UNVEILING THE WORLDWIDE UV-METARESISTOME: METAGENOMIC ANALYSIS OF MICROBIOMES EXPOSED TO DIFFERENT UV-B REGIMES.

Daniel G. Alonso-Reyes[1] , Maria Eugenia Farias[1] and Virginia Helena Albarracín[1, 3]*

[1]Laboratorio de Microbiología Ultraestructural y Molecular, Centro de Investigaciones y Servicios de Microscopía Electrónica (CISME, CONICET, UNT) CCT, CONICET, Tucumán, Argentina.

[2]Laboratorio de Investigaciones Microbiológicas de Lagunas Andinas (LIMLA), Planta Piloto de Procesos Industriales y Microbiológicos (PROIMI), CCT, CONICET, Tucumán, Argentina.

[3]Facultad de Ciencias Naturales e Instituto Miguel Lillo, Universidad Nacional de Tucumán, Tucumán, Argentina.

Running headline: UV-METARESISTOME OF MICROBIOMES UNDER DIFFERENT UV-B REGIMES

Keywords: UV, MICROBIOMES, UV-METARESISTOME, METAGENOMICS, EXTREMOPHILES

*Corresponding author:

Virginia Helena Albarracín, Centro de Investigaciones y Servicios de Microscopía Electrónica (CISME, CONICET, UNT) Camino de Sirga s/n. Finca El Manantial, Yerba Buena (4107). Tucumán, Argentina.

ABSTRACT

During evolution, microorganisms exposed to high UV doses have developed a complex battery of physiological and molecular mechanisms to cope with UV stress and damage recently called as the UV-resistome. As a precedent, the UV-resistome was analyzed at the genomic and proteomic level in poly-extremophiles bacteria isolated from High-Altitude Andean Lakes. In this work, we go further by exploring the impact of UV-B radiation on microbiomes in different geographic regions across the globe. The abundance of photoprotection and DNA repair pathways in each microbiome was use to configure the world-wide UV-metaresistome. Metagenomics combined with georeferenced climate data indicated that the higher the UV-B dose suffered by the microbiome the higher the abundance and diversity of the UV-metaresistome genes. In contrast, a substantial depletion on microbial diversity was observed in higher irradiated environments. A positive correlation between CPF abundance and radiation intensity/photoperiod was detected. CPF genes were investigated in the global dataset, being present in most abundant organisms of communities that are facing significant exposure to sunlight. Three extra clades not identified so far and a widespread occurrence of cryptochromes-related genes were reported.

INTRODUCTION

Radiation, although essential for the existence of complex life, can become adverse when its natural dose is altered. In the past decades, a rise of biologically harmful UV radiation (UV) at the Earth´s surface was evident together with its detrimental effect on all life forms. This rising was a natural consequence of the drastic reduction of stratospheric ozone caused by increased concentrations of chlorofluorocarbons (CFCs) (Aucamp, 2007) and other halon gases in the upper atmosphere (Russell III *et al.*, 1996).

The amount of UV radiation reaching the ground comprises only a small proportion of global radiation, about 6–7 % of UV-A (320–400 nm) and less than 1.0 % of UV-B (280–315 nm) (Hu *et al.*, 2008). Biological damage is wavelength dependent: UV-A causes only indirect damage to DNA, proteins, and lipids through reactive oxygen intermediates. On the other hand, UV-B and UV-C (100 to 280 nm) cause both indirect and direct damage to DNA because of the strong absorption at wavelengths below 320 nm by the DNA molecule (Mitchell and Karentz, 1993). Thus, UV-A/UV-B are biologically relevant as their actual effects on living organisms can be of ecological impact. Numerous studies reported the UV-B effects on plants (Searles *et al.*, 2001; Xiong and Day, 2001; Robinson *et al.*, 2005; Ruhland *et al.*, 2005; Yan *et al.*, 2012), animals (Robson *et al.*, 2005; Bao *et al.*, 2014) and microorganisms (Zaller *et al.*, 2002; Avery *et al.*, 2004; Rinnan *et al.*, 2005; Piccini *et al.*, 2009). Increased UV-B is known to alter photosynthetic capacity and reduce biomass production in plants (Xiong and Day, 2001; Jansen *et al.*, 2010). In marine environments, UV-B drives the dynamics of larval krill, affects the coral–algal symbiosis and alters bacterial assemblages impacting higher trophic levels (Häder *et al.*, 2011). UV-B has the potential to cause negative effects on bacterioplankton since bacteria possess simple

haploid genomes and their small size precludes effective cellular shading and reduces the benefits of protective pigmentation (Garcia-Pichel, 1994). These organisms play a central role in the cycling of nutrients as bacteria may account for up to 90% of the cellular DNA in aquatic environments. In fact, they constitute a fundamental block chain in carbon flow so any influence of UV on bacteria in the environment will impact the rest of the ecosystem food chain (Karentz, 1995; Joux *et al.*, 1999). At soil level, UV-B influences decomposition processes by altering microbial diversity (Ballaré *et al.*, 2011), thus affecting crops.

During evolution, microorganisms exposed to high UV doses have developed a range of strategies to cope with UV stress and damage. Recently, it was pointed out the need to understand the integral UV-response as a complex battery of physiological and molecular mechanisms known as UV-resistome (Kurth *et al.*, 2015; Portero *et al.*, 2019). Thus, the UV-resistome depends on the expression of a diverse set of genes devoted to evade or repair the damage provoked direct or indirectly. Ideally, it encompasses the following subsystems: (1) UV avoidance and protection strategies; (2) UV or other stress sensors with their corresponding response regulators; (3) Damage tolerance and oxidative stress response; and (4) DNA damage repair. As a precedent, the UV-resistome was analyzed at the genomic and proteomic level in a poly-extremophile *Acinetobacter sp.* Ver3, a gamaproteobacterium isolated from High-Altitude Andean Lakes (HAAL). The genes from most of the described subsystems were identified and supported the outstanding UV-resistance profile observed for this strain (Kurth *et al.*, 2015). In a follow-up work, the UV-B resistome subsystems on avoidance, protection strategies and DNA damage repair

were explored for Ver 3 as well in HAAL Gram positive bacteria: *Exiguobacterium sp.* S17 and *Nesterenkonia sp.* Act20 (Portero *et al.*, 2019).

An example of UV avoidance and protection strategies can be found in motile biomass producers and consumers, which uses vertical migration in aquatic ecosystems to avoid excessive radiation. Sessile (attached) organisms rely on selection of habitat to limit solar exposure. On the other hand, different taxonomic groups have developed a number of photoprotective substances such as melanins, mycosporines, mycosporine-like amino acids (MAAs), scytonemin, carotenoids, phycobiliproteins, ectoine and several other UVR absorbing compounds of yet unknown chemical structure (Häder *et al.*, 2011). In most cases, screening pigments are not effective in bacteria, because of their small size, which would require the concentration of these screening compounds to be excessively high to provide sufficient absorption. In *Exiguobacterium sp.* S17 and *Nesterenkonia sp.* Act20, pigment extraction indicated the presence of carotenoid-like compounds in cells suggesting an antioxidative defense or protective role for them (Portero *et al.*, 2019).

A great number of specific and highly conserved DNA repair mechanisms has been developed against DNA damage, as there are photoreactivation (PR), excision repair (NER and BER), mismatch repair (MMR) and homologous repair (HR). In addition, damage tolerance (dimer bypass), SOS (save our soul) response, checkpoint activation, and programmed cell death (PCD) or apoptosis efficiently act against DNA lesions ensuring the genomic integrity.

Of particular importance is PR, executed by photoreactivating enzymes known as "photolyases". Photolyases are monomeric proteins of 53–66 kDa that contain flavin adenine dinucleotide (FAD) as cofactors and antenna pigments such as deazaflavin or methenyltetrahydrofolate derivatives (S Weber, 2005). Photolyases are considered among the earliest solutions of nature to the threat of DNA damage by high UV-irradiation intensity (Sancar, 2000, 2003). These enzymes target the most abundant products formed by UV-B, which are cyclobutane pyrimidine dimers (CPD). Such CPD lesions bring polymerases to a standstill, eventually leading to cell death. These enzymes bind tightly to CPDs in the dark and can be activated by different wavelengths, such as UV-A and photosynthetic active radiation (400 to 700 nm). The activation causes the splitting of the two C–C bonds in the CPD unit and results in the re-formation of the two separate pyrimidine bases (S Weber, 2005). Photolyases together with the structurally closely related cryptochromes (Cry) form a divergent family of photoactive proteins present in all three biological domains of life, called as cryptochrome/photolyase family (CPF). However Cryptochromes (Cry) have no photolyase activity and function as signaling molecules regulating diverse biological responses such as entrainment of circadian rhythms in plants and animals (Roenneberg and Merrow, 2005; Harmer, 2009). In all HAAL models for UV-resistome a nearly full set of family members was founded, including CPD photolyases supplemented by iron–sulfur bacterial cryptochromes and photolyases (FeS-BCP), plus a chryptochrome DASH in *Exiguobacterium* sp. S17 (Albarracín *et al.*, 2012, 2014; Kurth *et al.*, 2015; Portero *et al.*, 2019).

Investigations at metagenome-level about the effects of sunlight on microbial communities are mostly confined to photoreceptors such as microbial rhodopsins

(Pushkarev and Béjà, 2016; Pushkarev *et al.*, 2018) and LOV-domains (Pathak *et al.*, 2012). Singh et al. even reported abundances for several light-related genes in microbiomes from different environments (Singh *et al.*, 2009). However, no work investigated the proposed components of the so-called UV-resistome on microbiomes coming from a broad UV intensity range. In this work, we aim to characterize the occurrence and diversity of molecular components from two sub-systems of the UV-resistome -pigment protection and DNA repair-. Using a metagenomic approach, we studied microbiomes exposed to different intensities of UV-B and photoperiods. This contribution will aid to the understanding of the effects that climate change and loss of atmospheric ozone are able to cause on microbial communities in the short-term. In addition, we focus this work on microbial communities from extreme environments, models for early life sciences and astrobiology, and important sources for novel extremoenzymes/extremolytes with, otherwise, offer promising biotechnological applications.

RESULTS

Comparison of UV-B intensity profiles between worldwide microbiomes. An essential aspect of our analysis was the linkage of metagenomic data with UV-B georeferenced values obtained from glUV datasets. The maps processed by QGIS (Fig. 1) indicated the existence of three groups according the UV-exposure regimes: High (UV_{High}), Mid (UV_{Mid}) and low-exposed (UV_{Low}). Belonging to the first group, Lake Diamante red biofilm (DM), Socompa stromatolites (SS) and Tibetan Plateau sediment (TB) were

exposed to UV-B intensities of 9677, 9536 and 8885 J/m^2/day respectively. In contrast, UV_{Low} microbiomes from Olkiluoto Island groundwater (OK), Lake Montjoie (MT) and Greenland cryoconite (CR) were linked to low UV-B intensities of 77, 719 and 1759 J/m^2/day respectively. In between, the microbiomes of the Amazon River (AM), Lake Rauer (RA) and Dewar Creek hots spring (HT), with an intensity of 5630, 2289 and 2281 J/m^2/day grouped together.

The above intensity values were the expected considering the expedition dates; the samples of the UV_{High} group were taken during summer, when the insolation is maximum on the high altitude environments of the Argentinean Puna and the Tibetan Plateau (February and August) and the incidence of UV-B in both regions become the highest of the planet. On the other hand, the UV_{Low} samples were taken during August, November and February which correspond to the autumn and winter seasons in the northern hemisphere, darker than their equals in the southern hemisphere. In turn, the UV_{Mid} comprised the AM microbiome belonging to a tropical environment, while RA and HT correspond to extreme environments (cold and hot, respectively) at high latitudes.

