *Coleofasciculus* (genus) and sulphide-oxidising phototrophic Alphaproteobacteria (class) were present in both aforementioned habitats. However, due to the low number of *Coleofasciculus* genomes from both habitats, I expanded the genome collection to the closely related cyanobacteria genus *Microcoleus* and I searched in public databases for genomes that were isolated from arid soils and non-freshwater aquatic samples, ranging from brackish to hypersaline.

5.1 Results

5.1.1 Cyanobacteria pangenome

Filamentous cyanobacteria of the genus *Coleofasciculus* were found in both, BSC (one MAG) and microbial mat (three MAGs), metagenomes. Additionally, I collected from public databases *Coleofasciculus* genomes obtained from arid soils, aquatic brackish, saline and hyper-saline environments and freshwater (Supplementary table S1). Due to the scarcity of *Coleofasciculus* genomes in the databases, I included in the analysis genomes from the genus *Microcoleus*, that was present in the Negev BSC metagenome. *Microcoleus* is also a filamentous Cyanobacteria genus, with a similar lifestyle (i.e. both are phototrophic, filamentous cyanobacteria that aggregate to solid matrices by secreting high amounts of EPS) and that has also been found in microbial mats and BSCs in xeric environments [130]. I collected 22 genomes; 10 classified as *Coleofasciculus* and 12 as *Microcoleus* (Supplementary table S1) according to the metadata.

Taxonomic classification and completion estimation

I checked the taxonomy of the retrieved genomes with GTDB-Tk, that classified and placed the 22 sequences into the tree of bacterial genomes based on concatenated alignment of 120 marker protein sequences [88]. The ten *Microcoleus* sequences were classified as *Microcoleus* in the GTDB bacterial tree. However, only four out of twelve "*Coleofasciculus*" genomes were actually placed in this genus. Five were placed in the genus PCC7113, that belongs to the family Coleofasciculaceae, two were placed in the Coleofasciculaceae family but were not assigned a genus and one was placed outside of it, in the Chamaesiphonaceae family (Table 5.1).

Metagenome assembly and binning are challenging processes; none of the available algorithms has achieved to reconstruct genomes from an environmental sample with 100% completeness [131]. We used CheckM to analyse the quality of the MAGs using lineage-specific marker genes for each MAG. Nine genomes had completeness values over 90%, eight genomes were between 70% and 90% complete and five genomes were between 50% and 70% complete (Table 5.1). The sum of contamination (presence of marker genes from distantly related organisms) and strain heterogeneity (presence of marker genes from closely related organisms) was below 5% in seventeen genomes and it did not reach 10% in any genome.

Genomes completeness and taxonomy				
Genome name	GTDB taxonomy	Completeness (%)	Contamination (%)	Strain heterogeneity (%)
bin38	s *M.* sp013299165	98.72	0.00	0.11
bin48	s *M.* sp013299185	99.27	0.55	0.00
C1bin4	g *Microcoleus*	69.82	1.43	1.43
C3bin4	g *PCC7113*	54.25	1.39	1.18
Co250	g *PCC7113*	87.83	4.46	0.71
Cobin12	g *Microcoleus*	52.22	2.51	3.95
Cobin14	g *PCC7113*	60.96	1.54	1.35
Coleofasciculus01	g *PCC7113*	89.74	1.05	0.17
FACHB1120	f Chamaesiphonaceae	100.00	0.37	0.00
LEGE07076	g *Microcoleus*	98.48	1.34	0.19
LEGE07081	g *Coleofasciculus*	99.17	1.70	0.00
LEGE07092	g *Coleofasciculus*	99.56	1.26	0.00
Microcoleus01	g *Microcoleus*	90.07	1.11	6.86
OSALTMAG134	f Coleofasciculaceae	97.22	3.28	0.35
OSALTMAG256	f Coleofasciculaceae	85.24	3.05	1.14
OSALTMAG82	g *Coleofasciculus*	76.67	0.18	0.89
PCC7420	s *C. chthonoplastes*	98.93	0.80	0.16
SCS111	g *PCC7113*	65.70	3.41	0.93
SCS153	g *Microcoleus*	76.55	1.93	4.38
SCS76	g *Microcoleus*	79.36	0.71	3.72
T1bin1	g *Microcoleus*	75.86	3.39	1.00
T3bin5	g *Microcoleus*	70.26	6.66	1.29

Table 5.1: Taxonomic placement and completeness and contamination estimation of the analysed Cyanobacteria MAGs. MAGs were placed at species (s), genus (g) or family (f) level in the bacterial tree of life.

Pangenomic analysis

Anvi'o pangenomic analysis associates annotated homologs in groups based on Diamond amino acid alignments [99]. Therefore, the functional working unit in this analysis is not the gene anymore, but the cluster of homologous genes. These genes are classified in the same cluster under the premise that they will have the same function. We compared the functional potential of the genomes based on presence-absence of homologous genes.

Anvi'o pangenomic workflow produced 24,762 gene clusters, that were reduced to 10,835 after removing the gene clusters that were found in only one genome (singletons). When I sorted the genomes based on their presence/absence pattern of homologous genes (gene clusters), they were split in two branches corresponding to *Microcoleus* and Coleofasciculaceae genomes. At lower ranks saline Coleofasciculaceae (OSAL MAG134, 256 and 82, LEGE07081 and 07092 and PCC7420) clustered together (Figure 5.1).

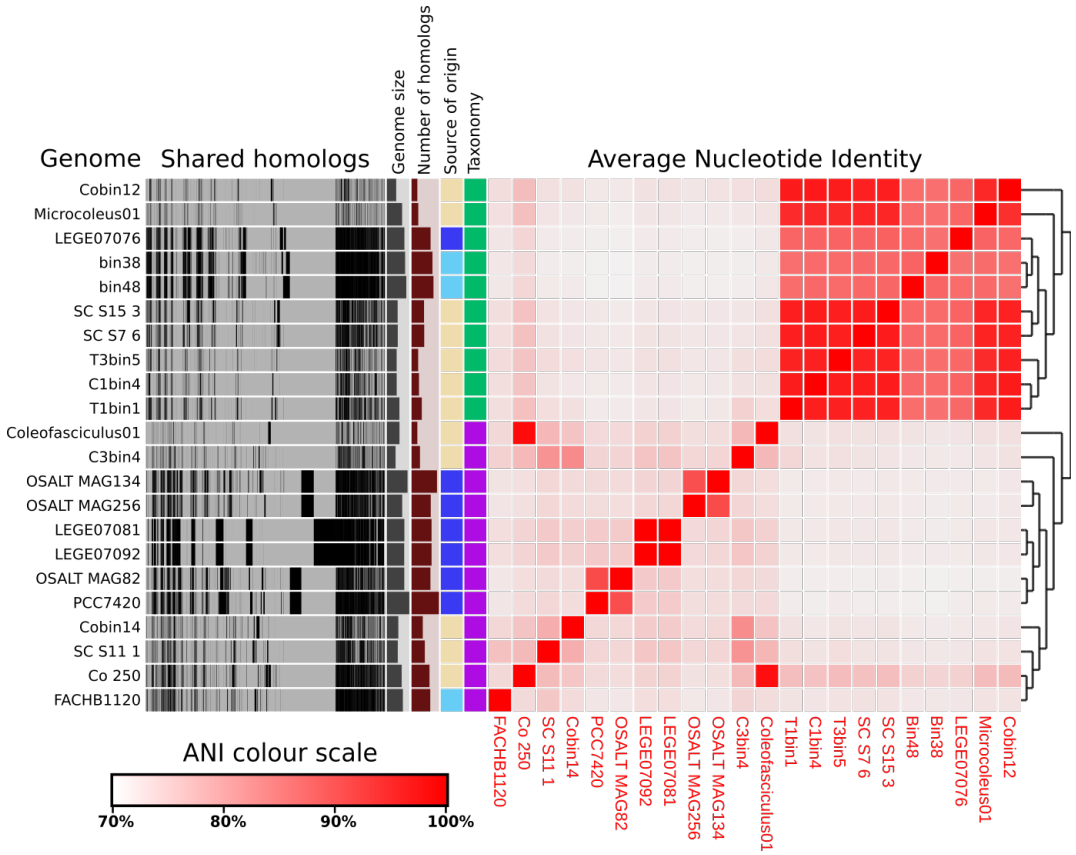

Figure 5.1: Anvi'o pangenomic analysis of 22 genomes belonging to *Microcoleus* (green) and Coleofasciculaceae (plus the Chamaesiphonaceae FACHB1120) (purple) binned from desert (brown), saline (dark blue) and freshwater (light blue) metagenomes. 75,947 genes were allocated in 10,835 gene clusters (black bars) after removing singletons. ANIb was calculated with pyANI v0.2.10 [103] and ranges from 70% (white) to 100% (red). Genomes were ordered according to presence/absence of gene clusters. Dendrogram was built with Fasttree v2.1.10 [132].

Anvi'o enrichment analysis

Additionally, I performed an enrichment analysis, that resulted in a list of gene functions that are predominantly found in one habitat (Section 2.9). This allows us to find functions that are characteristic of saline or arid environments.

One hundred and forty-four protein functions were reported to be more frequent in saline genomes than in desert and freshwater ones. Several putative halotolerance functions, selected among the top enriched functions based on literature research [1, 3, 7, 9, 133, 134, 135], were found in at least six out of seven saline genomes and one or less non-saline ones. These functions were: catalase KatE (COG0753), Na^+/proline symporter PutP (COG0591), Glycerol-3-phosphate dehydrogenase GlpA (COG0578), trehalose-6 phosphate synthase OtsA (COG0380) and Na^+/H^+ antiporter NhaP (COG0025) (Figure 5.3). Although these functions were present in most saline cyanobacteria, they were covered by different protein families (e.g. four gene clusters were associated to KatE, each one found in different genomes). Only three proteins were reported to be more frequent in desert genomes than in saline. And none of them had a high enrichment value.

Even though the genomes were classified by their functional potential (gene clusters presence/absence), they clustered based on phylogeny. For this reason, I performed a species-specific functional enrichment in order to check how many functions were characteristic of each taxon (i.e. *Microcoleus* or Coleofasciculaceae) and if there were "core functions" present in the genomes from the same family/genus. One hundred and sixty-seven functions were associated to one of the taxa: 111 to Coleofasciculaceae and 56 to *Microcoleus*. Only seven gene clusters associated to three functions were present in all Coleofasciculaceae and absent in all *Microcoleus* and none was present in all *Microcoleus* and absent in Coleofasciculaceae. According to this result, there are no core functions characteristic of each species, but this analysis only used the gene clusters that were associated to a COG function (61% of the gene clusters).

5.1.2 Sulphur-oxidising Phototrophic Alphaproteobacteria pangenome

Sulphide oxidising phototrophic Alphaproteobacteria were found in MAGs sets generated from both metagenomes - BSC and hyper-saline mats. I included in this group any Alphaproteobacteria that harboured in its genome any of the *soxABCXYZ* genes and chlorophyll synthesis genes. In total I collected 19 genomes; two were sequenced from BSC isolates, four were binned from the BSC metagenome and thirteen from the hyper-saline mat metagenome.

Taxonomic classification and completion estimation

GTDB-Tk placed the 19 genomes in class Alphaproteobacteria, in the orders: Kiloniellales (7), Rhodobacterales (4), Acetobacterales (4), Rhizobiales (3) and Azospirillales (1). Rhizobiales was the only order found in both habitats. CheckM was used to calculate completion and contamination values of the genomes with marker genes specific for the Alphaproteobacteria class. Four genomes were over 90% complete, six had completion values between 70% and 90%, nine genomes were between 50% and 70% complete. The sum of contamination and strain heterogeneity was below 5% in eleven genomes and between 5% and 10% in eight genomes.