An interesting fact not quite evident from the UV-datasets is revealed by the calculated day-length for each georeferenced metagenome (https://www.suncalc.org). Two microbiomes coincided with extreme photoperiods; the OK microbiome with the shorter photoperiod in the study, which was only 7.8 hours, and RA which was sampled when Antarctica went through its longest period of sunlight in the year, a complete day under the sun. Important to remark is that we assumed that the environments were exposed to similar conditions of indicated solar irradiance for a considerable wide period of time already before sampling. In consequence, the diversity of genes and species will be a reflection of the ecological pressure of radiation on the environment (among other factors) as the

geographic conditions (latitude, altitude, and orography) did not change considerably in those sampled regions for decades. In addition, the intensity of solar irradiation over each microbiome will also depend on their *on-site* spatial disposition being maximum for samples taken from soils or shallow water of lakes, springs, oceans and streams but much lower in sediments, groundwater or deep water.

Using the solar irradiation data, we assessed the effect of UV-B in microbiomes exposed to different intensity/time of insolation. The response variables evaluated were 1) microbial diversity, 2) content and diversity of CPF genes 3) abundance and diversity of other DNA repair mechanisms, such as NER, BER, MMR and Homologous Repair and 4) abundance of genes related to photoprotective pigments such as carotenoids and ectoine.

Microbial diversity. Microbial diversity present in each microbiome was evaluated using species richness and Shannon's diversity index parameters. Interestingly, both indexes decreased as the UV-B irradiation on the microbiome´s environment increases (Fig. 2). In microbiomes exposed to intensities below 4000 J/m^2/day, species richness oscillates within a wide range (11-103) while SS microbiome with intensities above 8000 J/m2/day only had 21 species. Shannon index indicated poor diversity on the three microbiomes of the UV_{High} group. All UV_{High} samples and HT (UV_{Mid} sample) presented the lower indices of the whole study.

Quantitative analysis of the UV-metaresistome. In order to evaluate the abundance of the UV resistance genes in the complete set of microbiomes, a database of such genes was built using Uniprot sequences linked to KEGG orthology numbers, which allowed us to assemble the metabolic pathways. The resulting abundances for each gene and metabolic pathway were analyzed (Table S1-S7) together with the abundance for each DNA repair/photoprotection strategy (Fig. 3).

Microbial communities of the UV_{High} group, TB and SS, reported the largest number of functions together with the highest abundance of the whole set of functions (UV-metaresistome). The case of DM microbiome is interesting as despite being the most exposed environment, its microbiome presented the lowest diversity of resistance strategies. However, this result may be due to technical and taxonomical biases. The archaeon *Halorubrum* sp., which constitutes the dominant genus detected by the taxonomic analysis performed on DM microbiome (Fig. 7), lack the type of polymerase proposed by KEGG in NER and BER pathways. KEGG prokaryote models for NER and BER establish DNA polymerase I (K02335) as the responsible of the final gap filling, but this enzyme seems not to be present in the genus, which probably uses another. This fact is evidenced by a gene (K02335) abundance of one or two orders of magnitude lower in DM compared with the rest of microbiomes (Table S1 and S3). Moreover, a rapid blast of K02335 linked sequences of Swiss-Prot against *Halorubrum sp.* genomes from the NCBI does not show any significant alignment. As the NER and BER pathways require DNA polymerase I (K02335) to be considered existent in our metagenomic analysis, they are not visible in the DM stacked bar. In the case of Homologous Recombination pathway, we also used as a reference the prokaryotic set of genes proposed by KEGG, while is known that Archaea uses a different one (Seitz *et al.*, 1998; Rzechorzek *et al.*, 2014). Finally, It was already stated that MMR is absent in phylum Euryarchaeota (Anderson *et al.*, 2009; Grasso and Tell, 2014) which includes *Halorubrum sp*.

SS was the community that had the whole set of molecular resistance pathways, even for those resistance mechanisms poorly represented or null in others. It is noteworthy the abundance of Ectoine Biosynthesis (180 e^{-02} RPKG) representing 20.29% of the SS metaresistome. This function has only a limited presence in RA (17 e^{-02} RPKG) with a

contribution of 2.44%. Ectoine is a compound synthetized by bacteria in response to high salinity, which helps on the hydration of proteins and cell membranes, while used as an active component of some sunscreens. Its presence in SS is consistent with the hypersaline nature of the microbiome environment. The significant presence of mismatch repair (MMR) were also noticed in the SS microbiome, with an abundance of 47 e^{-02} RPKG and a representativeness of 5.37% of the metaresistome. This mechanism of DNA repair has little or no presence in the rest of the microbiomes.

The distribution of CPF genes in the three-irradiation groups followed an ecological significant pattern: they were flagrant in microbiomes with UV_{High} and UV_{Mid}, while insignificant or null in communities with UV_{Low} or null exposure -OK and GU respectively. In DM CPF genes became more than relevant. Their abundances were 249 e^{-02} RPKG, which were greater than in the rest of the microbiomes. Furthermore, they contribution to the DM metaresistome was 68.40%, the highest of the whole study followed by RA with 34.14%.

The relationship between abundance of CPF and intensity/time of insolation was studied (Figure 4.a) showing an upward trend of CPF abundance as UV-B irradiation increases. Particularly, RA, an UV_{Mid} microbiome, completely desviate from the trend due to an outstanding abundance of this class of genes. We had previously mentioned that RA has the longest photoperiod of the study, so we set out to verify if there is a relationship between the abundance of CPF and the photoperiod. Figure 4.b shows that there is upward trend of CPF as the photoperiod increases, finding RA quite in line with the trend. Thus, both factors, the intensity and the photoperiod could be influencing the abundance of CPF in the microbial communities.

Assembled CPF genes Analysis. Alignment editing and subsequent phylogenetic analysis resulted in a tree with 214 CPF sequences and numerous clades (Fig. 5), which mostly correspond to subfamilies already studied in previous works: CPD photolyases classes I, II and III, DASH cryptochromes which include single strand photolyases (Selby and Sancar, 2006), FeS-BCPs group which consist of proteins having an iron-sulfur cluster either prokaryotic [6-4] photolyases or cryptochromes (Graf *et al.*, 2015) and thermostable CPD photolyases which have FMN as antenna cofactor and a thermostable nature (Ueda *et al.*, 2005), a group of unstable position in phylogeny (Portero *et al.*, 2019).

In addition, this work reports three extra clades not identified so far and called as unidentified I, II and III (UI, UII and UIII) with 41%, 88% and 96% of bootstrap support, respectively. In total, these groups represent 24.76% of the global sequences. Fig. 6 shows the distribution of each subfamily by microbiome. Interestingly, the clades UI and UII have most of their components coming from a single microbiome; 85.7% of UI is constituted by sequences from the RA microbiome while 87.5% of UII corresponds to MT sequences. It is possible to observe that UI subfamily is the unique in DM and is dominant in RA, both communities with the highest peak of CPF abundance. Furthermore, UI and UII clades have a homogeneous taxonomic consistency, with their sequences being classified as Halobacteria and Actinobacteria, respectively. On the other hand, UIII seems to be a clade formed by sequences of different taxonomic origins, including Proteobacteria, Bateroidetes and Verrucomicrobia phylum.

It is clear the widespread presence of cryptochrome-holding subfamilies cry-DASH and Fes-BCPs among the microbiomes (Fig. 6). Although the relative abundance of Fes-BCPs clade remained similar among the microbiomes, the abundance of cry-DASH

followed a pattern of affinity for light, being abundant in SS and TB communities (UVB_{High}), AM (UVB_{Mid}) and RA (longest photoperiod).

Taxonomic identification of each gene was carried out using the BLAST algorithm. CPF sequences were classified by genus and paired with the same information previously obtained for each microbiome through MetaPhlAn. The CPF sequences that were grouped together with their homologue with an identity lower than 70% by BLAST were clustered as an unclassified category. Both, the genus diversity from the whole microbiome (top bar) and the set of CPF sequences recovered from that microbiome (lower bar), are shown in Figure 7.The classification rate of the CPF genes was generally high (>60%) in the samples DM, SS, TB, AM, RA and GU but lower in the rest. The lowest percentage of sequences was registered at MT, with barely 23% of the sequences being assigned to some genus. This may be because the community itself has a poorly referenced diversity (top bar). Despite this, MT had the greatest diversity of genus (12) assigned to CPF along with SS and RA (11 each).

By pairing the information of both bars we evaluate the relevance of CPF in each community. The most abundant taxa in DM, SS, RA and HT communities possess CPF genes. In AM, only *Ralstonia* and *Vibrio* display CPF; both genuses together represent 35% of the community. In the case of CR, only *Hymenobacter* and *Agromyces* possess these genes, adding both a representativeness of 7%. *Pedobacter* was the unique taxon in TB with CPF genes and it represents less than 4% of the community. Finally, *Escherichia* which has an abundance of just 0.44% was the unique CPF contributor in GU. Neither in MT nor OK found matches between the MetaPhlAn and BLAST classifications.

DISCUSSION

UV-B is a recognized driver of ecological processes regulating numerous biological patterns and processes. The availability of a new and specific set of global UV-B surfaces (glUV) containing monthly mean UV-B data and six derived UV-B surfaces allow us to study the effect of UV exposure on selected environments worldwide. This is the first work in which UV-B was considered as an ecological variable in a sequenced-based metagenomic study of microbial communities.

It is known that UV-B negatively affects microbial diversity (Ballaré *et al.*, 2011) which concur with our findings in this work. Species richness (Fig. 2A) was quite variable at low radiation intensities probably due to the different environmental conditions that can limit or promote species diversity in each community before UV-B intensity reaches a critical role. Once a certain threshold has been exceeded -above 6000™ J/m^2/day- when UV-B becomes a limiting factor, the number of species decreases due to selective pressure on those that do not possess efficient molecular mechanisms to defend themselves from UV-B, being unable to adapt to the new assemblage of species. A similar situation was observed when applying the Shannon index (Fig. 2B) which incorporates equitability of the species abundances in addition to its ability of detecting rare species. Shannon index, although with variable values in UV_{Low} microbiomes, inevitably decreases in UV_{High} ones. This suggests that the microbes with full capacity to defend themselves against UV-B rays become dominant, relegating others to a less substantial place in the community.

By performing quantitative metagenomics we evaluated diversity and abundance of each DNA repair/photoprotection pathway in all the communities (Fig. 3). SS proved to be the most diverse as contains the complete set of pathways, even with a nearly equitable

distribution of its abundances. The SS microbiome corresponds to a non-lithified modern stromatolites growing at the shore of the remote volcanic lake Socompa at 3570 m a.s.l. in the Puna (Argentinean Andes) (Farías *et al.*, 2011) on a desert area that withstand the most elevated doses of global solar radiation on Earth (Piacentini *et al.*, 2003; Albarracín *et al.*, 2015). These complex communities of stratified microbial diversity along physico-chemical gradients (Toneatti *et al.*, 2017) developed under the pressure of extreme environmental factors similar to the ones present in Early Earth's atmosphere. The bacteria isolated from these stromatolites are poly-extremophiles able to resist severe stress conditions, including UV-B, heavy metals, salinity and arsenic.

The SS microbiome, which is the most complex and diverse of those belonging to UV_{High} group, it is also complex and diverse in its way of defending itself against UV-B. SS maintains a relatively high percentage of gene copies dedicated to the synthesis of pigments such as carotenoids and ectoine. Ectoine is an amino acid derivative produced mainly by aerobic, chemoheterotrophic and halophilic bacteria in response to osmotic stress and unfavorable environmental conditions. Ectoine is also known for its practical use for enzyme stabilization, human skin protection, anti-inflammatory treatment and inhibitory effects in neurodegenerative diseases (Bownik and St□pniewska, 2016). It stabilizes cell membranes, enzymes, and nucleic acids at extreme temperatures or higher salt concentrations in some bacteria (Smiatek *et al.*, 2012). Furthermore it absorbs UV-A radiation, preserves DNA from breaking down in various cell types and reduce oxidative stress caused by UV (Buenger and Driller, 2004; Botta *et al.*, 2008; Sajjad *et al.*, 2018). It is remarkable that SS were singular in regard to his significant presence of ectoine synthase, since this fact is compatible with the stressful environment in which it was sampled,

including high salt concentration in the close hypersaline Socompa lake (17%) and high exposure to UV irradiation leading to high osmotic and oxidative stress respectively. Further study of how ectoine contributes to the functioning of these microbial communities in extreme environments could shed light on new biotechnological applications.

The significant abundance of MMR in SS can be related to several factors. MMR proteins provide DNA stability under the stress of environmental fluctuations, which in the case of Lake Socompa may be temperature, arsenic, hypersalinity and UV. It has become clear that the major process influencing evolution in DNA is MMR (Džidić *et al.*, 2003). It controls mutation rates and interspecies recombination. Inhibition of MMR allows in-vivo recombination between diverged DNA fragments and rises mutation rates (10^2 to 10^3 fold) creating *de novo* variation, which is the genesis of new biodiversity and novel biosynthetic compounds. Strains will easily acquire beneficial mutations by reacquiring MMR wild type alleles in horizontal gene transfer through their hyper-recombination phenotype. The abundance of this pathway by the SS community may be beneficial as it can provide high mutation-rates in these communities through cycles of lose-and-reacquire of the *mut* genes, otherwise needed to rapidly evolve and adapt under fluctuating ecological circumstances.

TB, DM and RA also show signs of genomic adaptation to radiation. TB was the third most exposed microbiome which showed a high abundance of gene copies assigned to 6 of the 7 pathways examined, DM is a different case, since in spite of being the first in radiation intensity, it has the lowest number of defense strategies. However, this result may be due to taxonomical and technical biases. As mentioned above, the community is dominated by the archaeon *Halorubrum* sp., which uses a different polymerase from that proposed by KEGG for the prokaryotic NER and BER metabolic pathways. This made both

pathways invisible in our search, although it is very likely that both exist because they are important DNA repair systems in Archaea (Grasso and Tell, 2014; Grogan, 2015). Additionally, HR pathway was not detected because the respective archaeal genes were not used. On the other hand, MMR is a type of DNA repair that has not been reported for the Euryarchaeota phylum, including *Halorubrum sp.* We believe that a same analysis we performed in this work, made with the right genes, should unveil an abundant and diverse UV-metaresistome in the DM case compared with other archaeal-biased metagenomes of lower insolation.

The relevance of CPF in the ecology of microbial communities becomes more evident when its tendency along UV-B intensity gradients and photoperiod is studied (Fig. 4). Our study reveals a rising trend of CPF abundance in microbial communities as their environments receive greater radiation or day duration extends. This is likely due to an increase in DNA damage caused by UV-B, which are mainly pyrimidine dimers. In such a situation, populations could increase the expression of CPF by increasing the gene copy number. In addition, those species of the community, which lack CPF in their genome, would be replaced by species that contain these genes. In either case, the overall increase in CPF copy abundance would indicate that the community improves its defense capabilities against UVB using the highly efficient mechanism of photoreactivation or modulating enzymatic mechanisms triggered by light sensing by cryptochromes.

The CPF group of proteins comprises mostly genes of photolyases with different specializations, and cryptochromes whose functions are largely unknown. The family has been divided into different subgroups considering phylogeny, kind of chromophore, specialization, host organism, structure, etc. After assembling the reads and aligning the

sequences corresponding to CPF, we were able to classify our sequences in the main subgroups discovered to date and we also detected the presence of other unknown subgroups (Fig. 5). It is interesting that the sum of the sequences belonging to these last unknown groups comprise approximately a quarter of the total, thus offering us a large pool of candidates for new functions, specializations or molecular specificities.

As already mentioned, the unique photolyase present in DM belongs to clade UI (Fig. 6), apparently novel and also present in RA. We already know that DM and RA are the samples with the highest radiation intensity and the highest photoperiod respectively. Apparently, *Halorubrum sp.*, the sole organism detected by our analysis in DM, could cope with UV-B mainly using this photoreceptor without any other complement from the CPF family. We propose this gene as a candidate to further studies of photoreactivation, characterization and molecular structure analysis in order to shed light on this portion of the phylogenetic tree of CPF which has not been studied yet.

It is a fact the presence of the Cry-DASH class in SS, TB, AM and RA while of lower relative abundance or absent in other samples of less exposure to light. It has been reported in previous works that these cryptochromes are in fact photolyases with affinity for single-stranded DNA and in some cases RNA (Selby and Sancar, 2006). It is possible that Cry-DASH plays a complementary role of standard photolyases with affinity for double-stranded DNA, contributing to the global increase of photoreactivation in these microbiomes. The FeS-BCPs group had a remarkable behavior, being present in almost all the communities. Previously, it has been suggested that this class complements other photolyases by performing the function of a [6-4] photolyase (Zhang *et al.*, 2013; Graf *et al.*, 2015), avoiding or decreasing in this way the use of the inefficient NER system. Our

work gives more support to this claim by showing its ubiquitous and abundant presence in microbial communities.

Another interesting fact is that CPF and thus photoreactivation are usually present in the most abundant organisms of the UV_{High} microbiomes (Fig. 7). In addition, SS and RA show a high number of genus in which photoreactivation is present. This suggest that photoreactivation may act as a successful system for assuring survival and predominance of taxa in UV-stressed environments. This fact contrasts with those less insolated samples, where natural selection does not act in favor of photoreactivation promoting neither abundance nor diversity of individuals carrying these genes.

CONCLUSIONS

This work showed the correlation of high UV exposure of a given microbiome with its low microbial diversity and high specific UV-resistance molecular components in a world-wide study. Thus, a major expression of certain components of the UV-resistome was observed in response to higher radiation intensities. A downward trend in microbial diversity and an upward trend of CPF abundance were consistently observed as UV-B irradiation increases. Likewise, cryptochrome-like genes were found abundant in most exposed microbiomes indicating a complementary role to standard photolyases. Also, we observed that CPF are more likely present in most abundant organisms of the UV_{High} microbiomes, suggesting an evolutionary important force for survival and dominance in highly irradiated environments. In accordance, this work reported three novel CPF clades not identified in previous analyses.

Finally, metagenomics proved to be an excellent tool useful to reveal an important correlation of microbiomes UV-exposure with diversifying resistance mechanisms and the increase of their gene copies gene diversity. Additional methods such as metatranscriptomics and metaproteomics should be implemented in order to unveil the molecular dynamics of the proposed UV- metaresistomes.

ACKNOWLEDGEMENTS

VHA and MEF are staff researchers from the National Research Council (CONICET) in Argentina. DA is recipients of a doctoral fellowship from CONICET. The authors have produced this manuscript in spite of the delays of funding execution from National Agencies in Argentina, mainly FONCyT (PICT 2013 2991) and CONICET (PIP 2015 0519, PIO-UNCA y PICT V 3825-2016), and the drastic devaluation of Argentinean currency which began in 2016.

EXPERIMENTAL PROCEDURES

UVB data processing and metagenome selection. Monthly mean UVB glUV datasets (Beckmann *et al.*, 2014) were utilized as guides to choose samples using sample information. The glUV datasets from the Helmholtz Centre for Environmental Research are daily measurements summarized into monthly mean UVB Erythemal Daily Dose values and averaged across all years in the period of 2004-2013. These layers cover global land and marine areas and have a spatial resolution of 15 arc-minutes. They are in the latitude/longitude coordinate reference system and the datum is WGS84[26].

The datasets were processed using QGIS (www.qgis.org) and worldwide colored maps were generated by month. The maps were employed to select potential samples to be analyzed from different world places according to a visual criteria of irradiance. The sample information provided by the database ("biosample" linked section of each SRA entry), specifically collect date and geographic location, becomes crucial here as the monthly mean irradiation value for each microbial community in the time when it was sampled can be known.

The metagenomes selected from NCBI using biosample information were the following, each one with its respective SRA entry: Lake Diamante (ERR1824222) (Rascovan *et al.*, 2016), Socompa stromatolite (SRR3341855) (Kurth *et al.*, 2017), Tibetan Plateau sediment (SRR3322106) (Chen *et al.*, 2016), Amazon River (SRR1790676) (Satinsky *et al.*, 2015), Lake Rauer (SRR6129205) (Tschitschko *et al.*, 2018), Dewar Creek hot spring (SRR5580900), Greenland cryoconite (SRR5275901) (Hauptmann *et al.*, 2017), Lake Montjoie (SRR5818193) (Tran *et al.*, 2018), Olkiluoto Island groundwater (SRR6976411) and human gut (SRR6517782) (Ye *et al.*, 2018). The metagenome of the human gut is assumed as UVB free environment, acting as a negative control.

General overview of the metagenomes. DNA from all microbial communities was obtained and sequenced by shotgun strategy through Illumina technology except DM, which was sequenced with 454 GS FLX Titanium instrument. Quality filtering and merging yielded a range between 0.64 and 13.66 Gb for further analysis (Table 1). Only TB and AM datasets reported high percentage of merging with FLASH (Magoč and Salzberg, 2011), however the low-merged dataset from OK was also used for downstream steps as it ended with a reasonable size (between the range mentioned above). MEGAHIT (Liu *et al.*, 2015)

was used for assembly since it can deal with large and complex datasets in a time- and cost-efficient manner. Protein prediction with Prodigal (Hyatt *et al.*, 2010) over contigs with >999 bp outputted a rate of 1000-1500 proteins / Mb, which is congruent with the high coding density expected for microbial species.

Quality control and assembly. The pre-processing and assembly of the metagenomes one by one implied the usage of several bioinformatic tools. Adapters were removed from Illumina raw reads using fastq-mcf tool of ea-utils v1.04.676 (Aronesty, 2011). This step was not needed for Lake Diamante metagenome, as it was sequenced with 454 technology. The quality filtering and trimming was performed with the same program using parameters l>50 and q>20. The program kneaddata v0.6.1 (https://bitbucket.org/biobakery/kneaddata/wiki/Home) was used with the --bypass-trim option to clean contaminant sequences from human. Pair end reads were merged with FLASH v1.2.11 (Magoč and Salzberg, 2011) in order to recover unpaired longer reads. The pair end reads with low percentage of merging were leaved in paired state for assembly. Assembly of filtered reads was performed with MEGAHIT v1.1.2 (Liu *et al.*, 2015). Assembled contigs were annotated with Prodigal v2.6.2 (Hyatt *et al.*, 2010) which outputted translated protein sequences.

To perform quantitative metagenomics using the reads directly, the pair end reads that remained with low percentage of merging were concatenated together by the forward and reverse using a script available at the GitHub (https://github.com/LangilleLab/microbiome_helper/blob/master/concat_paired_end.pl).

Reference database building. Considering the main functional orthologues of the UV-resistome subsystems main components, a reference protein database was built. KEGG Orthology number corresponding to our pathway of interest was linked with the UniProt database, and then filtered results by EC number, known molecular function or biological process, uniref 90 clusters, and sequence length >100 subsequently. The single, concatenated and merged reads were aligned to the aforementioned databases with PALADIN which outputted read counts. The counts were normalized to Reads per Kilobase per Genome using the following formula: (Read counts÷Number of Genomes)÷Gene Length.