Genomes completeness and taxonomy				
Genome name	GTDB taxonomy	Completness (%)	Contamination (%)	Strain heterogeneity (%)
Acetobacteraceae01	f Acetobacteraceae	82.80	2.34	3.89
Acetobacteraceae02	f Acetobacteraceae	68.81	1.16	2.59
AvdatX09	f Acetobacteraceae	95.27	0.75	0.00
AvdatX14	f Beijerinckiaceae	92.50	0.00	0.18
Craurococcus01	f Acetobacteraceae	97.96	1.00	0.00
Microvirga01	f Beijerinckiaceae	69.20	3.69	0.34
OSALTMAG03	f Rhodovibrionaceae	93.71	0.67	0.89
OSALTMAG129	f Rhodobacteraceae	79.28	3.52	0.25
OSALTMAG141	f Rhodovibrionaceae	62.31	4.15	2.97
OSALTMAG172	f Rhodovibrionaceae	77.37	1.80	3.59
OSALTMAG185	f Inquilinaceae	54.23	6.36	1.78
OSALTMAG187	f Rhodomicrobiaceae	64.87	4.79	1.15
OSALTMAG228	f Rhodobacteraceae	71.02	7.88	0.98
OSALTMAG275	f Rhodovibrionaceae	57.97	0.76	0.76
OSALTMAG52	f Rhodovibrionaceae	68.08	1.61	2.89
OSALTMAG56	f Rhodovibrionaceae	69.11	1.83	3.96
OSALTMAG71	f Rhodobacteraceae	77.65	2.21	1.47
OSALTMAG83	f Rhodobacteraceae	57.14	1.09	1.96
OSALTMAG89	f Rhodovibrionaceae	72.27	3.48	1.74

Table 5.2: Taxonomic placement and completion and contamination estimation of the analysed Alphaproteobacteria MAGs. All the MAGs were placed at family (f) level in the bacterial tree of life.

Pangenomic analysis

Anvi'o pangenomic workflow generated 34,936 gene clusters that were reduced to 11,698 after removing the singletons. I sorted the genomes based on their presence/absence pattern of gene clusters. The genomes split into two clusters: one formed by six Rhodovibrionaceae genomes belonging to the hyper-saline habitat, and the other one including the remaining genomes from both habitats. Interestingly, one Rhodovibrionaceae was functionally more similar to Rhodomicrobiaceae, Beijerinckiaceae and Inquilinaceae than to the other Rhodovibrionaceae. At lower ranks, the four Rhodobacteraceae (OSALT MAG129, 228, 71 and 83) and the four Acetobacteraceae (Acetobacteraceae 01 and 02, Craurococcus 01 and Avdat X09) clustered together respectively.

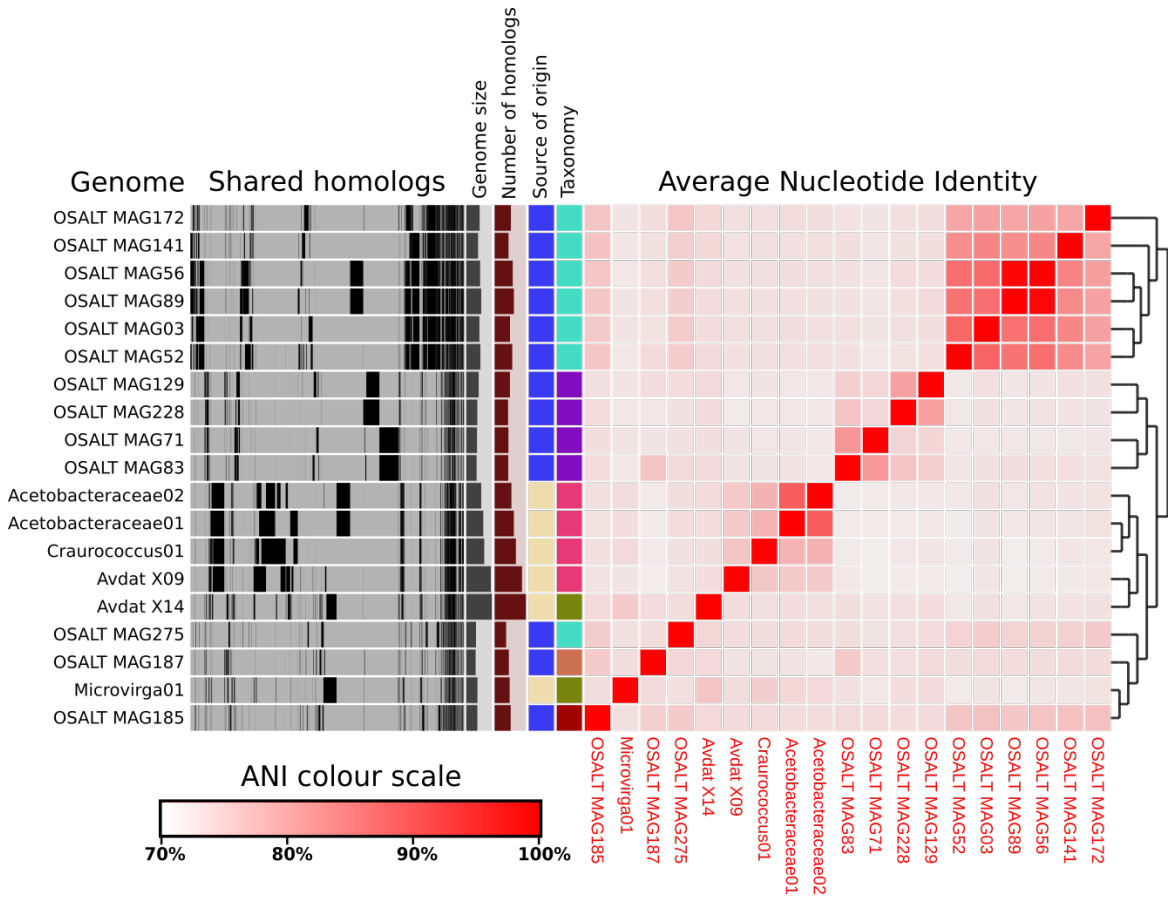

Figure 5.2: Anvi'o pangenomic analysis of 19 genomes belonging to the class Alphaproteobacteria binned from desert (brown) and saline (dark blue) metagenomes. The genomes belonged to the families: Rhodovibrionaceae (light blue), Rhodobacteraceae (purple), Acetobacteraceae (pink), Beijerinckiaceae (green), Rhodomicrobiaceae (brown) and Inquilinaceae (dark red). 50,576 genes were allocated in 11,698 gene clusters after removing singletons. ANIb was calculated with pyANI v0.2.10 [103] and ranges from 70% (white) to 100% (red). Genomes were ordered according to presence/absence of gene clusters. Dendrogram was built with Fasttree v2.1.10 [132].

Anvi'o enrichment analysis

Enrichment analysis of Alphaproteobacteria genomes produced 102 COG functions enriched in desert genomes, 15 of which were present in all desert genomes and in no saline one. Thirty-three functions were enriched in saline genomes, from which 7 were found in 12/13 saline genomes and in no desert genome.

Among the most enriched functions in MAGs from Negev Desert BSC some could be involved in coping with xeric stress according to literature research [1, 3, 7, 9, 133, 136, 137]: K^+ uptake protein Kup (COG3158) and Na^+/H^+ antiporter NhaA (COG3004). The most enriched functions in MAGs from the hyper-saline mats included: choline-glycine betaine transporter BetT (COG1292), Multisubunit Na^+/H^+ antiporter (subunit MnhC) (COG1006), ABC-type proline/glycine betaine transport system ProVW (COG4176) and Ca^{2+}/Na^+ antiporter ECM27 (COG0530) (Figure 5.3).

When genomes were sorted based on their presence/absence pattern of gene clusters, they grouped by families: six out of seven Rhodovibrionaceae together, the four Rhodobacteraceae together and three out of four Acetobacteraceae together (Figure 5.2). I looked for functions that are characteristic of each family and could explain this distribution. Two functions were found in all Rhodovibrionaceae and absent in the rest of the groups. Twenty four functions were found in all Acetobacteraceae and absent in the rest of the genomes. Seven functions were found in all Rhodobacteraceae and were absent in all the other families.

In contrast to the functional enrichment by habitat, in which the top enriched functions were related to xeric stress, family enriched functions were mostly linked to the cell cycle and metabolism: DNA-binding transcriptional regulator PadR, thiamine kinase CotS in Rhodovibrionaceae, 2-keto-4-pentenoate hydratase MhpD and β-glucosidase BglB in Acetobacteraceae, and putative NADPH-quinone reductase MdaB and Serine transporter YbeC in Rhodobacteraceae.

5.2 Discussion

5.2.1 Multi-locus sequence comparison is required for Cyanobacteria taxonomy classification

Prior to analysing their genomic potential, we placed all the MAGs in the GTDB tree. Out of the twelve genomes that were collected from the databases as "*Coleofasciculus*" only four were placed in the *Coleofasciculus* genus branch, seven were placed within the Coleofasciculaceae family and one was placed outside of it (Table 5.1). Cyanobacteria have been traditionally classified based on morphological traits that do not reflect their phylogeny, leading to polyphyletic taxa which require revision [138]. This was the case of *Microcoleus*, that was reclassified into several families and genera based on their 16S-23S ITS region [139], one of them being the genus *Coleofasciculus* [140]. In this study we could use the genomes even if they didn't belong to the genus *Coleofasciculus* since our aim was to look for habitat adaptations of Cyanobacteria with a similar lifestyle. Nevertheless, this result addresses the importance of using taxonomic classification methods based on concatenated marker proteins instead of only 16S rRNA sequence alignment.

Placement of the genomes in the GTDB tree was in concordance with the ANI values of the MAGs. ANI values among *Microcoleus* sequences were higher than those among "*Coleofasciculus*" and formed a solid cluster (Figure 5.1). ANI values within the *Microcoleus* cluster were above the 94% species threshold [141] in the seven genomes corresponding to desert *Microcoleus*, and dropped to 86.0-87.6% in the three *Microcoleus* genomes corresponding to saline and freshwater environments. GTDB-Tk placed the ten genomes at the same genus level.

Genomes that belonged to the family Coleofasciculaceae did not form a solid cluster. Furthermore, the desert genome Co250, that was classified by GTDB-Tk as PCC7133 (family Coleofasciculaceae), had slightly higher ANI values when compared to desert *Microcoleus* genomes (family Microcoleaceae) (ANI of 75.8% or higher) than to any saline or freshwater genome belonging to its family (ANI of 73.9% or lower) (Figure 5.1). This highlights again the limitations of 16S rRNA classification, according to which all the Coleofasciulaceae genomes belonged to the same genus.

5.2.2 Genomes are classified based on phylogeny instead of habitat

We sorted the genomes based on their gene cluster presence/absence pattern, expecting that those microorganisms inhabiting the same environment and facing the same stresses would have a more similar set of functions. However, the genomes grouped based on phylogeny and only at lower ranks split by habitat (Figures 5.1 and 5.2). It is possible that habitat adaptation is determined by few gene clusters. These clusters would have less influence in the presence/absence grouping than clusters containing core genes.