The protein families and pathways chosen fall in two UV-resistome subsystems: DNA repair and photoprotection. At the DNA repair level these were photolyase/chryptochrome family (K06876, K01669), base excision repair (K01247, K03649, K10563, K10773, K10800, K01246, K03648, K01151, K01142, K02335, K01972), nucleotide excision repair (K03701, K03702, K03703, K02335, K01972), mismatch repair (K03555, K03572, K03573, K01141, K10857, K03601, K07462, K02337, K01972), homologous recombination repair (K03629, K03584, K03553, K02337, K02335, K03550, K03551, K01159), while at the photoprotection level we screened for carotenoid biosynthesis (K15745, K14595, K09847, K09844, K09845, K09846, K10027, K10208, K02292, K10209, K10210, K10211, K02294, K09836, K02294, K22502, K14746, K02291, K09835, K14605, K09839), and ectoine biosynthesis (K06720). Finally, the fasta headers of each file were labeled with the name of its respective gene and then files were merged in a single one.

Estimation of gene abundance, pathway abundance and diversity. To search and quantify the abundance of the genes in the different metagenomes, the preprocessed unpaired reads were aligned against the UV-resistant reference database created before. The protein alignment was performed with PALADIN v1.3.1 (Westbrook *et al.*, 2017) which outputs a table with alignment counts for each gene. Alignments were filtered by maximum quality = 60 and the abundances were normalized in reads (or counts) per kilobase per genome (RPKG) having estimated first the number of genomes for each metagenome through MicrobeCensus v1.1.0 (Nayfach and Pollard, 2015).

The abundance of the complete pathways was quantified in a following manner. First the pathways were divided in several component reactions or steps according to KEGG pathway maps and added the abundances of the genes that catalyze the same step with the porpoise of compute the abundance for each step separately. Then, was selected the step with the minimum value as the representative value for the whole pathway (Table S1-7). In the case of carotenoid biosynthesis pathway, was added all the abundances globally as each gene can produce a different variant of carotenoid.

Diversity analysis of each metagenome was performed with MetaPhlAn v2.7.7 (Truong *et al.*, 2015) with the --bt2_ps parameter set to 'sensitive'. MetaPhlAn profiles the composition of microbial communities from metagenomic shotgun sequencing data with species-level. It relies on ~1M unique clade-specific marker genes identified from ~17,000 reference genomes (~13,500 bacterial and archaeal, ~3,500 viral, and ~110 eukaryotic). Shannon diversity index was obtained with Vegan 2.5-2 R package.

Analysis of the annotated genes. The protein sequences were aligned using Diamond v0.9.22 software (Buchfink *et al.*, 2014) against the CPF genes of the reference database

built earlier. Alignment parameters were 50% identity, 70% query coverage and e < 10^{-5}. The retrieved files coming from each metagenome were modified at the sequence headers to hold the name of its respective metagenome. Next, all files were merged in a single one and filtered by sequence length > 400 residues.

The filtered sequences were used for a phylogenetic analysis. The phylogenetic tree was built with FastME v2.0 (Lefort *et al.*, 2015) using the Jones–Taylor–Thornton rate matrix (Jones *et al.*, 1992) with 1000 bootstrap replicates. A consensus of the 1000 resulting trees was selected for further processing and visualization using iTOL (Letunic and Bork, 2016). Additionally, a taxonomical identification of the sequences was performed through BLAST web server (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

REFERENCES

Albarracín, V., Simon, J., Pathak, G., … L.V.-P.&, and 2014, U. (2014) First characterisation of a CPD-class I photolyase from a UV-resistant extremophile isolated from High-Altitude Andean Lakes. *pubs.rsc.org*.

Albarracín, V.H., Kurth, D., Ordoñez, O.F., Belfiore, C., Luccini, E., Salum, G.M., et al. (2015) High-Up: A Remote Reservoir of Microbial Extremophiles in Central Andean Wetlands. *Front. Microbiol.* **6**: 1404.

Albarracín, V.H., Pathak, G.P., Douki, T., Cadet, J., Borsarelli, C.D., Gärtner, W., and Farias, M.E. (2012) Extremophilic Acinetobacter Strains from High-Altitude Lakes in Argentinean Puna: Remarkable UV-B Resistance and Efficient DNA Damage Repair. *Orig. Life Evol. Biosph.* **42**: 201–221.

Anderson, I., Ulrich, L.E., Lupa, B., Susanti, D., Porat, I., Hooper, S.D., et al. (2009) Genomic Characterization of Methanomicrobiales Reveals Three Classes of

Methanogens. *PLoS One* **4**: e5797.

Aronesty, E. (2011) ea-utils: Command-line tools for processing biological sequencing data.

Aucamp, P.J. (2007) Questions and answers about the effects of the depletion of the ozone layer on humans and the environment. *Photochem. Photobiol. Sci.* **6**: 319–330.

Avery, L., C. Thorpe, P., Thompson, K., Paul, N., Grime, J., and M. West, H. (2004) Physical disturbance of an upland grassland influences the impact of elevated UV-B radiation on metabolic profiles of below-ground micro-organisms.

Ballaré, C.L., Caldwell, M.M., Flint, S.D., Robinson, S.A., and Bornman, J.F. (2011) Effects of solar ultraviolet radiation on terrestrial ecosystems. Patterns, mechanisms, and interactions with climate change. *Photochem. Photobiol. Sci.* **10**: 226–241.

Bao, X., Li, Q., Hua, J., Zhao, T., and Liang, W. (2014) Interactive effects of elevated ozone and UV-B radiation on soil nematode diversity. *Ecotoxicology* **23**: 11–20.

Beckmann, M., Václavík, T., Manceur, A.M., Šprtová, L., von Wehrden, H., Welk, E., and Cord, A.F. (2014) glUV: a global UV-B radiation data set for macroecological studies. *Methods Ecol. Evol.* **5**: 372–383.

Botta, C., Di Giorgio, C., Sabatier, A.-S., and De Méo, M. (2008) Genotoxicity of visible light (400–800nm) and photoprotection assessment of ectoin, l-ergothioneine and mannitol and four sunscreens. *J. Photochem. Photobiol. B Biol.* **91**: 24–34.

Bownik, A. and St□pniewska, Z. (2016) Ectoine as a promising protective agent in humans and animals. *Arh. Hig. Rada Toksikol.* **67**: 260–265.

Buchfink, B., Xie, C., and Huson, D.H. (2014) Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**: 59.

Buenger, J. and Driller, H. (2004) Ectoin: An Effective Natural Substance to Prevent UVA-

Induced Premature Photoaging. *Skin Pharmacol. Physiol.* **17**: 232–237.

Chen, B., Yuan, K., Chen, X., Yang, Y., Zhang, T., Wang, Y., et al. (2016) Metagenomic Analysis Revealing Antibiotic Resistance Genes (ARGs) and Their Genetic Compartments in the Tibetan Environment. *Environ. Sci. Technol.* **50**: 6670–6679.

Džidić, S., Bačun-Duržina, V., and Petranović, M. (2003) The role of mismatch repair in bacterial evolution. *Food Technol. Biotechnol.* **41**: 177–182.

Farías, M.E., Contreras, M., Rasuk, M.C., Kurth, D., Flores, M.R., Poiré, D.G., et al. (2014) Characterization of bacterial diversity associated with microbial mats, gypsum evaporites and carbonate microbialites in thalassic wetlands: Tebenquiche and La Brava, Salar de Atacama, Chile. *Extremophiles* **18**: 311–329.

Farías, M.E., Poiré, D.G., Arrouy, M.J., and Albarracin, V.H. (2011) Modern Stromatolite Ecosystems at Alkaline and Hypersaline High-Altitude Lakes in the Argentinean Puna BT - STROMATOLITES: Interaction of Microbes with Sediments. In, Tewari,V. and Seckbach,J. (eds). Springer Netherlands, Dordrecht, pp. 427–441.

Farías, M.E., Rascovan, N., Toneatti, D.M., Albarracín, V.H., Flores, M.R., Poiré, D.G., et al. (2013) The Discovery of Stromatolites Developing at 3570 m above Sea Level in a High-Altitude Volcanic Lake Socompa, Argentinean Andes. *PLoS One* **8**:.

Garcia-Pichel, F. (1994) A model for internal self-shading in planktonic organisms and its implications for the usefulness of ultraviolet sunscreens. *Limnol. Oceanogr.* **39**: 1704–1717.

Graf, D., Wesslowski, J., Ma, H., Scheerer, P., Krauß, N., Oberpichler, I., et al. (2015) Key Amino Acids in the Bacterial (6-4) Photolyase PhrB from Agrobacterium fabrum. *PLoS One* **10**: e0140955.

Grasso, S. and Tell, G. (2014) Base excision repair in Archaea: Back to the future in DNA

repair. *DNA Repair (Amst).* **21**: 148–157.

Grogan, D.W. (2015) Understanding DNA Repair in Hyperthermophilic Archaea□: Persistent Gaps and Other Reasons to Focus on the Fork. *Archaea* **2015**:.

Häder, D.-P., Helbling, E.W., Williamson, C.E., and Worrest, R.C. (2011) Effects of UV radiation on aquatic ecosystems and interactions with climate change. *Photochem. Photobiol. Sci.* **10**: 242–260.

Harmer, S.L. (2009) The Circadian System in Higher Plants. *Annu. Rev. Plant Biol.* **60**: 357–377.

Hauptmann, A.L., Sicheritz-Pontén, T., Cameron, K.A., Bælum, J., Plichta, D.R., Dalgaard, M., and Stibal, M. (2017) Contamination of the Arctic reflected in microbial metagenomes from the Greenland ice sheet. *Environ. Res. Lett.* **12**:.

Hu, B., Wang, Y., and Liu, G. (2008) Influences of the clearness index on UV solar radiation for two locations in the Tibetan Plateau-Lhasa and Haibei. *Adv. Atmos. Sci.* **25**: 885–896.

Hyatt, D., Chen, G.-L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

Jansen, M.A.K., Martret, B. Le, and Koornneef, M. (2010) Variations in constitutive and inducible UV-B tolerance; dissecting photosystem II protection in Arabidopsis thaliana accessions. *Physiol. Plant.* **138**: 22–34.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**: 275–282.

Joux, F., Jeffrey, W.H., Lebaron, P., and Mitchell, D.L. (1999) Marine bacterial isolates display diverse responses to UV-B radiation. *Appl. Environ. Microbiol.* **65**: 3820–

3827.

Karentz, D. (1995) Considerations for evaluating ultraviolet radiation-induced genetic damage relative to Antarctic ozone depletion.

Kurth, D., Amadio, A., Ordoñez, O.F., Albarracín, V.H., Gärtner, W., and Farías, M.E. (2017) Arsenic metabolism in high altitude modern stromatolites revealed by metagenomic analysis. *Sci. Rep.* **7**: 1024.

Kurth, D., Belfiore, C., Gorriti, M.F., Cortez, N., Farias, M.E., and Albarracín, V.H. (2015) Genomic and proteomic evidences unravel the UV-resistome of the poly-extremophile Acinetobacter sp. Ver3. **6**: 1–18.

Lefort, V., Desper, R., and Gascuel, O. (2015) FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program: Table 1. *Mol. Biol. Evol.* **32**: 2798–2800.

Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**: W242–W245.

Liu, C.-M., Li, D., Sadakane, K., Luo, R., and Lam, T.-W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**: 1674–1676.

Magoč, T. and Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**: 2957–2963.

Mitchell, D.L. and Karentz, D. (1993) The Induction and Repair of DNA Photodamage in the Environment. In, *Environmental UV Photobiology*. Springer US, Boston, MA, pp. 345–377.

Nayfach, S. and Pollard, K.S. (2015) Average genome size estimation improves

comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol.* **16**: 51.

Pathak, G.P., Losi, A., and Gärtner, W. (2012) Metagenome-based screening reveals worldwide distribution of LOV-domain proteins. *Photochem. Photobiol.* **88**: 107–118.

Piacentini, R.D., Cede, A., and Bárcena, H. (2003) Extreme solar total and UV irradiances due to cloud effect measured near the summer solstice at the high-altitude desertic plateau Puna of Atacama (Argentina). *J. Atmos. Solar-Terrestrial Phys.* **65**: 727–731.

Piccini, C., Conde, D., Pernthaler, J., and Sommaruga, R. (2009) Alteration of chromophoric dissolved organic matter by solar UV radiation causes rapid changes in bacterial community composition. *Photochem. Photobiol. Sci.* **8**: 1321–1328.

Portero, L.R., Alonso-Reyes, D.G., Zannier, F., Vazquez, M.P., Farías, M.E., Gärtner, W., and Albarracín, V.H. (2019) Photolyases and Cryptochromes in UV-resistant Bacteria from High-altitude Andean Lakes. *Photochem. Photobiol.* **95**: 315–330.

Pushkarev, A. and Béjà, O. (2016) Functional metagenomic screen reveals new and diverse microbial rhodopsins. *ISME J.* **10**: 2331–2335.

Pushkarev, A., Inoue, K., Larom, S., Flores-Uribe, J., Singh, M., Konno, M., et al. (2018) A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* **558**: 595–599.

Rascovan, N., Maldonado, J., Vazquez, M.P., and Eugenia Farías, M. (2016) Metagenomic study of red biofilms from Diamante Lake reveals ancient arsenic bioenergetics in haloarchaea. *ISME J.* **10**: 299–309.

Rinnan, R., Keinänen, M.M., Kasurinen, A., Asikainen, J., Kekki, T.K., Holopainen, T., et al. (2005) Ambient ultraviolet radiation in the Arctic reduces root biomass and alters microbial community composition but has no effects on microbial biomass. *Glob.*

Chang. Biol. **11**: 564–574.

Robinson, S.A., Turnbull, J.D., and Lovelock, C.E. (2005) Impact of changes in natural ultraviolet radiation on pigment composition, physiological and morphological characteristics of the Antarctic moss, Grimmia antarctici. *Glob. Chang. Biol.* **11**: 476–489.

Robson, T.M., Pancotto, V.A., Scopel, A.L., Flint, S.D., and Caldwell, M.M. (2005) Solar UV-B influences microfaunal community composition in a Tierra del Fuego peatland. *Soil Biol. Biochem.* **37**: 2205–2215.

Roenneberg, T. and Merrow, M. (2005) Circadian clocks—the fall and rise of physiology. *Nat. Rev. Mol. Cell Biol.* **6**: 965–971.

Ruhland, C.T., Xiong, F.S., Clark, W.D., and Day, T.A. (2005) The Influence of Ultraviolet-B Radiation on Growth, Hydroxycinnamic Acids and Flavonoids of Deschampsia antarctica during Springtime Ozone Depletion in Antarctica¶†. *Photochem. Photobiol.* **81**: 1086–1093.

Russell III, J.M., Luo, M., Cicerone, R.J., and Deaver, L.E. (1996) Satellite confirmation of the dominance of chlorofluorocarbons in the global stratospheric chlorine budget. *Nature* **379**: 526–529.

Rzechorzek, N.J., Blackwood, J.K., Bray, S.M., Maman, J.D., Pellegrini, L., and Robinson, N.P. (2014) Structure of the hexameric HerA ATPase reveals a mechanism of translocation-coupled DNA-end processing in archaea. *Nat. Commun.* **5**: 5506.

S Weber (2005) Light-driven enzymatic catalysis of DNA repair: a review of recent biophysical studies on photolyase. *Elsevier*.

Sajjad, W., Qadir, S., Ahmad, M., Rafiq, M., Hasan, F., Tehan, R., et al. (2018) Ectoine: a compatible solute in radio-halophilic Stenotrophomonas sp. WMA-LM19 strain to

prevent ultraviolet-induced protein damage. *J. Appl. Microbiol.* **125**: 457–467.

Sancar, A. (2003) Structure and Function of DNA Photolyase and Cryptochrome Blue-Light Photoreceptors. *Chem. Rev.* **103**: 2203–2238.

Sancar, G.B. (2000) Enzymatic photoreactivation: 50 years and counting. *Mutat. Res. Mol. Mech. Mutagen.* **451**: 25–37.

Satinsky, B.M., Fortunato, C.S., Doherty, M., Smith, C.B., Sharma, S., Ward, N.D., et al. (2015) Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011. *Microbiome* **3**: 39.

Searles, P.S., Flint, S.D., and Caldwell, M.M. (2001) A meta-analysis of plant field studies simulating stratospheric ozone depletion. *Oecologia* **127**: 1–10.

Seitz, E.M., Brockman, J.P., Sandler, S.J., Clark, A.J., and Kowalczykowski, S.C. (1998) RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. *Genes Dev.* **12**: 1248–1253.

Selby, C.P. and Sancar, A. (2006) A cryptochrome/photolyase class of enzymes with single-stranded DNA-specific photolyase activity. *Proc. Natl. Acad. Sci.* **103**: 17696–17700.

Singh, A.H., Doerks, T., Letunic, I., Raes, J., and Bork, P. (2009) Discovering Functional Novelty in Metagenomes: Examples from Light-Mediated Processes. *J. Bacteriol.* **191**: 32 LP-41.

Smiatek, J., Harishchandra, R.K., Rubner, O., Galla, H.-J., and Heuer, A. (2012) Properties of compatible solutes in aqueous solution. *Biophys. Chem.* **160**: 62–68.

Toneatti, D.M., Albarracín, V.H., Flores, M.R., Polerecky, L., and Farías, M.E. (2017) Stratified bacterial diversity along physico-chemical gradients in high-altitude modern stromatolites. *Front. Microbiol.* **8**: 1–12.

Tran, P., Ramachandran, A., Khawasik, O., Beisner, B.E., Rautio, M., Huot, Y., and Walsh, D.A. (2018) Microbial life under ice: Metagenome diversity and in situ activity of Verrucomicrobia in seasonally ice-covered Lakes. *Environ. Microbiol.* **20**: 2568–2584.

Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., et al. (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**: 902.

Tschitschko, B., Erdmann, S., DeMaere, M.Z., Roux, S., Panwar, P., Allen, M.A., et al. (2018) Genomic variation and biogeography of Antarctic haloarchaea. *Microbiome* **6**: 113.

Ueda, T., Kato, A., Kuramitsu, S., Terasawa, H., and Shimada, I. (2005) Identification and characterization of a second chromophore of DNA photolyase from Thermus thermophilus HB27. *J. Biol. Chem.* **280**: 36237–36243.

Westbrook, A., Ramsdell, J., Bergeron, R.D., Normington, L., MacManes, M.D., Thomas, W.K., and Schuelke, T. (2017) PALADIN: protein alignment for functional profiling whole metagenome shotgun data. *Bioinformatics* **33**: 1473–1478.

Xiong, F.S. and Day, T.A. (2001) Effect of solar ultraviolet-B radiation during springtime ozone depletion on photosynthesis and biomass production of Antarctic vascular plants. *Plant Physiol.* **125**: 738–751.

Yan, A., Pan, J., An, L., Gan, Y., and Feng, H. (2012) The response of trichome mutants to enhanced ultraviolet-B radiation in Arabidopsis thaliana.

Ye, Z., Zhang, N., Wu, C., Zhang, X., Wang, Q., Huang, X., et al. (2018) A metagenomic study of the gut microbiome in Behcet's disease. *Microbiome* **6**: 135.

Zaller, J.G., Caldwell, M.M., D. Flint, S., Scopel, A.L., Salo, O.E., and Ballaré, C.L. (2002) Solar UV-B radiation affects below-ground parameters in a fen ecosystem in Tierra

del Fuego, Argentina: implications of stratospheric ozone depletion. *Glob. Chang. Biol.* **8**: 867–871.

Zhang, F., Scheerer, P., … I.O.-P. of the, and 2013, U. (2013) Crystal structure of a prokaryotic (6-4) photolyase with an Fe-S cluster and a 6, 7-dimethyl-8-ribityllumazine antenna chromophore. *Natl. Acad Sci.*

Tables

Table 1.

Metagenome	Reference	Size of Raw Metagenomic Data (Gb)	Size after merging and quality filtering (Gb)	N° of assembled Contigs (>999 bp)	Assembly size (bp)	N° of annotated proteins	N° of proteins per Mb
Lake Diamante sediment (DM)	Rascovan et al., 2016	0,74	0,64	19537	30047993	47731	1588,492
Socompa stromatolite (SS)	Kurth et al., 2017	11,4	7,57	77652	190723415	219571	1151,254
Tibetan Plateau sediment (TB)	Chen et al., 2016	3,4	2,59	42453	80284053	105553	1314,744
Amazon River (AM)	Brandon et al., 2015	9,1	5,7	41430	71904038	108943	1515,117
Lake Rauer (RA)	Tschitschko et al., 2018	5,1	5,01	64101	177770567	216623	1218,554
Dewar Creek hot spring (HT)	-	9,9	9,47	26840	162485558	169888	1045,558
Greenland cryoconite (CR)	Hauptmann et al., 2017	8,7	7,77	45376	150415849	173231	1151,68
Lake Montjoie (MT)	Tran et al., 2018	14,2	13,66	240051	545748448	693357	1270,47
Olkiluoto Island groundwater (OK)	-	24,9	8,85	74876	159143965	231244	1453,049
Human gut (GU)	Ye et al., 2018	10	9,91	47652	205076299	210369	1025,808

Figure and table legends

Figure 1. Worldwide metagenomic sampling linked to UV-B georeferenced data obtained from glUV datasets (https://www.ufz.de/gluv/). The metagenomes were selected from NCBI using biosample information: DM, Lake Diamante (ERR1824222); SS, Socompa stromatolite (SRR3341855); TB, Tibetan Plateau sediment (SRR3322106); AM, Amazon River (SRR1790676); RA, Lake Rauer (SRR6129205); HT, Dewar Creek hot spring

(SRR5580900); CR, Greenland cryoconite (SRR5275901); MT, Lake Montjoie (SRR5818193); OK, Olkiluoto Island groundwater (SRR6976411).

Figure 2. Scatterplots showing the impact of UV-B on the microbial diversity in the whole metagenomic dataset. Two parameters were measured: species richness (A) and the Shannon index (B). Abbreviations: DM, Lake Diamante; SS, Socompa stromatolite; TB, Tibetan Plateau sediment; AM, Amazon River; RA, Lake Rauer; HT, Dewar Creek hot spring; CR, Greenland cryoconite; MT, Lake Montjoie; OK, Olkiluoto Island groundwater; GU, human gut.

Figure 3. The stacked bar chart shows for all metagenomes the abundance of each DNA repair/photoprotection strategy in colors. The right vertical axis provides the numerical values of the two variables plotted as arrows: UV-B intensity ($J/m^2/day$) and Day length (minutes x 5). Abbreviations: **DM**, Lake Diamante; **SS**, Socompa stromatolite; **TB**, Tibetan Plateau sediment; **AM**, Amazon River; **RA**, Lake Rauer; **HT**, Dewar Creek hot spring; **CR**, Greenland cryoconite; **MT**, Lake Montjoie; **OK**, Olkiluoto Island groundwater; **GU**, human gut.

Figure 4. Scatterplots showing the impact of UV-B (A) and Day length (B) on the CPF abundance in the whole metagenomic dataset. Abbreviations: **DM**, Lake Diamante; **SS**, Socompa stromatolite; **TB**, Tibetan Plateau sediment; **AM**, Amazon River; **RA**, Lake Rauer; **HT**, Dewar Creek hot spring; **CR**, Greenland cryoconite; **MT**, Lake Montjoie; **OK**, Olkiluoto Island groundwater; **GU**, human gut.

Figure 5. Phylogenetic tree built with the CPF sequences from the whole metagenomic dataset.