5.2.3 Catalases are taxon- and habitat-specific

Catalase activity is present in all organisms that are exposed to O_2; it protects them from ROS by decomposing H_2O_2 to H_2O and O_2 [142]. Enrichment analysis of the pangenomes showed that the catalase KatE was exclusively found in saline Cyanobacteria and in desert inhabiting Alphaproteobacteria (Figure 5.3). KatE is not exclusively found in xeric environments, as it is encoded by *Escherichia coli* [133]. Expression of *E. coli* KatE in the freshwater Cyanobacterium *Synechococcus* sp. PCC 7942 conferred it salt stress tolerance [133].

It is surprising that, even though the gene is present in both environments, it was conserved by different taxa in each one: by Cyanobacteria in saline environments and by Alphaproteobacteria in arid ones. This could be due to the presence of additional catalases in the pangenomes that could carry out the same function. The most common catalase (CotJC) in both genome collections contains Mn instead of haem group as prosthetic group. This catalase was widely present in both taxa and environment except for saline Alphaproteobacteria (Figure 5.3). Mn group has a higher thermal and oxidation stability than haem group, and is preferred over it in specific conditions such as extreme heat [142]. In my Scientific Practice, Mn-containing catalase was preferred over haem-catalases in desert and temperate environment inhabiting Alphaproteobacteria [143]. Our result has shown, in concordance, a preference for Mn-containing catalase in desert-inhabiting Alphaproteobacteria, but not in saline inhabiting ones.

Another catalase was encoded by the majority of the Alphaproteobacteria but it was absent in the Cyanobacteria pangenome, the haem-containing catalase KatG. It was the major catalase in saline Alphaproteobacteria, compensating for the absence of KatE and CotJC. No catalase was found in desert Cyanobacteria MAGs C3bin4 and SCS111. This could be due to their low completeness values, 55.11% and 65.88% respectively (Table 5.1).

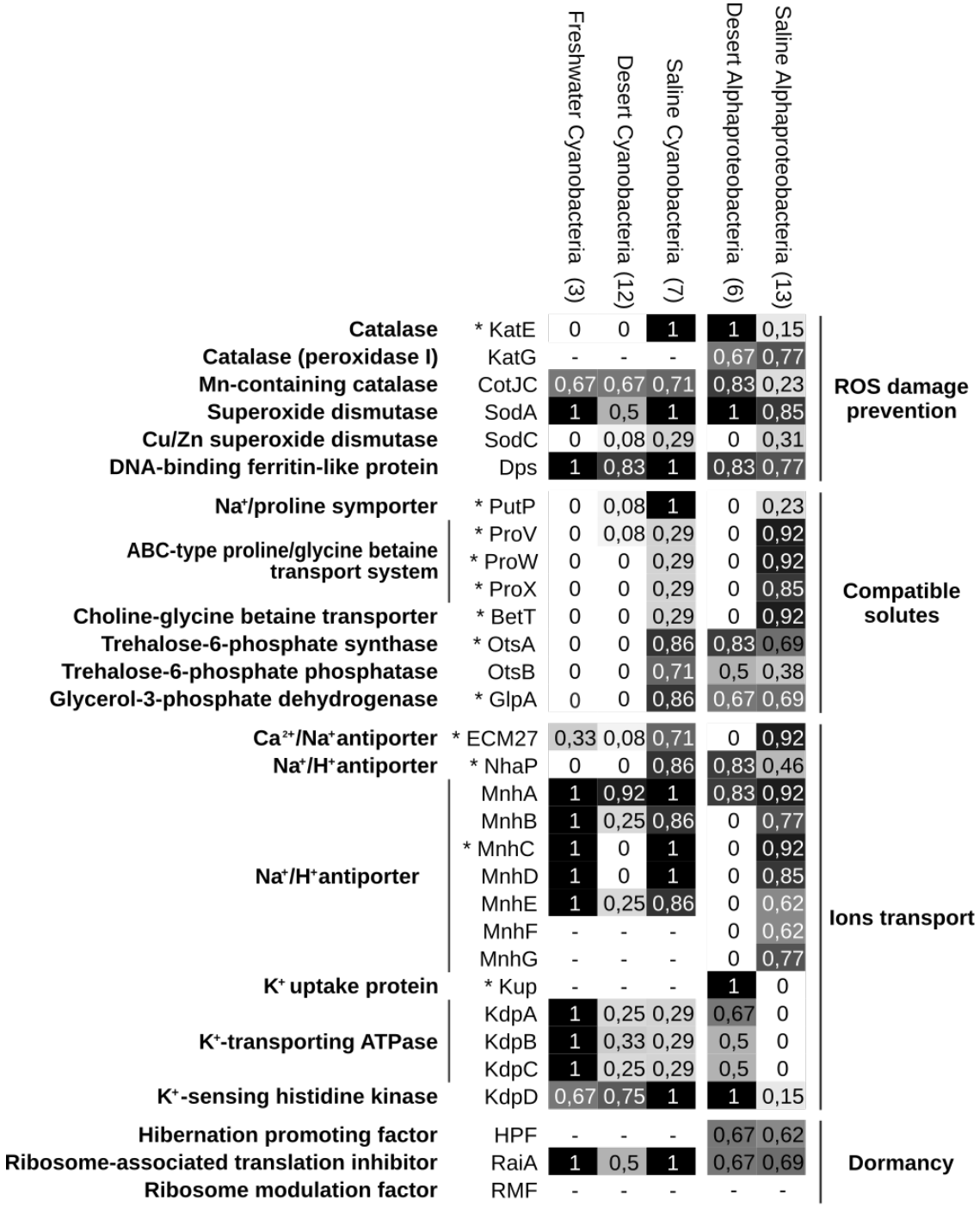

Figure 5.3: Proportion of genomes from arid, saline and freshwater environments that encode the selected xeric stress resistance genes. The number of genomes per group is indicated in parentheses. Genes enriched in any of the groups are indicated with an asterisk (*) adjacent to the name of the protein they encode.

5.2.4 Compatible solutes transporters were exclusively encoded by saline-inhabiting bacteria

Most microorganisms do not possess active water transporters [21]. Thus, in order to retain water under hyper-osmotic conditions they must accumulate osmolytes in their cytoplasm [1]. Compatible solutes are compounds that increase the osmolarity of the cytoplasm without interfering with the central metabolism [21]. We found several functions related to the biosynthesis, transport and accumulation of these compounds in both genome collections.

The proteins OtsA, OtsB and GlpA, associated to compatible solutes biosynthesis, were found in the saline Cyanobacteria and Alphaproteobacteria and the desert Alphaproteobacteria (Figure 5.3). OtsA and OtsB are involved in the synthesis of trehalose [134], whereas GlpA is the key enzyme of glycerol synthesis [135]. In *Saccharomyces cerevisiae*, glycerol production is the main response to osmotolerance and the expression of glycerol-3-phosphate deshydrogenase is enhanced under hyper-osmotic conditions [144].

Proteins associated to compatible solutes transport were only encoded in saline bacteria genomes. Saline Cyanobacteria were enriched in Na^+/proline symporter (PutP), whereas saline Alphaproteobacteria encoded transporters for choline-glycine betaine (ProVWX) and proline/glycine betaine (BetT), as well as the Ca^{2+}/Na^+ antiporter ECM27. The deletion of ECM27 in *S. cerevisiae* was linked to inability to maintain trehalose levels [137], and it could play a similar role in bacteria.

Compatible solutes have two mechanisms of action under xeric stress conditions: they retain water inside the cell by increasing the osmolarity [21] and they induce the vitrification of the cytoplasm by replacing water in its interactions with intracellular macromolecules [22], protecting their structure. Bacteria inhabiting arid environments could rely on vitrification instead of retaining water under desiccation conditions, when the extracellular osmolarity is too high to be countered. Furthermore, the transport of compatible solutes could be limited under desiccation conditions due to lack of a solvent. Thus, desiccated microorganisms would have to rely on endogenous biosynthesis despite being energetically unfavourable in comparison with transport [23].

5.2.5 Ion transporters were common in both habitats

Na^+/H^+ antiporters have several roles in bacteria [145]. The most important ones in hyper-osmotic environments are avoiding the accumulation of Na^+ in the cytoplasm, which is toxic in high concentrations, and establishing an electrochemical potential of Na^+ in order to couple it with the symport of compatible solutes [145]. Interestingly, the three groups of bacteria that had enriched functions were able to extrude Na^+, but the same function was carried out by different proteins. In desert Alphaproteobacteria and saline Cyanobacteria, this function was carried out by NhaP protein. However, saline Alphaproteobacteria encoded the complete multi-subunit protein Mnh instead. An incomplete Mnh complex was encoded by freshwater and saline Cyanobacteria too (Figure 5.3). Based on genomic information, saline Cyanobacteria have the potential to use Na^+ extrusion as a way to import proline, since the proteins NhaP and PutP were found in most of the genomes of the group.

The K^+ uptake protein Kup was exclusively found in all desert-inhabiting Alphaproteobacteria and absent in the rest of the genomes (Figure 5.3). Xerotolerant microorganisms import K^+ as a first response to osmotic shock, since it is faster than accumulation of compatible solutes [136]. Furthermore, Kup has been associated with the osmotic shock response in *E. coli* and *Bacillus subtilis* [136]. We looked in the pangenomes for additional K^+ uptake proteins associated with hyper-osmolarity in the literature [7]. The K^+ transporter Trk was not found in the pangenomes, whereas the K^+-transporting ATPase KdpABC was encoded by a fraction of the xerotolerant genomes and by the three freshwater Cyanobacteria genomes, suggesting that does not play a crucial role in adaptation to xeric conditions.

5.2.6 Assessing dormancy potential in the pangenomes

No metabolic dormancy protein was highlighted in the enrichment analysis. Nevertheless, we studied dormancy potential in the pangenomes by searching for the hibernation factors HPF, RMF and RaiA, which are reported to be ubiquitous within the bacterial domain [15]. The ribosome modulation factor (RMF) was not found in any of the pangenomes. This was expected since it has been reported to be confined to the Gammaproteobacteria [15]. The ribosome associated translation inhibitor (RaiA) was found in at least 50% of the genomes from both pangenomes, whereas the hibernation promoting factor (HPF) was only present in Alphaproteobacteria genomes. Both RaiA and HPF are homologs and induce ribosomal inactivation and growth arrest under certain circumstances, such as stress or transition to stationary phase [15]. RaiA binds to 70S ribosomes and inactivates them, whereas HPF induces 70S ribosomes dimerisation into 100S inactive dimers [15]. According to our results, most of the analysed genomes present metabolic dormancy potential regardless their habitat. Microbial dormancy is not restricted to extreme xeric environments; periods of nutrients starvation or unfavourable conditions can be experienced in any natural environment [146].

5.2.7 Perspectives: comparative metapangenomics as a way to study potential microbial survival strategies in xeric environments

We used the open source software platform Anvi'o to link the genomic content of closely related bacterial genomes to xeric environments adaptation. Our results show the feasibility of the method to associate genomic information to environmental data, in concordance with previous works [98, 99, 147]. However, the performance varied between the Cyanobacteria and the Alphaproteobacteria pangenomes. Our results were more informative for Alphaproteobacteria, showing potential exclusive strategies for arid- and saline-inhabiting microorganisms, whereas only saline-specific strategies could be inferred from the Cyanobacteria pangenome. Disparity in the results could be caused by different annotation coverage of the pangenomes: only 61% of the gene clusters (6619/10835) in the Cyanobacteria pangenome were associated to a COG20 function in comparison to 84.2% (9854/11698) in the Alphaproteobacteria pangenome. Acquired adaptation genes are encoded in the flexible genome, which is poorly annotated [66]. Therefore we could be missing a fraction of the xeric stress-related genes encoded in the pangenome. Furthermore, genes associated to the same COG function were split in several clusters, suggesting that not all the homologs could be grouped together. Therefore, Anvi'o analysis could perform better with genome collections related at species or genus level.