Figure 6. Distribution of each subfamily by metagenome. Abbreviations: **DM**, Lake Diamante; **SS**, Socompa stromatolite; **TB**, Tibetan Plateau sediment; **AM**, Amazon River;

RA, Lake Rauer; **HT**, Dewar Creek hot spring; **CR**, Greenland cryoconite; **MT**, Lake Montjoie; **OK**, Olkiluoto Island groundwater; **GU**, human gut.

Figure 7. Comparative barplot showing for each metagenome the genus diversity both for whole microbial community (top bar) and the set of CPF sequences recovered from the same metagenome (lower bar). The diversity of the microbial community was estimated with Metaphlan, while the diversity of the set of CPF sequences were computed with Blast online platform. Abbreviations: **DM**, Lake Diamante; **SS**, Socompa stromatolite; **TB**, Tibetan Plateau sediment; **AM**, Amazon River; **RA**, Lake Rauer; **HT**, Dewar Creek hot spring; **CR**, Greenland cryoconite; **MT**, Lake Montjoie; **OK**, Olkiluoto Island groundwater; **GU**, human gut.

Table 1. General traits of the metagenomes used in the current study.

Supplementary table S1. Estimation of the Base Excision Repair pathway abundance.

Supplementary table S2. Estimation of the Cryptochrome/Photolyase Family abundance

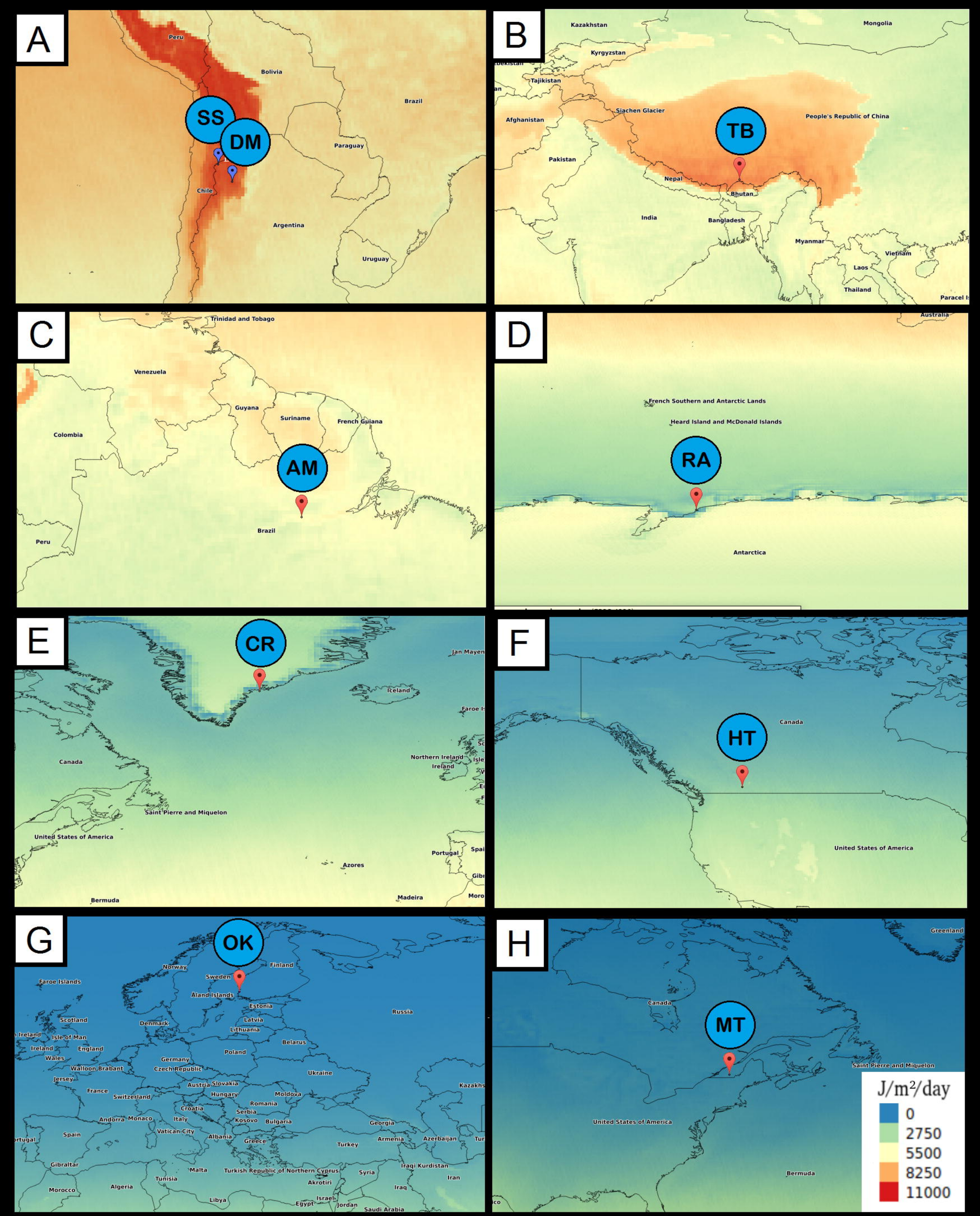
Supplementary table S3. Estimation of the Nucleotide Excision Repair pathway abundance.

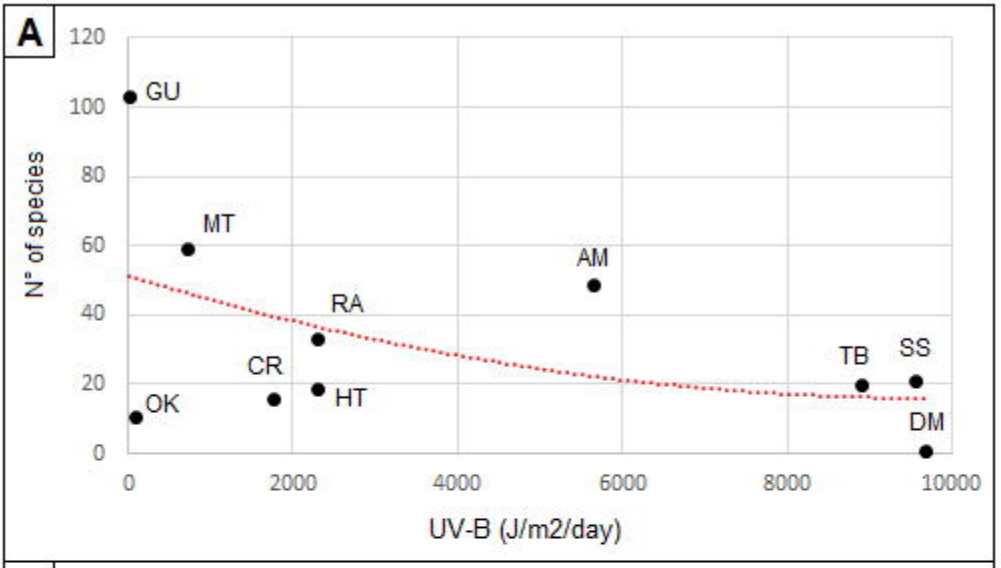
Supplementary table S4. Estimation of the Mismatch Repair Pathway abundance.

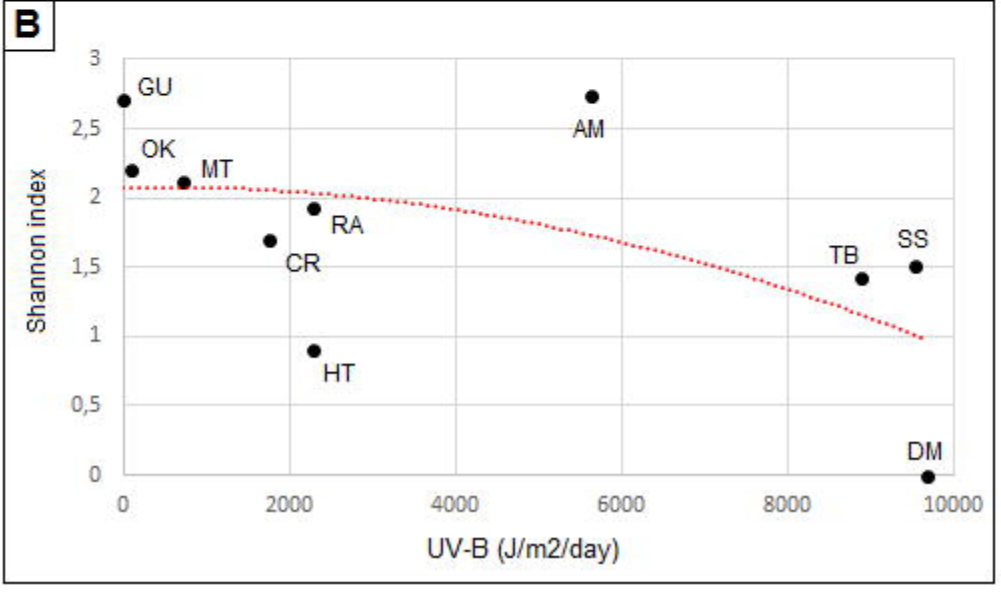
Supplementary table S5. Estimation of the Homologous Repair pathway abundance.

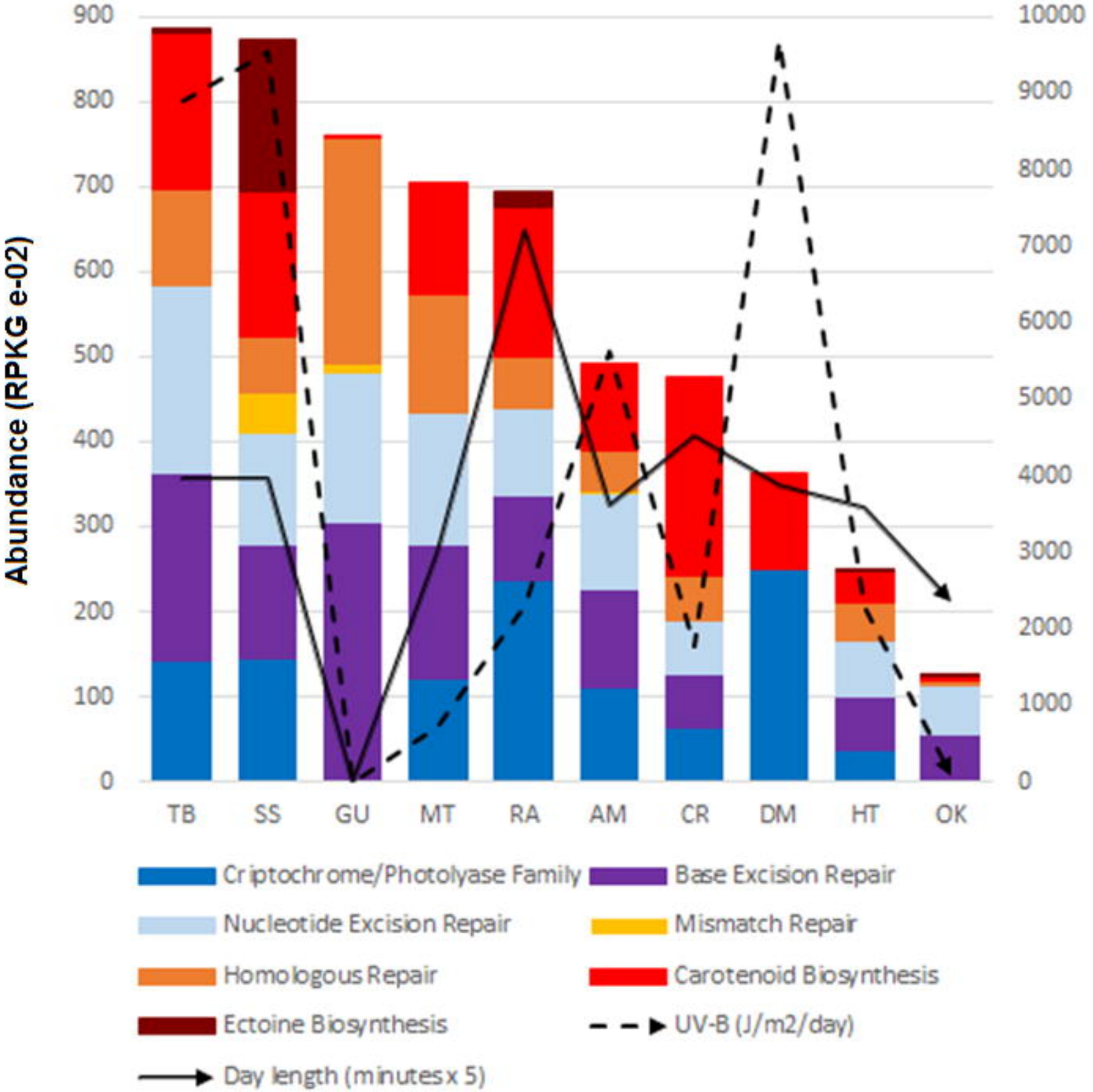
Supplementary table S6. Estimation of Carotenoid Biosynthesis pathway abundance.