Another shortcoming of the analysis was experienced during the selection of genomes. We had enough phototrophic S-oxidising Alphaproteobacteria genomes from arid (Negev BSC) and hyper-saline (Oman microbial mats) environments. However, lack of sufficient hyper-saline inhabiting cyanobacteria genomes lead to the inclusion of a range of brackish and saline inhabiting cyanobacteria genomes (Table S1). Brackish water is more saline than freshwater, but cannot be considered an extreme environment that requires adaptations to hyper-osmolarity. Thus, the inclusion of brackish cyanobacteria into the "saline" category could partially explain why the Alphaproteobacteria comparison, in which only genomes from true xerotolerant and halotolerant bacteria were employed, was more informative than the Cyanobacteria comparison. It has been reported that 20% of the metagenomes produced for scientific publications are not publicly available [148]; efforts should be made to increase the accessibility of the genomic data in order to produce high quality comparative metagenomics surveys. We also included freshwater cyanobacteria genomes as a "negative control". The inclusion of genomes associated to non-xeric environments in the analysis could serve as a reference for organisms not adapted to low water activity. However, fresh water organisms inhabiting certain environments such as river beds can experience xeric stress periodically (e.g. the river dries out in summer) and thus be equipped with genes to cope with this stress. For this reason, metadata associated to publicly available genomes and MAGs needs to be as complete and precise as possible in order to link environmental conditions to genomic content.

The comparative pangenomics analysis allows to screen for xeric adaptation strategies in publicly available genomic data and formulate hypotheses. However, this strategy cannot replace cultivation-dependent methods. As shown in the discussion of this

work, the potential screened strategies are supported by low throughput *in vivo* studies. Although it goes beyond the scope of this thesis, downstream analysis can be carried out for each potential xeric-stress resistance gene cluster, assessing phylogenetic relations between the genes included in the cluster and looking for potential polymorphisms that could explain differences between habitats.

From this analysis we can conclude that phylogeny is more important than adaptation to the environment in terms of the genomic content of closely related bacteria. However, the most differentiated functions between environments were stress-associated, therefore how they cope with stress is the main difference between bacteria adapted to arid and saline environments. Additionally, we found strategies that were exclusively found in saline environments but not in arid ones, providing new support to the idea that desiccation and hyper-salinity are different stresses.

Chapter 6

Conclusions

Microorganisms have developed several strategies to cope with low water activity. However, response to low stress is heterogeneous and varies between taxa and habitats. In this thesis we studied microbial adaptations to low bioavailable water from different approaches: fluorescence *in situ* hybridisation and comparative genomics, identifying several limitations in the current methods used in microbial ecology research.

Studying xerotolerant microorganisms is a challenging task since their adaptation strategies could be interfering with molecular assays. We could not compare FISH intensity values between two assays using the same BSC samples due to variability associated to the FISH protocol. FISH signal is proportional to the number of ribosomes per cell. However, signal intensity can differ between assays due to variation in the reagents and human bias, which hinders the reproducibility of the method. The performance of CARD-FISH in hyper-saline microbial mats failed with specific probes and mat layer material. Thus, (CARD-)FISH protocols need to be optimised for complex environments undergoing extreme conditions in order to fully cover the microbial communities that inhabit them and understand their dynamics.

Advances in sequencing technologies allow us to study the vast fraction of the microbial world that cannot be grown in the laboratory. However, amplicon sequencing and metagenomic reads diversity estimations present biases that should be taken into consideration in order to assess the real microbial abundances in the community. Although the number of metagenomes and MAGs that are available in public databases is increasing vertiginously, parallel efforts should be made to improve the classification methods of the genomic data (by using multi-locus sequence alignment) and to include complete and accurate environmental metadata in order to make robust associations between microbial genomes and environmental conditions. Furthermore, functional genomics is limited by the high proportion of unknown coding sequences in the metagenomes as a result of relying on our knowledge about cultivable microorganisms. We need strategies to screen for potential functions in environmental metagenomes besides protein alignment to already known functions.

Despite the aforementioned limitations, a combined approach involving single-cell microscopy and computational biology can provide insights into microbial survival strategies in hostile environments. From our results we can conclude that protein synthesis potential increased at community and single-cell level in a BSC upon hydration and we did not find support to the hypothesis that long-term storage lead to ribosome degradation. We could verify disagreement between sequencing- and microscopy-based abundance estimations in the same microbial mat material. We concluded as well that the main difference between Alphaproteobacteria inhabiting arid and hyper-saline environments is how they cope with stress. Therefore, hyper-salinity and desiccation are different stresses and have different impact on the genome, despite leading both to lower water activity.

Chapter 7

Bibliography

[1] M Potts. Desiccation tolerance of prokaryotes. *Microbiological Reviews*, 58(4):755–805, December 1994.

[2] Stephen B. Pointing and Jayne Belnap. Microbial colonization and controls in dryland systems. *Nature Reviews Microbiology*, 10(8):551–562, August 2012.

[3] Pedro H. Lebre, Pieter De Maayer, and Don A. Cowan. Xerotolerant bacteria: Surviving through a dry spell. *Nature Reviews. Microbiology*, 15(5):285–296, May 2017.

[4] W. D. Grant. Life at low water activity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359(1448):1249–1266; discussion 1266–1267, August 2004.

[5] M. Chen and M. Alexander. Survival of soil bacteria during prolonged desiccation. *Soil Biology and Biochemistry*, 5(2):213–221, March 1973.

[6] Jason Bosch, Gilda Varliero, John E. Hallsworth, Tiffany D. Dallas, David Hopkins, Beat Frey, Weidong Kong, Pedro Lebre, Thulani P. Makhalanyane, and Don A. Cowan. Microbial anhydrobiosis. *Environmental Microbiology*, 23(11):6377–6390, 2021.

[7] Sabine Zeidler and Volker Müller. Coping with low water activities and osmotic stress in Acinetobacter baumannii: Significance, current status and perspectives. *Environmental Microbiology*, 21(7):2212–2230, 2019.

[8] Jan Vriezen, Frans Bruijn, and Klaus Nüsslein. Responses of Rhizobia to Desiccation in Relation to Osmotic Stress, Oxygen, and Temperature. *Applied and environmental microbiology*, 73:3451–9, June 2007.

[9] Nina Gunde-Cimerman, Ana Plemenitaš, and Aharon Oren. Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS Microbiology Reviews*, 42(3):353–375, May 2018.

[10] Roberto Orellana, Constanza Macaya, Guillermo Bravo, Flavia Dorochesi, Andrés Cumsille, Ricardo Valencia, Claudia Rojas, and Michael Seeger. Living at the Frontiers of Life: Extremophiles in Chile and Their Potential for Bioremediation. *Frontiers in Microbiology*, 9, 2018.

[11] Emily S. C. Rittershaus, Seung-hun Baek, and Christopher M. Sassetti. The Normalcy of Dormancy. *Cell host & microbe*, 13(6):643–651, June 2013.

[12] Jay T. Lennon and Stuart E. Jones. Microbial seed banks: The ecological and evolutionary implications of dormancy. *Nature Reviews Microbiology*, 9(2):119–130, February 2011.

[13] Pok Man Leung, Sean K. Bay, Dimitri V. Meier, Eleonora Chiri, Don A. Cowan, Osnat Gillor, Dagmar Woebken, and Chris Greening. Energetic Basis of Microbial Growth and Persistence in Desert Ecosystems. *mSystems*, 5(2), April 2020.

[14] E K Fuge, E L Braun, and M Werner-Washburne. Protein synthesis in long-term stationary-phase cultures of Saccharomyces cerevisiae. *Journal of Bacteriology*, 176(18):5802–5813, September 1994.

[15] Thomas Prossliner, Kristoffer Skovbo Winther, Michael Askvad Sørensen, and Kenn Gerdes. Ribosome Hibernation. *Annual Review of Genetics*, 52(1):321–348, November 2018.

[16] Jason M. Hansen, Young-Mi Go, and Dean P. Jones. Nuclear and mitochondrial compartmentation of oxidative stress and redox signaling. *Annual Review of Pharmacology and Toxicology*, 46:215–234, 2006.

[17] M. B. França, A. D. Panek, and E. C. A. Eleutherio. Oxidative stress and its effects during dehydration. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 146(4):621–631, April 2007.

[18] Lara Rajeev, Ulisses Nunes da Rocha, Niels Klitgord, Eric G. Luning, Julian Fortney, Seth D. Axen, Patrick M. Shih, Nicholas J. Bouskill, Benjamin P. Bowen, Cheryl A. Kerfeld, Ferran Garcia-Pichel, Eoin L. Brodie, Trent R. Northen, and Aindrila Mukhopadhyay. Dynamic cyanobacterial response to hydration and dehydration in a desert biological soil crust. *The ISME Journal*, 7(11):2178–2191, November 2013.

[19] A. Martinez and R. Kolter. Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps. *Journal of Bacteriology*, 179(16):5188–5194, August 1997.

[20] Breanne Shirkey, Don Paul Kovarcik, Deborah J. Wright, Gabriel Wilmoth, Todd F. Prickett, Richard F. Helm, Eugene M. Gregory, and Malcolm Potts. Active Fe-Containing Superoxide Dismutase and Abundant sodF mRNA in Nostoc commune (Cyanobacteria) after Years of Desiccation. *Journal of Bacteriology*, 182(1):189–197, January 2000.

[21] B. Kempf and E. Bremer. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Archives of Microbiology*, 170(5):319–330, October 1998.

[22] J. H. Crowe, J. F. Carpenter, and L. M. Crowe. The role of vitrification in anhydrobiosis. *Annual Review of Physiology*, 60:73–103, 1998.

[23] Aharon Oren. Bioenergetic Aspects of Halophilism. *Microbiology and Molecular Biology Reviews*, 63(2):334–348, June 1999.

[24] Erwin A. Galinski and Hans G. Trüper. Microbial behaviour in salt-stressed ecosystems. *FEMS Microbiology Reviews*, 15(2-3):95–108, October 1994.

[25] Stephan H. Saum, Friedhelm Pfeiffer, Peter Palm, Markus Rampp, Stephan C. Schuster, Volker Müller, and Dieter Oesterhelt. Chloride and organic osmolytes: A hybrid strategy to cope with elevated salinities by the moderately halophilic, chloride-dependent bacterium Halobacillus halophilus. *Environmental Microbiology*, 15(5):1619–1633, 2013.

[26] J. Wolfe and G. Bryant. Freezing, drying, and/or vitrification of membrane-solute-water systems. *Cryobiology*, 39(2):103–129, September 1999.

[27] Geoffrey R Brown, Iain C Sutcliffe, David Bendell, and Stephen P Cummings. The modification of the membrane of Oceanomonas baumanniiT when subjected to both osmotic and organic solvent stress. *FEMS Microbiology Letters*, 189(2):149–154, August 2000.

[28] L. J. Halverson and M. K. Firestone. Differential effects of permeating and nonpermeating solutes on the fatty acid composition of Pseudomonas putida. *Applied and Environmental Microbiology*, 66(6):2414–2421, June 2000.

[29] E. T. Houang, R. T. Sormunen, L. Lai, C. Y. Chan, and A. S. Leong. Effect of desiccation on the ultrastructural appearances of Acinetobacter baumannii and Acinetobacter lwoffii. *Journal of Clinical Pathology*, 51(10):786–788, October 1998.

[30] J. Belnap, B. Büdel, and O. L. Lange. Biological Soil Crusts: Characteristics and Distribution. In Jayne Belnap and Otto L. Lange, editors, *Biological Soil Crusts: Structure, Function, and Management*, Ecological Studies, pages 3–30. Springer, Berlin, Heidelberg, 2003.

[31] Wolfgang Elisabeth Krumbein, Anna A. Gorbushina, and Elisabeth Holtkamp-Tacken. Hypersaline Microbial Systems of Sabkhas: Examples of Life's Survival in "Extreme" Conditions. *Astrobiology*, 4(4):450–459, December 2004.

[32] Jayne Belnap. The world at your feet: Desert biological soil crusts. *Frontiers in Ecology and the Environment*, 1(4):181–189, 2003.

[33] Woo-Suk Chang, Martijn van de Mortel, Lindsey Nielsen, Gabriela Nino de Guzman, Xiaohong Li, and Larry J. Halverson. Alginate Production by Pseudomonas putida Creates a Hydrated Microenvironment and Contributes to Biofilm Architecture and Stress Tolerance under Water-Limiting Conditions. *Journal of Bacteriology*, 189(22):8290, November 2007.

[34] Henk Bolhuis, Mariana Silvia Cretoiu, and Lucas J. Stal. Molecular ecology of microbial mats. *FEMS Microbiology Ecology*, 90(2):335–350, November 2014.

[35] Hans van Gemerden. Microbial mats: A joint venture. *Marine Geology*, 113(1):3–25, July 1993.

[36] David J. Des Marais. The Biogeochemistry of Hypersaline Microbial Mats. In J. Gwynfryn Jones, editor, *Advances in Microbial Ecology*, pages 251–274. Springer US, Boston, MA, 1995.

[37] Lan-Zhou Chen, Dun-Hai Li, Li-Rong Song, Chun-Xiang Hu, Gao-Hong Wang, and Yong-Ding Liu. Effects of Salt Stress on Carbohydrate Metabolism in Desert Soil Alga Microcoleus vaginatus Gom. *Journal of Integrative Plant Biology*, 48(8):914–919, 2006.

[38] Oumarou Malam Issa, Christian Défarge, Yves Le Bissonnais, Béatrice Marin, Odile Duval, Ary Bruand, Luigi Paolo D'Acqui, Sofie Nordenberg, and Marcus Annerman. Effects of the inoculation of cyanobacteria on the microstructure and the structural stability of a tropical soil. *Plant and Soil*, 290(1):209–219, January 2007.

[39] Dimitri V. Meier, Stefanie Imminger, Osnat Gillor, and Dagmar Woebken. Distribution of Mixotrophy and Desiccation Survival Mechanisms across Microbial Genomes in an Arid Biological Soil Crust Community. *mSystems*, 6(1), February 2021.

[40] R I Amann, W Ludwig, and K H Schleifer. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, 59(1):143–169, March 1995.

[41] Rudolf Amann and Bernhard M. Fuchs. Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nature Reviews Microbiology*, 6(5):339–348, May 2008.

[42] Niculina Musat, Rachel Foster, Tomas Vagner, Birgit Adam, and Marcel M. M. Kuypers. Detecting metabolic activities in single cells, with emphasis on nanoSIMS. *FEMS Microbiology Reviews*, 36(2):486–511, March 2012.

[43] K. C. Schuster, E. Urlaub, and J. R. Gapes. Single-cell analysis of bacteria by Raman microscopy: Spectral information on the chemical composition of cells and on the heterogeneity in a culture. *Journal of Microbiological Methods*, 42(1):29–38, September 2000.

[44] Claude Lechene, Francois Hillion, Greg McMahon, Douglas Benson, Alan M Kleinfeld, J Patrick Kampf, Daniel Distel, Yvette Luyten, Joseph Bonventre, Dirk Hentschel, Kwon Moo Park, Susumu Ito, Martin Schwartz, Gilles Benichou, and Georges Slodzian. High-resolution quantitative imaging of mammalian and bacterial cells using stable isotope mass spectrometry. *Journal of Biology*, 5(6):20, 2006.

[45] Stephanie A. Eichorst, Florian Strasser, Tanja Woyke, Arno Schintlmeister, Michael Wagner, and Dagmar Woebken. Advancements in the application of NanoSIMS and Raman microspectroscopy to investigate the activity of microbial cells in soils. *FEMS Microbiology Ecology*, 91(10):fiv106, October 2015.

[46] Christine Nims, Brandi Cron, Maxwell Wetherington, Jennifer Macalady, and Julie Cosmidis. Low frequency Raman Spectroscopy for micron-scale and in vivo characterization of elemental sulfur in microbial samples. *Scientific Reports*, 9(1):7971, May 2019.

[47] Luke C. Burow, Dagmar Woebken, Ian PG Marshall, Erika A. Lindquist, Brad M. Bebout, Leslie Prufert-Bebout, Tori M. Hoehler, Susannah G. Tringe, Jennifer Pett-Ridge, Peter K. Weber, Alfred M. Spormann, and Steven W. Singer. Anoxic carbon flux in photosynthetic microbial mats as revealed by metatranscriptomics. *The ISME Journal*, 7(4):817–829, April 2013.

[48] Estelle Couradeau, Joelle Sasse, Danielle Goudeau, Nandita Nath, Terry C. Hazen, Ben P. Bowen, Romy Chakraborty, Rex R. Malmstrom, and Trent R. Northen. Probing the active fraction of soil microbiomes using BONCAT-FACS. *Nature Communications*, 10(1):2770, June 2019.

[49] R. I. Amann, L. Krumholz, and D. A. Stahl. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *Journal of Bacteriology*, February 1990.

[50] Holger Daims, Andreas Brühl, Rudolf Amann, Karl-Heinz Schleifer, and Michael Wagner. The Domain-specific Probe EUB338 is Insufficient for the Detection of all Bacteria: Development and Evaluation of a more Comprehensive Probe Set. *Systematic and Applied Microbiology*, 22(3):434–444, September 1999.

[51] D. Gade, H. Schlesner, F. O. Glöckner, R. Amann, S. Pfeiffer, and M. Thomm. Identification of Planctomycetes with Order-, Genus-, and Strain-Specific 16S rRNA-Targeted Probes. *Microbial Ecology*, 47(3):243–251, April 2004.

[52] L. Kerkhof and B. B. Ward. Comparison of Nucleic Acid Hybridization and Fluorometry for Measurement of the Relationship between RNA/DNA Ratio and Growth Rate in a Marine Bacterium. *Applied and Environmental Microbiology*, May 1993.

[53] L. K. Poulsen, G. Ballard, and D. A. Stahl. Use of rRNA fluorescence in situ hybridization for measuring the activity of single cells in young and established biofilms. *Applied and Environmental Microbiology*, May 1993.

[54] Annelie Pernthaler, Jakob Pernthaler, and Rudolf Amann. Fluorescence In Situ Hybridization and Catalyzed Reporter Deposition for the Identification of Marine Bacteria. *Applied and Environmental Microbiology*, June 2002.

[55] Tatsuhiko Hoshino, L Safak Yilmaz, Daniel R Noguera, Holger Daims, and Michael Wagner. Quantification of target molecules needed to detect microorganisms by fluorescence in situ hybridization (FISH) and catalyzed reporter deposition-FISH. *Applied and environmental microbiology*, 74(16):5068–5077, August 2008.

[56] David Andrew Fike, Crystal Lynn Gammon, Wiebke Ziebis, and Victoria Jeanne Orphan. Micron-scale mapping of sulfur cycling across the oxycline of a cyanobacterial mat: A paired nanoSIMS and CARD-FISH approach. *The ISME Journal*, 2(7):749–759, July 2008.

[57] Steven J. Blazewicz, Romain L. Barnard, Rebecca A. Daly, and Mary K. Firestone. Evaluating rRNA as an indicator of microbial activity in environmental communities: Limitations and uses. *The ISME Journal*, 7(11):2061–2068, November 2013.

[58] Lisa Boden. Microscopic insights into the microbial community of biological soil crusts from the Negev Desert, Israel. Master's thesis, Universität Wien, 2020.

[59] Boris Zarda, D. Hahn, Antonis Chatzinotas, Wilhelm Schönhuber, Alexander Neef, Rudolf I. Amann, and Josef Zeyer. Analysis of bacterial community structure in bulk soil by in situ hybridization. *Archives of Microbiology*, 168(3):185–192, September 1997.

[60] Henrik Christensen, Michael Hansen, and Jan Sørensen. Counting and Size Classification of Active Soil Bacteria by Fluorescence In Situ Hybridization with an rRNA Oligonucleotide Probe. *Applied and Environmental Microbiology*, April 1999.

[61] Mitchell L. Sogin, Hilary G. Morrison, Julie A. Huber, David Mark Welch, Susan M. Huse, Phillip R. Neal, Jesus M. Arrieta, and Gerhard J. Herndl. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences*, 103(32):12115–12120, August 2006.

[62] Douglas W. Fadrosh, Bing Ma, Pawel Gajer, Naomi Sengamalay, Sandra Ott, Rebecca M. Brotman, and Jacques Ravel. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, 2(1):6, February 2014.

[63] Masayuki Ushio, Kobayashi Makoto, Jonatan Klaminder, Hiroyuki Takasu, and Shin-ichi Nakano. High-throughput sequencing shows inconsistent results with a microscope-based analysis of the soil prokaryotic community. *Soil Biology and Biochemistry*, 76:53–56, September 2014.

[64] Tomáš Větrovský and Petr Baldrian. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS ONE*, 8(2):e57923, February 2013.

[65] Vito Pecoraro, Karolin Zerulla, Christian Lange, and Jörg Soppa. Quantification of Ploidy in Proteobacteria Revealed the Existence of Monoploid, (Mero-)Oligoploid and Polyploid Species. *PLOS ONE*, 6(1):e16392, January 2011.

[66] Martin F. Polz, Eric J. Alm, and William P. Hanage. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends in Genetics*, 29(3):170–175, March 2013.

[67] Hervé Tettelin, David Riley, Ciro Cattuto, and Duccio Medini. Comparative genomics: The bacterial pan-genome. *Current Opinion in Microbiology*, 11(5):472–477, October 2008.

[68] Omer Murik, Nadav Oren, Yoram Shotland, Hagai Raanan, Haim Treves, Isaac Kedem, Nir Keren, Martin Hagemann, Nadin Pade, and Aaron Kaplan. What distinguishes cyanobacteria able to revive after desiccation from those that cannot: The genome aspect. *Environmental Microbiology*, 19(2):535–550, 2017.

[69] Dimitri V. Meier, Andreas J. Greve, Arjun Chennu, Marit R. van Erk, Thirumahal Muthukrishnan, Raeid M. M. Abed, Dagmar Woebken, and Dirk de Beer. Limitation of Microbial Processes at Saturation-Level Salinities in a Microbial Mat Covering a Coastal Salt Flat. *Applied and Environmental Microbiology*, June 2021.

[70] Janina C. Vogt, Raeid M. M. Abed, Dirk C. Albach, and Katarzyna A. Palinska. Bacterial and Archaeal Diversity in Hypersaline Cyanobacterial Mats Along a Transect in the Intertidal Flats of the Sultanate of Oman. *Microbial Ecology*, 75(2):331–347, February 2018.

[71] Frank Oliver Glöckner, Rudolf Amann, Albin Alfreider, Jakob Pernthaler, Roland Psenner, Karlheinz Trebesius, and Karl-Heinz Schleifer. An In Situ Hybridization Protocol for Detection and Identification of Planktonic Bacteria. *Systematic and Applied Microbiology*, 19(3):403–406, October 1996.