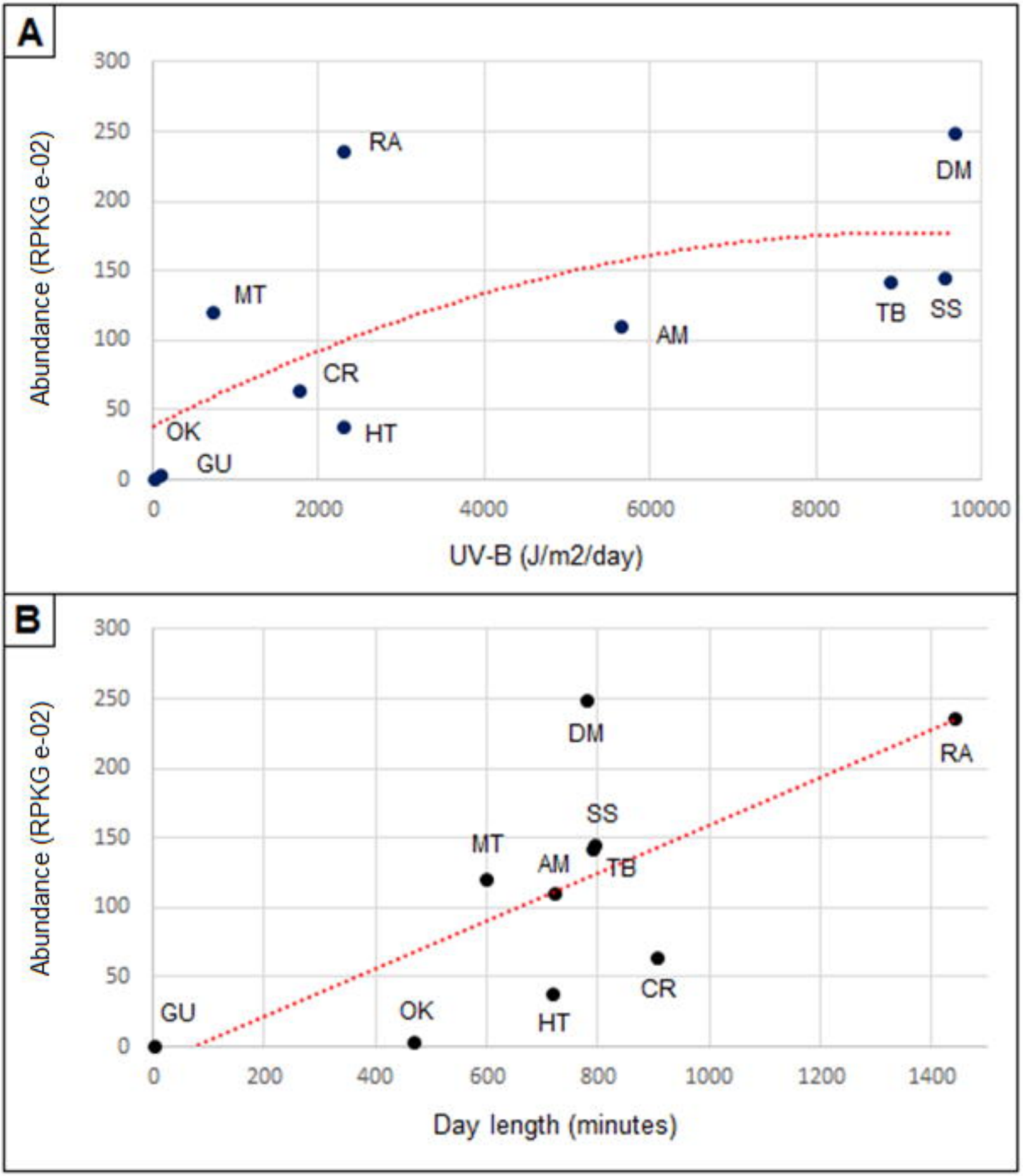
Supplementary table S7. Estimation of the Ectoine Biosynthesis pathway abundance.

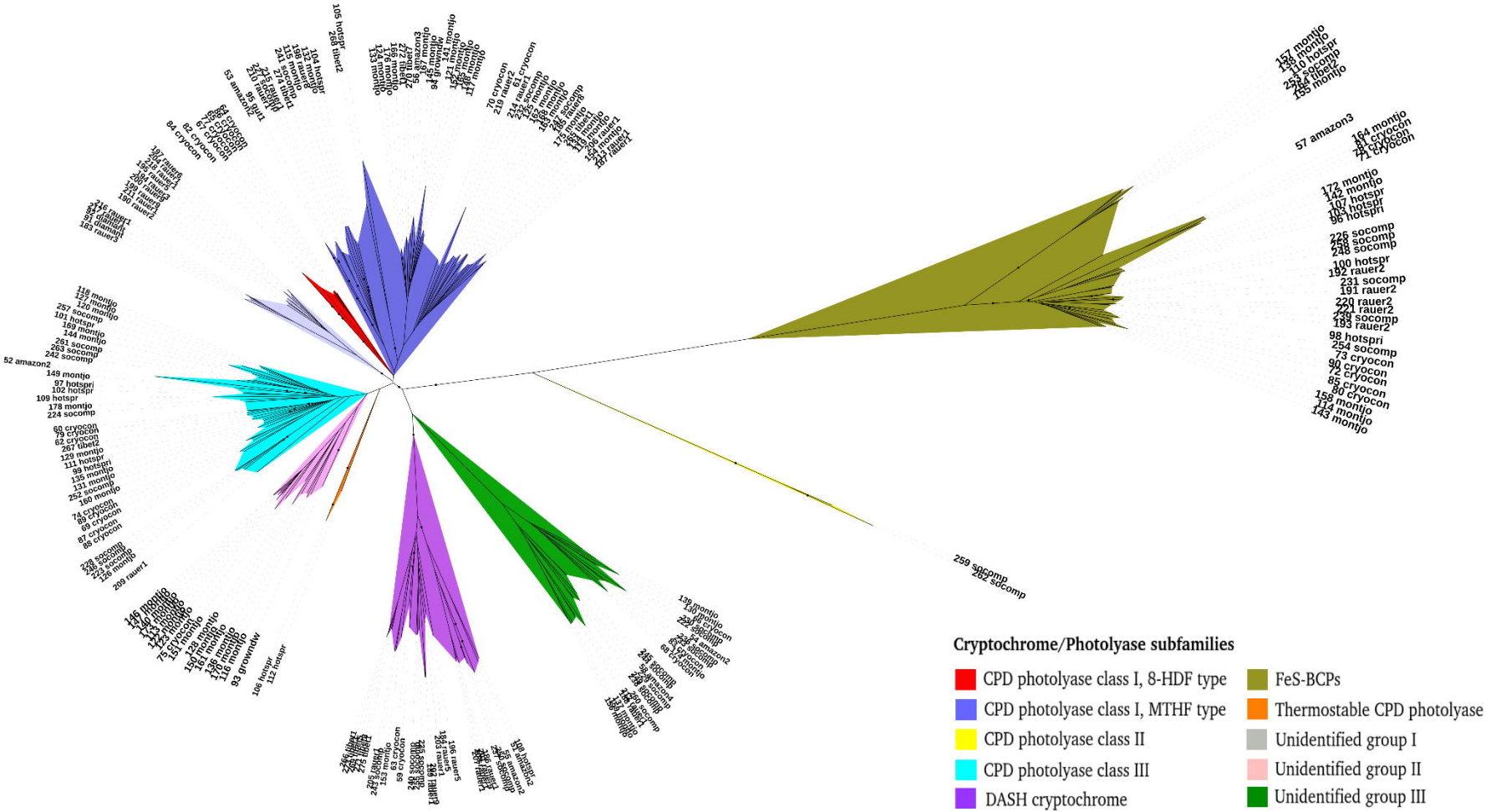

A
Peru
Bolivia
Brazil
SS
DM
Paraguay
Chile
Argentina
Uruguay
B
Kazakhstan
Mongolia
Kyrgyzstan
Tajikistan
Siachen Glacier
Afghanistan
TB
People's Republic of China
Pakistan
Nepal
Bhutan
India
Bangladesh
Myanmar
Vietnam
Laos
Thailand
C
Trinidad and Tobago
Venezuela
Guyana
Suriname
French Guiana
Colombia
AM
Brazil
Peru
D
Australia
French Southern and Antarctic Lands
Heard Island and McDonald Islands
RA
Antarctica
E
CR
Jan Mayen
Iceland
Canada
Northern Ireland
Ireland
Saint Pierre and Miquelon
United States of America
Portugal
Azores
Madeira
Bermuda
F
HT
Canada
United States of America
G
OK
Norway
Finland
Sweden
Faroe Islands
Åland Islands
Estonia
Latvia
Lithuania
Russia
Scotland
Denmark
Isle of Man
Ireland
England
Wales
Belarus
Poland
Germany
Walloon Brabant
Czech Republic
Ukraine
Jersey
Austria
Slovakia
France
Switzerland
Hungary
Moldova
Romania
Croatia
Serbia
Andorra
Monaco
Italy
Kosovo
Bulgaria
Vatican City
Georgia
Spain
Albania
Greece
Armenia
Azerbaijan
Turkey
Gibraltar
Malta
Turkish Republic of Northern Cyprus
Syria
Iraqi Kurdistan
Tunisia
Akrotiri
Iran
Morocco
Algeria
Iraq
Libya
Israel
Egypt
Jordan
Saudi Arabia
H
Greenland
Canada
MT
Saint Pierre and Miquelon
United States of America
Bermuda
J/m²/day
0
2750
5500
8250
11000