[72] Kousuke Ishii, Marc Mußmann, Barbara J. MacGregor, and Rudolf Amann. An improved fluorescence in situ hybridization protocol for the identification of bacteria and archaea in marine sediments. *FEMS Microbiology Ecology*, 50(3):203–213, November 2004.

[73] R. I. Amann, B. J. Binder, R. J. Olson, S. W. Chisholm, R. Devereux, and D. A. Stahl. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and Environmental Microbiology*, June 1990.

[74] Andreas Neef. Anwendung der in situ-Einzelzell-identifizierung von bakterien zur populationsanalyse in komplexen mikrobiellen Biozonosen. *Ph. D. thesis. Technische Universitat Munichen*, 1997.

[75] Lovisa Björnsson, Philip Hugenholtz, Gene W. Tyson, and Linda L.YR 2002 Blackall. Filamentous Chloroflexi (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal. *Microbiology*, 148(8):2309–2318, 2002.

[76] Josefa Antón, Ramón Rosselló-Mora, Francisco Rodríguez-Valera, and Rudolf Amann. Extremely Halophilic Bacteria in Crystallizer Ponds from Solar Salterns. *Applied and Environmental Microbiology*, 66(7):3052, July 2000.

[77] Günter Wallner, Rudolf Amann, and Wolfgang Beisker. Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry*, 14(2):136–143, 1993.

[78] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean-Yves Tinevez, Daniel James White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak, and Albert Cardona. Fiji: An open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682, July 2012.

[79] Holger Daims, Sebastian Lücker, and Michael Wagner. Daime, a novel image analysis program for microbial ecology and biofilm research. *Environmental Microbiology*, 8(2):200–213, 2006.

[80] Philip Dixon. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, 14(6):927–930, 2003.

[81] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[82] RStudio Team. *RStudio: Integrated Development Environment for r*. RStudio, PBC., Boston, MA, 2020.

[83] National Center for Biotechnology Information, U. S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, and 20894 Usa. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/.

[84] Sean K. Bay, David W. Waite, Xiyang Dong, Osnat Gillor, Steven L. Chown, Philip Hugenholtz, and Chris Greening. Chemosynthetic and photosynthetic bacteria contribute differentially to primary production across a steep desert aridity gradient. *The ISME Journal*, pages 1–18, May 2021.

[85] Anton Bankevich, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, Sergey I. Nikolenko, Son Pham, Andrey D. Prjibelski, Alexey V. Pyshkin, Alexander V. Sirotkin, Nikolay Vyahhi, Glenn Tesler, Max A. Alekseyev, and Pavel A. Pevzner. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, 19(5):455–477, May 2012.

[86] Sergey Nurk, Dmitry Meleshko, Anton Korobeynikov, and Pavel A. Pevzner. metaSPAdes: A new versatile metagenomic assembler. *Genome Research*, 27(5):824–834, January 2017.

[87] Brian Bushnell, Jonathan Rood, and Esther Singer. BBMerge – Accurate paired shotgun read merging via overlap. *PLOS ONE*, 12(10):e0185056, October 2017.

[88] Pierre-Alain Chaumeil, Aaron J. Mussig, Philip Hugenholtz, and Donovan H. Parks. GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, 36(6):1925–1927, March 2020.

[89] Genome Taxonomy Database. https://gtdb.ecogenomic.org/.

[90] Donovan H. Parks, Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W. Tyson. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25(7):1043–1055, January 2015.

[91] Doug Hyatt, Gwo-Liang Chen, Philip F. Locascio, Miriam L. Land, Frank W. Larimer, and Loren J. Hauser. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, 11:119, March 2010.

[92] Sara El-Gebali, Jaina Mistry, Alex Bateman, Sean R. Eddy, Aurélien Luciani, Simon C. Potter, Matloob Qureshi, Lorna J. Richardson, Gustavo A. Salazar, Alfredo Smart, Erik L. L. Sonnhammer, Layla Hirsh, Lisanna Paladin, Damiano Piovesan, Silvio C. E. Tosatto, and Robert D. Finn. The Pfam protein families database in 2019. *Nucleic Acids Research*, 47(D1):D427–D432, January 2019.

[93] M. Kanehisa and S. Goto. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, January 2000.

[94] Roman L. Tatusov, Michael Y. Galperin, Darren A. Natale, and Eugene V. Koonin. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 28(1):33–36, January 2000.

[95] Benjamin Buchfink, Chao Xie, and Daniel H. Huson. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1):59–60, January 2015.

[96] Jaime Huerta-Cepas, Kristoffer Forslund, Luis Pedro Coelho, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering, and Peer Bork. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, 34(8):2115–2122, August 2017.

[97] Jaime Huerta-Cepas, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K. Forslund, Helen Cook, Daniel R. Mende, Ivica Letunic, Thomas Rattei, Lars J. Jensen, Christian von Mering, and Peer Bork. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1):D309–D314, January 2019.

[98] A. Murat Eren, Özcan C. Esen, Christopher Quince, Joseph H. Vineis, Hilary G. Morrison, Mitchell L. Sogin, and Tom O. Delmont. Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ*, 3:e1319, October 2015.

[99] Tom O. Delmont and A. Murat Eren. Linking pangenomes and metagenomes: The Prochlorococcus metapangenome. *PeerJ*, 6:e4320, January 2018.

[100] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, October 1990.

[101] Matthew N. Benedict, James R. Henriksen, William W. Metcalf, Rachel J. Whitaker, and Nathan D. Price. ITEP: An integrated toolkit for exploration of microbial pan-genomes. *BMC Genomics*, 15(1):8, January 2014.

[102] Stijn Van Dongen. Graph Clustering Via a Discrete Uncoupling Process. *SIAM Journal on Matrix Analysis and Applications*, 30(1):121–141, January 2008.

[103] Leighton Pritchard, Rachel H. Glover, Sonia Humphris, John G. Elphinstone, and Ian K. Toth. Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Analytical Methods*, 8(1):12–24, December 2015.

[104] Johan Goris, Konstantinos T Konstantinidis, Joel A Klappenbach, Tom Coenye, Peter Vandamme, and James M Tiedje. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International journal of systematic and evolutionary microbiology*, 57(Pt 1):81–91, January 2007.

[105] Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402, September 1997.

[106] Knud Nierhaus and Daniel Wilson. *Protein Synthesis and Ribosome Structure*. Wiley-VCH, January 2006.

[107] Michael A. Zundel, Georgeta N. Basturea, and Murray P. Deutscher. Initiation of ribosome degradation during starvation in Escherichia coli. *RNA*, 15(5):977–983, January 2009.

[108] Hideji Yoshida and Akira Wada. The 100S ribosome: Ribosomal hibernation induced by stress. *WIREs RNA*, 5(5):723–732, 2014.

[109] Assaf Sukenik, Ruth N. Kaplan-Levy, Jessica Mark Welch, and Anton F. Post. Massive multiplication of genome and ribosomes in dormant cells (akinetes) of Aphanizomenon ovalisporum (Cyanobacteria). *The ISME Journal*, 6(3):670–679, March 2012.

[110] Adam Šťovíček, Minsu Kim, Dani Or, and Osnat Gillor. Microbial community response to hydration-desiccation cycles in desert soil. *Scientific Reports*, 7(1), May 2017.

[111] Tatsuya Akiyama, Kerry S. Williamson, Robert Schaefer, Shawna Pratt, Connie B. Chang, and Michael J. Franklin. Resuscitation of Pseudomonas aeruginosa from dormancy requires hibernation promoting factor (PA4463) for ribosome preservation. *Proceedings of the National Academy of Sciences*, 114(12):3204–3209, March 2017.

[112] Geneviève Filion, Christian Laflamme, Nathalie Turgeon, Jim Ho, and Caroline Duchaine. Permeabilization and hybridization protocols for rapid detection of Bacillus spores using fluorescence in situ hybridization. *Journal of Microbiological Methods*, 77(1):29–36, April 2009.

[113] Donald E. Canfield and David J. Des Marais. Biogeochemical cycles of carbon, sulfur, and free oxygen in a microbial mat. *Geochimica et Cosmochimica Acta*, 57(16):3971–3984, August 1993.

[114] Harald R. Gruber-Vodicka, Brandon K. B. Seah, and Elmar Pruesse. phyloFlash: Rapid Small-Subunit rRNA Profiling and Targeted Assembly from Metagenomes. *mSystems*, 5(5):e00920–20, October 2020.

[115] Dagmar Rother, Hans-Jürgen Henrich, Armin Quentmeier, Frank Bardischewsky, and Cornelius G. Friedrich. Novel Genes of the sox Gene Cluster, Mutagenesis of the Flavoprotein SoxF, and Evidence for a General Sulfur-Oxidizing System in Paracoccus pantotrophusGB17. *Journal of Bacteriology*, 183(15):4499–4508, August 2001.

[116] Stefan M Sievert, Wiebke Ziebis, Jan Kuever, and Kerstin Sahm. Relative abundance of Archaea and Bacteria along a thermal gradient of a shallow-water hydrothermal vent quantified by rRNA slot-blot hybridization. *Microbiology (Reading, England)*, 146 (Pt 6):1287–1293, June 2000.

[117] Susana Benlloch, Arantxa López-López, Emilio O. Casamayor, Lise Øvreås, Victoria Goddard, Frida Lise Daae, Gary Smerdon, Ramón Massana, Ian Joint, Frede Thingstad, Carlos Pedrós-Alió, and Francisco Rodríguez-Valera. Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. *Environmental Microbiology*, 4(6):349–360, 2002.

[118] Kasia Piwosz, Tanja Shabarova, Jürgen Tomasch, Karel Šimek, Karel Kopejtka, Silke Kahl, Dietmar H. Pieper, and Michal Koblížek. Determining lineage-specific bacterial growth curves with a novel approach based on amplicon reads normalization using internal standard (ARNIS). *The ISME Journal*, 12(11):2640–2654, November 2018.

[119] Kasia Piwosz, Tanja Shabarova, Jakob Pernthaler, Thomas Posch, Karel Šimek, Petr Porcal, and Michaela M. Salcher. Bacterial and Eukaryotic Small-Subunit Amplicon Data Do Not Provide a Quantitative Picture of Microbial Communities, but They Are Reliable in the Context of Ecological Interpretations. *mSphere*, March 2020.

[120] Silvia G. Acinas, Ramahi Sarma-Rupavtarm, Vanja Klepac-Ceraj, and Martin F. Polz. PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample. *Applied and Environmental Microbiology*, 71(12):8966–8969, December 2005.

[121] Silvia G. Acinas, Isabel Ferrera, Hugo Sarmento, Cristina Díez-Vives, Irene Forn, Clara Ruiz-González, Francisco M. Cornejo-Castillo, Guillem Salazar, and Josep M. Gasol. Validation of a new catalysed reporter deposition–fluorescence in situ hybridization probe for the accurate quantification of marine Bacteroidetes populations. *Environmental Microbiology*, 17(10):3557–3569, 2015.

[122] Sebastian Behrens, Caroline Rühland, Joao Inacio, Harald Huber, À.Á Fonseca, I Spencer-Martins, Bernhard Fuchs, and Rudolf Amann. In Situ Accessibility of Small-Subunit rRNA of Members of the Domains Bacteria, Archaea, and Eucarya to Cy3-Labeled Oligonucleotide Probes. *Applied and environmental microbiology*, 69:1748–58, April 2003.

[123] W. Schönhuber, B. Fuchs, S. Juretschko, and R. Amann. Improved sensitivity of whole-cell hybridization by the combination of horseradish peroxidase-labeled oligonucleotides and tyramide signal amplification. *Applied and Environmental Microbiology*, 63(8):3268–3273, August 1997.

[124] Yadhu Kumar, Ralf Westram, Sebastian Behrens, Bernhard Fuchs, Frank Oliver Glöckner, Rudolf Amann, Harald Meier, and Wolfgang Ludwig. Graphical representation of ribosomal RNA probe accessibility data using ARB software package. *BMC Bioinformatics*, 6:61, March 2005.

[125] Raju Sekar, Annelie Pernthaler, Jakob Pernthaler, Falk Warnecke, Thomas Posch, and Rudolf Amann. An improved protocol for quantification of freshwater Actinobacteria by fluorescence in situ hybridization. *Applied and Environmental Microbiology*, 69(5):2928–2935, May 2003.

[126] M. Piuri, C. Sanchez-Rivas, and S.m. Ruzal. Cell wall modifications during osmotic stress in Lactobacillus casei. *Journal of Applied Microbiology*, 98(1):84–95, 2005.

[127] Aude Fourçans, Tirso García de Oteyza, Andrea Wieland, Antoni Solé, Elia Diestra, Judith van Bleijswijk, Joan O. Grimalt, Michael Kühl, Isabel Esteve, Gerard Muyzer, Pierre Caumette, and Robert Duran. Characterization of functional bacterial groups in a hypersaline microbial mat community (Salins-de-Giraud, Camargue, France). *FEMS microbiology ecology*, 51(1):55–70, December 2004.

[128] Aude Fourçans, Antoni Solé, Ella Diestra, Anthony Ranchou-Peyruse, Isabel Esteve, Pierre Caumette, and Robert Duran. Vertical migration of phototrophic bacterial populations in a hypersaline microbial mat from Salins-de-Giraud (Camargue, France). *FEMS Microbiology Ecology*, 57(3):367–377, 2006.

[129] Federico M. Ibarbalz, María Victoria Pérez, Eva L. M. Figuerola, and Leonardo Erijman. The Bias Associated with Amplicon Sequencing Does Not Affect the Quantitative Assessment of Bacterial Community Dynamics. *PLOS ONE*, 9(6):e99722, June 2014.

[130] Lira A. Gaysina, Aniket Saraf, and Prashant Singh. Chapter 1 - Cyanobacteria in Diverse Habitats. In A. K. Mishra, D. N. Tiwari, and A. N. Rai, editors, *Cyanobacteria*, pages 1–28. Academic Press, January 2019.

[131] Alexander Sczyrba, Peter Hofmann, Peter Belmann, David Koslicki, Stefan Janssen, Johannes Dröge, Ivan Gregor, Stephan Majda, Jessika Fiedler, Eik Dahms, Andreas Bremges, Adrian Fritz, Ruben Garrido-Oter, Tue Sparholt Jørgensen, Nicole Shapiro, Philip D Blood, Alexey Gurevich, Yang Bai, Dmitrij Turaev, Matthew Z DeMaere, Rayan Chikhi, Niranjan Nagarajan, Christopher Quince, Fernando Meyer, Monika Balvočiūtė, Lars Hestbjerg Hansen, Søren J Sørensen, Burton K H Chia, Bertrand Denis, Jeff L Froula, Zhong Wang, Robert Egan, Dongwan Don Kang, Jeffrey J Cook, Charles Deltel, Michael Beckstette, Claire Lemaitre, Pierre Peterlongo, Guillaume Rizk, Dominique Lavenier, Yu-Wei Wu, Steven W Singer, Chirag Jain, Marc Strous, Heiner Klingenberg, Peter Meinicke, Michael D Barton, Thomas Lingner, Hsin-Hung Lin, Yu-Chieh Liao, Genivaldo Gueiros Z Silva, Daniel A Cuevas, Robert A Edwards, Surya Saha, Vitor C Piro, Bernhard Y Renard, Mihai Pop, Hans-Peter Klenk, Markus Göker, Nikos C Kyrpides, Tanja Woyke, Julia A Vorholt, Paul Schulze-Lefert, Edward M Rubin, Aaron E Darling, Thomas Rattei, and Alice C McHardy. Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software. *Nature Methods*, 14(11):1063–1071, November 2017.

[132] Morgan N. Price, Paramvir S. Dehal, and Adam P. Arkin. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE*, 5(3):e9490, March 2010.

[133] Nobuo Kaku, Takashi Hibino, Yoshito Tanaka, Hiroshi Ishikawa, Etsuko Araki, Tetsuko Takabe, and Teruhiro Takabe. Effects of overexpression of Escherichia coli katE and bet genes on the tolerance for salt stress in a freshwater cyanobacterium Synechococcus sp. PCC 7942. *Plant Science*, 159(2):281–288, November 2000.

[134] Sabine Zeidler, Josephine Hubloher, Kim Schabacker, Pedro Lamosa, Helena Santos, and Volker Müller. Trehalose, a temperature- and salt-induced solute with implications in pathobiology of Acinetobacter baumannii. *Environmental Microbiology*, 19(12):5088–5099, 2017.

[135] J. Albertyn, S. Hohmann, J. M. Thevelein, and B. A. Prior. GPD1, which encodes glycerol-3-phosphate dehydrogenase, is essential for growth under osmotic stress in Saccharomyces cerevisiae, and its expression is regulated by the high-osmolarity glycerol response pathway. *Molecular and Cellular Biology*, June 1994.

[136] I.R. Booth and C.F. Higgins. Enteric bacteria and osmotic stress: Intracellular potassium glutamate as a secondary signal of osmotic stress? *FEMS Microbiology Reviews*, 6(2-3):239–246, June 1990.

[137] Rachel Klukovich and William E. Courchesne. Functions of Saccharomyces cerevisiae Ecm27p, a putative Na+/Ca2+ exchanger, in calcium homeostasis, carbohydrate storage and cell cycle reentry from the quiescent phase. *Microbiological Research*, 186–187:81–89, May 2016.

[138] Juline M. Walter, Felipe H. Coutinho, Bas E. Dutilh, Jean Swings, Fabiano L. Thompson, and Cristiane C. Thompson. Ecogenomics and Taxonomy of Cyanobacteria Phylum. *Frontiers in Microbiology*, 8:2132, November 2017.

[139] Sarah L. Boyer, Jeffrey R. Johansen, Valerie R. Flechtner, and Gwyn L. Howard. Phylogeny and genetic variance in terrestrial Microcoleus (Cyanophyceae) species based on sequence analysis of the 16S rRNA gene and associated 16S–23S ITS region 1. *Journal of Phycology*, 38(6):1222–1235, 2002.

[140] Maria A. Siegesmund, Jeffrey R. Johansen, Ulf Karsten, and Thomas Friedl. Coleofasciculus Gen. Nov. (cyanobacteria): Morphological and Molecular Criteria for Revision of the Genus Microcoleus Gomont1. *Journal of Phycology*, 44(6):1572–1585, 2008.

[141] Michael Richter and Ramon Rosselló-Móra. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences*, October 2009.

[142] James W. Whittaker. Non-heme manganese catalase – the 'other' catalase. *Archives of biochemistry and biophysics*, 525(2):111–120, September 2012.

[143] J. Bosco Gracia-Alvira. *Genome Comparison between Desert and Temperate Environment Bacteria Reveals Potential Desiccation Stress Resistance Genes.* Scientific Practice, Universität Wien, February 2021.

[144] A. Blomberg and L. Adler. Roles of glycerol and glycerol-3-phosphate dehydrogenase (NAD+) in acquired osmotolerance of Saccharomyces cerevisiae. *Journal of Bacteriology*, February 1989.

[145] Toshiaki Hiramatsu, Kazuyo Kodama, Teruo Kuroda, Tohru Mizushima, and Tomofusa Tsuchiya. A Putative Multisubunit Na+/H+ Antiporter fromStaphylococcus aureus. *Journal of Bacteriology*, 180(24):6642–6648, December 1998.

[146] Raphael Trösch and Felix Willmund. The conserved theme of ribosome hibernation: From bacteria to chloroplasts of plants. *Biological Chemistry*, 400(7):879–893, July 2019.

[147] Tom O. Delmont. Environmental genomics points to non-diazotrophic Trichodesmium species abundant and widespread in the open ocean. *bioRxiv*, page 2021.03.24.436785, March 2021.

[148] Ester M. Eckert, Andrea Di Cesare, Diego Fontaneto, Thomas U. Berendonk, Helmut Bürgmann, Eddie Cytryn, Despo Fatta-Kassinos, Andrea Franzetti, D. G. Joakim Larsson, Célia M. Manaia, Amy Pruden, Andrew C. Singer, Nikolina Udikovic-Kolic, and Gianluca Corno. Every fifth published metagenome is not available to science. *PLOS Biology*, 18(4):e3000698, April 2020.

[149] Nadia Eusebio, Adriana Rego, Nathaniel R. Glasser, Raquel Castelo-Branco, Emily P. Balskus, and Pedro N. Leão. Distribution and diversity of dimetal-carboxylate halogenases in cyanobacteria. *BMC genomics*, 22(1):633, August 2021.

[150] Meng-Yun Chen, Wen-Kai Teng, Liang Zhao, Chun-Xiang Hu, Yang-Kai Zhou, Bo-Ping Han, Li-Rong Song, and Wen-Sheng Shu. Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. *The ISME journal*, 15(1):211–227, January 2021.

[151] Karen Jordaan, Rachael Lappan, Xiyang Dong, Ian J. Aitkenhead, Sean K. Bay, Eleonora Chiri, Nimrod Wieler, Laura K. Meredith, Don A. Cowan, Steven L. Chown, and Chris Greening. Hydrogen-Oxidizing Bacteria Are Abundant in Desert Soils and Strongly Stimulated by Hydration. *mSystems*, 5(6), December 2020.

[152] H. S. Tee, D. Waite, L. Payne, M. Middleditch, S. Wood, and K. M. Handley. Tools for successful proliferation: Diverse strategies of nutrient acquisition by a benthic cyanobacterium. *The ISME journal*, 14(8):2164–2178, August 2020.