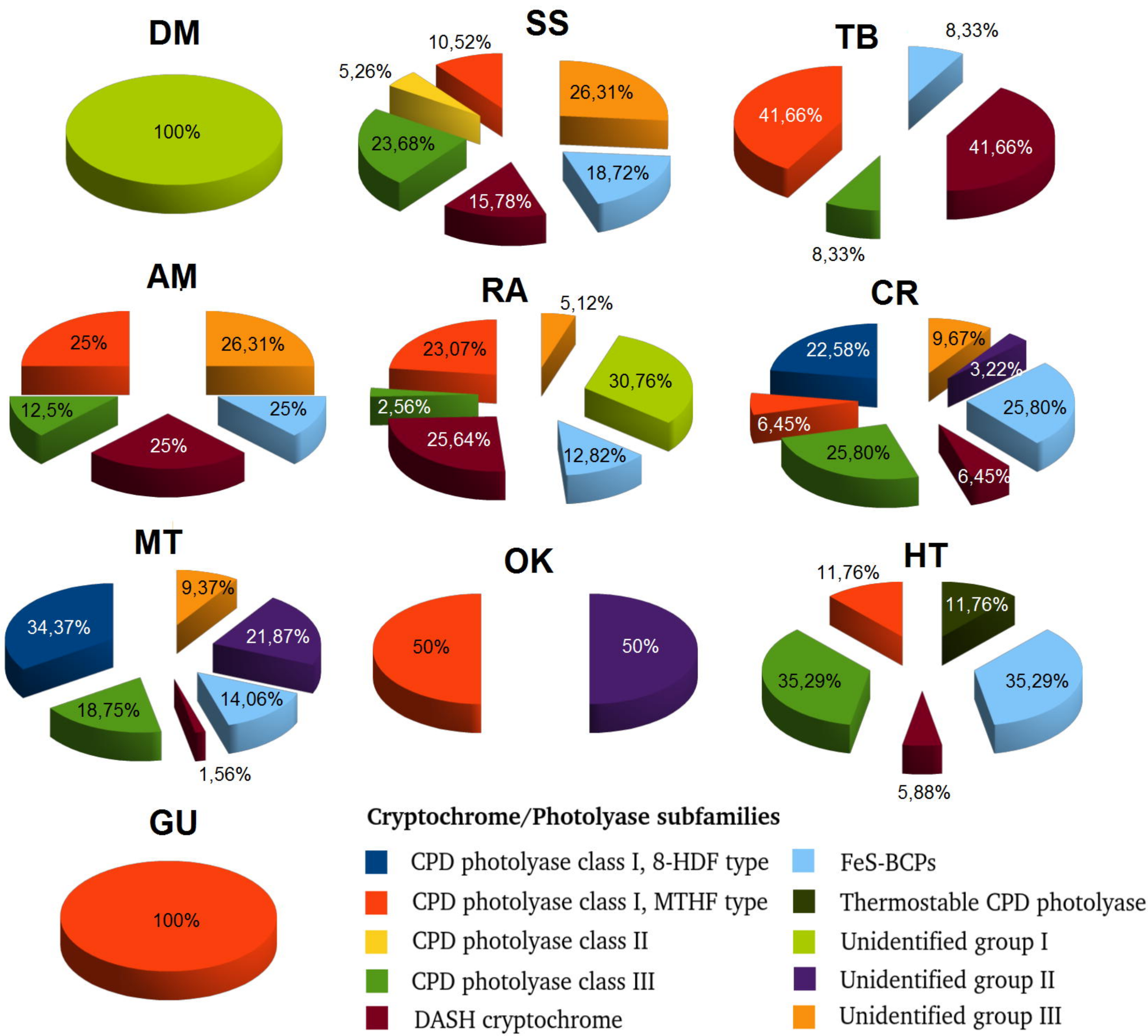

A
N° of species
120
100
80
60
40
20
0
GU
MT
AM
RA
CR
HT
OK
TB
SS
DM
0
2000
4000
6000
8000
10000
UV-B (J/m2/day)

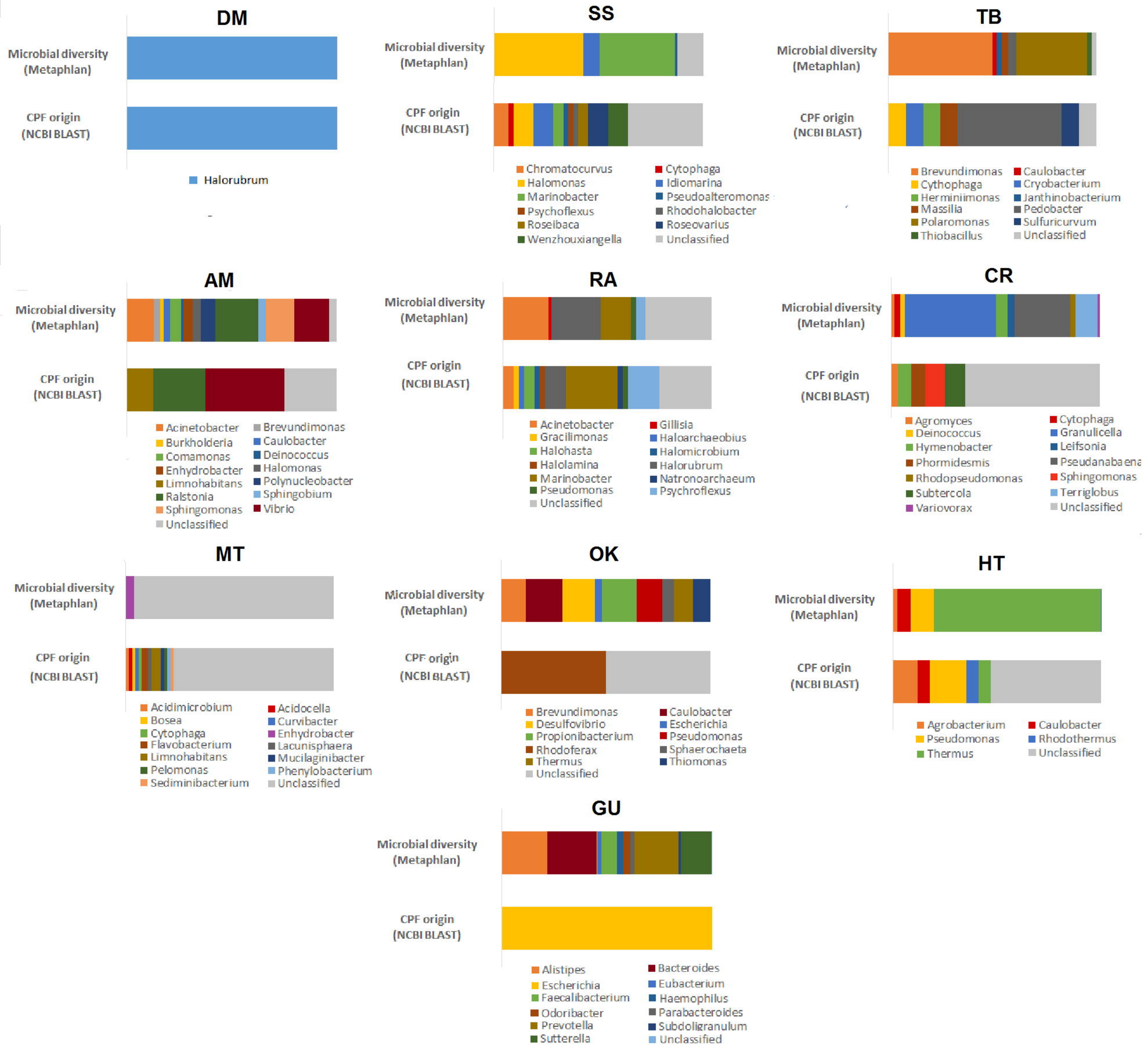

B
Shannon index
3
2,5
2
1,5
1
0,5
0
GU
OK
MT
AM
RA
CR
HT
TB
SS
DM
0
2000
4000
6000
8000
10000
UV-B (J/m2/day)


Abundance (RPKG e-02)
900
800
700
600
500
400
300
200
100
0
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
TB
SS
GU
MT
RA
AM
CR
DM
HT
OK
Criptochrome/Photolyase Family
Base Excision Repair
Nucleotide Excision Repair
Mismatch Repair
Homologous Repair
Carotenoid Biosynthesis
Ectoine Biosynthesis
UV-B (J/m2/day)
Day length (minutes x 5)


A
Abundance (RPKG e-02)
UV-B (J/m2/day)
RA
DM
MT
TB
SS
AM
CR
OK
HT
GU
B
Abundance (RPKG e-02)
Day length (minutes)
DM
RA
SS
MT
AM
TB
CR
GU
OK
HT


Cryptochrome/Photolyase subfamilies
CPD photolyase class I, 8-HDF type
CPD photolyase class I, MTHF type
CPD photolyase class II
CPD photolyase class III
DASH cryptochrome
FeS-BCPs
Thermostable CPD photolyase
Unidentified group I
Unidentified group II
Unidentified group III


DM
100%
SS
10,52%
5,26%
26,31%
23,68%
18,72%
15,78%
TB
8,33%
41,66%
41,66%
8,33%
AM
25%
26,31%
12,5%
25%
25%
RA
5,12%
23,07%
30,76%
2,56%
25,64%
12,82%
CR
22,58%
9,67%
3,22%
25,80%
6,45%
25,80%
6,45%
MT
9,37%
34,37%
21,87%
18,75%
14,06%
1,56%
OK
50%
50%
HT
11,76%
11,76%
35,29%
35,29%
5,88%
GU
100%
Cryptochrome/Photolyase subfamilies
CPD photolyase class I, 8-HDF type
CPD photolyase class I, MTHF type
CPD photolyase class II
CPD photolyase class III
DASH cryptochrome
FeS-BCPs
Thermostable CPD photolyase
Unidentified group I
Unidentified group II
Unidentified group III


DM
Microbial diversity (Metaphlan)
CPF origin (NCBI BLAST)
Halorubrum
SS
Microbial diversity (Metaphlan)
CPF origin (NCBI BLAST)
Chromatocurvus
Cytophaga
Halomonas
Idiomarina
Marinobacter
Pseudoalteromonas
Psychoflexus
Rhodohalobacter
Roseibaca
Roseovarius
Wenzhouxiangella
Unclassified
TB
Microbial diversity (Metaphlan)
CPF origin (NCBI BLAST)
Brevundimonas
Caulobacter
Cythophaga
Cryobacterium
Herminiimonas
Janthinobacterium
Massilia
Pedobacter
Polaromonas
Sulfuricurvum
Thiobacillus
Unclassified
AM
Microbial diversity (Metaphlan)
CPF origin (NCBI BLAST)
Acinetobacter
Brevundimonas
Burkholderia
Caulobacter
Comamonas
Deinococcus
Enhydrobacter
Halomonas
Limnohabitans
Polynucleobacter
Ralstonia
Sphingobium
Sphingomonas
Vibrio
Unclassified
RA
Microbial diversity (Metaphlan)
CPF origin (NCBI BLAST)
Acinetobacter
Gillisia
Gracilimonas
Haloarchaeobius
Halohasta
Halomicrobium
Halolamina
Halorubrum
Marinobacter
Natronoarchaeum
Pseudomonas
Psychroflexus
Unclassified
CR
Microbial diversity (Metaphlan)
CPF origin (NCBI BLAST)
Agromyces
Cytophaga
Deinococcus
Granulicella
Hymenobacter
Leifsonia
Phormidesmis
Pseudanabaena
Rhodopseudomonas
Sphingomonas
Subtercola
Terriglobus
Variovorax
Unclassified
MT
Microbial diversity (Metaphlan)
CPF origin (NCBI BLAST)
Acidimicrobium
Acidocella
Bosea
Curvibacter
Cytophaga
Enhydrobacter
Flavobacterium
Lacunisphaera
Limnohabitans
Mucilaginibacter
Pelomonas
Phenylobacterium
Sediminibacterium
Unclassified
OK
Microbial diversity (Metaphlan)
CPF origin (NCBI BLAST)
Brevundimonas
Caulobacter
Desulfovibrio
Escherichia
Propionibacterium
Pseudomonas
Rhodoferax
Sphaerochaeta
Thermus
Thiomonas
Unclassified
HT
Microbial diversity (Metaphlan)
CPF origin (NCBI BLAST)
Agrobacterium
Caulobacter
Pseudomonas
Rhodothermus
Thermus
Unclassified
GU
Microbial diversity (Metaphlan)
CPF origin (NCBI BLAST)
Alistipes
Bacteroides
Escherichia
Eubacterium
Faecalibacterium
Haemophilus
Odoribacter
Parabacteroides
Prevotella
Subdoligranulum
Sutterella
Unclassified