Figures list

Tables list

Appendices

Genomes list				
Bin name	Habitat	Location	Pangenome	Accession
OSALTMAG134	Hypersaline	Oman sabkha[69]	Cyanobacteria	PRJNA653500[c]
OSALTMAG256	Hypersaline	Oman sabkha[69]	Cyanobacteria	PRJNA653498[c]
OSALTMAG82	Hypersaline	Oman sabkha[69]	Cyanobacteria	PRJNA653498[c]
PCC7420	Saline	Salt marsh	Cyanobacteria	GCA_000155555.1
LEGE07081	Brackish	Mesotidal zone[149]	Cyanobacteria	GCA_015207445.1
LEGE07092	Brackish	Mesotidal zone[149]	Cyanobacteria	GCA_015207375.1
FACHB1120	Freshwater	Freshwater[150]	Cyanobacteria	GCA_014698845.1
Coleofasciculus01	Arid	Negev desert[39]	Cyanobacteria	GCA_902805485.1
Co250	Arid	Negev desert[84]	Cyanobacteria	-[a]
SCS111	Arid	Negev desert[84]	Cyanobacteria	-[a]
Cobin14	Arid	Arid sandy surface[151]	Cyanobacteria	GCA_014534335.1
C3bin4	Arid	Arid sandy surface[151]	Cyanobacteria	GCA_014534505.1
LEGE07076	Brackish	Mesotidal zone[149]	Cyanobacteria	GCA_015207455.1
bin38	Freshwater	Riverbed cobble biofilm[152]	Cyanobacteria	GCA_013299165.1
bin48	Freshwater	Riverbed cobble biofilm[152]	Cyanobacteria	GCA_013299185.1
Microcoleus01	Arid	Negev desert[39]	Cyanobacteria	GCA_902805765.1
SCS153	Arid	Negev desert[84]	Cyanobacteria	-[a]
SCS76	Arid	Negev desert[84]	Cyanobacteria	-[a]
C1bin4	Arid	Arid sandy surface[151]	Cyanobacteria	GCA_014534575.1
Cobin12	Arid	Arid sandy surface[151]	Cyanobacteria	GCA_014534405.1
T1bin1	Arid	Arid sandy surface[151]	Cyanobacteria	GCA_014534015.1
T3bin5	Arid	Arid sandy surface[151]	Cyanobacteria	GCA_014533905.1
OSALTMAG03	Hypersaline	Oman sabkha[69]	Alphaproteobacteria	PRJNA653497[c]
OSALTMAG141	Hypersaline	Oman sabkha[69]	Alphaproteobacteria	PRJNA653497[c]
OSALTMAG52	Hypersaline	Oman sabkha[69]	Alphaproteobacteria	PRJNA653499[c]
OSALTMAG187	Hypersaline	Oman sabkha[69]	Alphaproteobacteria	PRJNA653500[c]
OSALTMAG129	Hypersaline	Oman sabkha[69]	Alphaproteobacteria	PRJNA653500[c]
OSALTMAG172	Hypersaline	Oman sabkha[69]	Alphaproteobacteria	PRJNA653501[c]
OSALTMAG185	Hypersaline	Oman sabkha[69]	Alphaproteobacteria	PRJNA653501[c]
OSALTMAG56	Hypersaline	Oman sabkha[69]	Alphaproteobacteria	PRJNA653503[c]
OSALTMAG228	Hypersaline	Oman sabkha[69]	Alphaproteobacteria	PRJNA653505[c]
OSALTMAG83	Hypersaline	Oman sabkha[69]	Alphaproteobacteria	PRJNA653506[c]
OSALTMAG89	Hypersaline	Oman sabkha[69]	Alphaproteobacteria	PRJNA653506[c]
OSALTMAG71	Hypersaline	Oman sabkha[69]	Alphaproteobacteria	PRJNA653506[c]
OSALTMAG275	Hypersaline	Oman sabkha[69]	Alphaproteobacteria	PRJNA653506[c]
Microvirga01	Arid	Negev desert[39]	Alphaproteobacteria	GCA_902805785.1
Craurococcus01	Arid	Negev desert[39]	Alphaproteobacteria	GCA_902805545.1
Acetobacteraceae01	Arid	Negev desert[39]	Alphaproteobacteria	GCA_902805645.1
Acetobacteraceae02	Arid	Negev desert[39]	Alphaproteobacteria	GCA_902805625.1
AvdatX09	Arid	Negev desert[39]	Alphaproteobacteria	-[b]
AvdatX14	Arid	Negev desert[39]	Alphaproteobacteria	-[b]

Table S1: List of genomes included in the pangenomic analyses. The assembly accession number or the Bioproject accession number (if assembly is not available) are provided for each genome. Some genomes were not available in NCBI. Co250, SCS111, SCS153 and SCS76 were obtained from the supplementary material of Bay *et al.* 2021 [84] ([a]). AvdatX09 and AvdatX14 were isolated in pure culture, sequenced and assembled by Dr. Dimitri Meier ([b]). MAGs from hyper-saline mats were assembled and binned by Dr. Dimitri Meier and assembly accession is not available in NCBI yet ([c]).

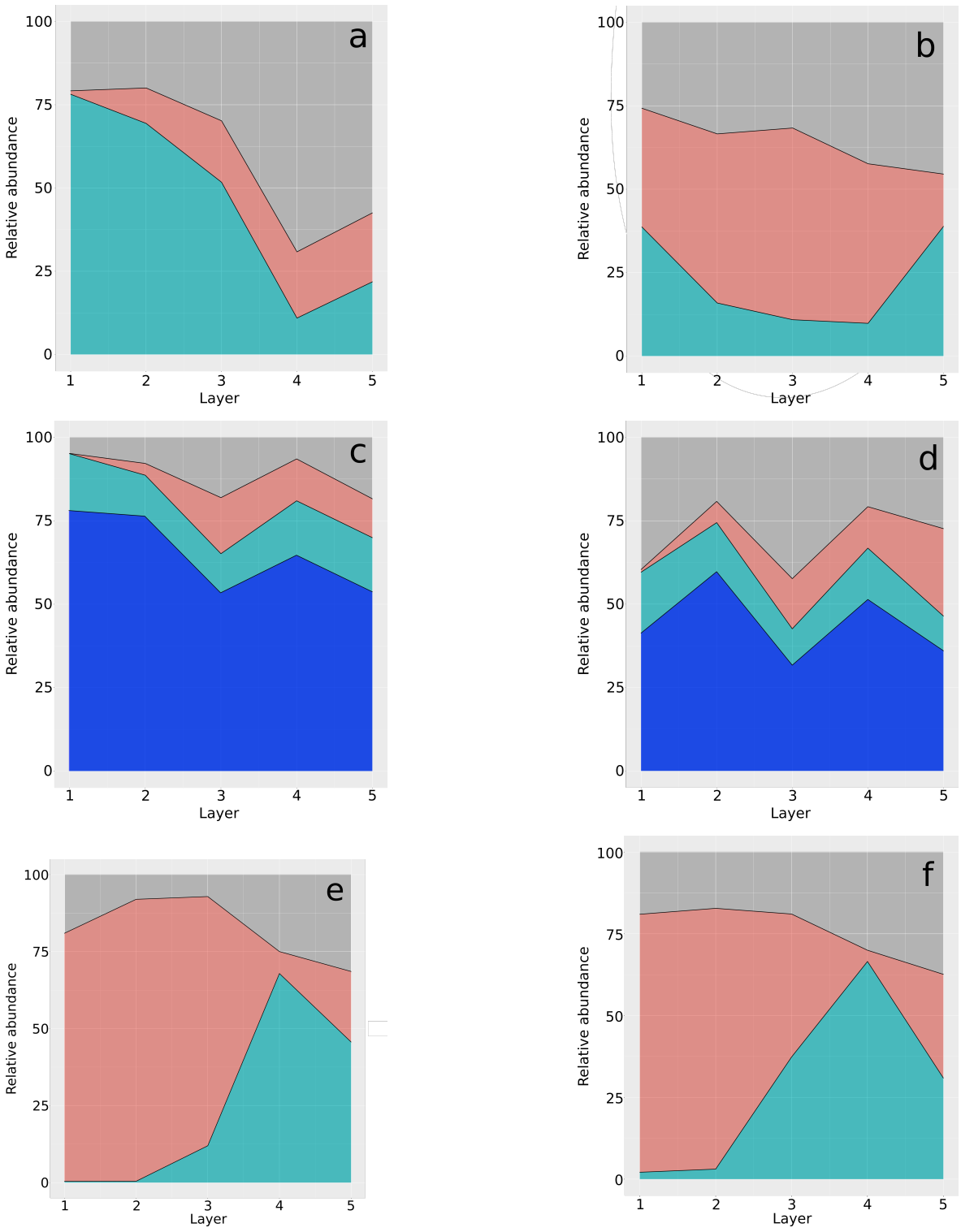

Figure S1: Stacked abundance of Alphaproteobacteria (a, b), Bacteroidetes (c, d) and Chloroflexi (e, f) in the low salinity (a, c, e) and high salinity (b, d, f) mats based on metagenomic reads abundance. Top: *Rhodovibrio* (blue), Rhodobacteraceae (red) and other Alphaproteobacteria (grey). Middle: *Salinibacter* (dark blue), Rhodothermaceae (light blue), Marinilabiliaceae (red) and other Bacteroidetes (grey). Bottom: SBR1031 (blue), *Ca.* Chlorothrix (red) and other Chloroflexi (grey).

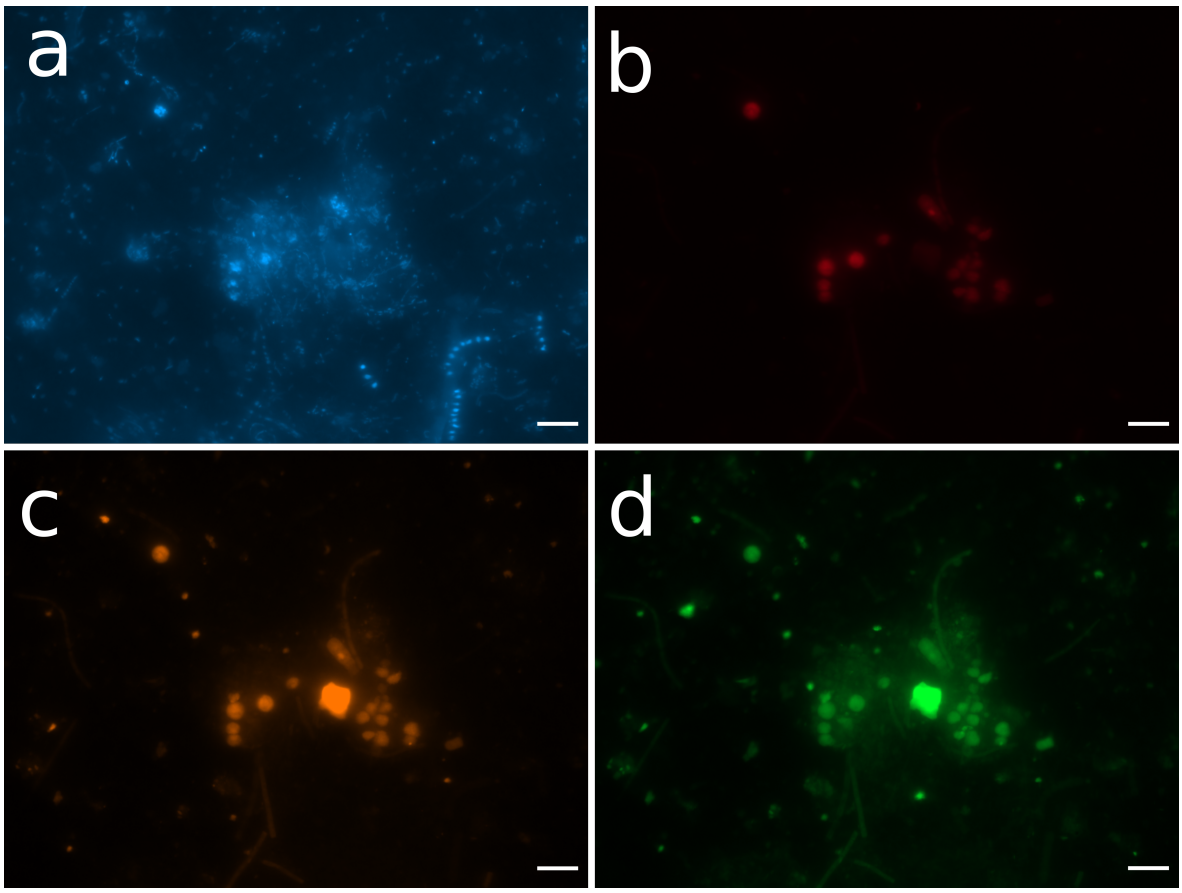

Figure S2: Epifluorescence microscopy images of hyper-saline mat material stained with DAPI. 405 nm (blue, DAPI) (a), 660 nm (far red) (b), 550 nm (orange) (c) and 480 nm (green) (d) channels. Scale bars of 10 μm