



Uncovering cryptochrome/photolyase gene diversity in aquatic microbiomes exposed to diverse UV-B regimes

Daniel G. Alonso-Reyes^{1,2}, Maria Eugenia Farias², Virginia Helena Albarracín^{1,2,3,*}

¹Laboratorio de Microbiología Ultraestructural y Molecular, Centro Integral de Microscopía Electrónica (CIME, CONICET, UNT) CCT, CONICET, 4107 Tucumán, Argentina

²Laboratorio de Investigaciones Microbiológicas de Lagunas Andinas (LIMLA), Planta Piloto de Procesos Industriales y Microbiológicos (PROIMI), CCT, CONICET, 4000 Tucumán, Argentina

³Facultad de Ciencias Naturales e Instituto Miguel Lillo, Universidad Nacional de Tucumán, 4000 Tucumán, Argentina

ABSTRACT: During evolution, microorganisms exposed to high amounts of UV-B irradiation developed fine-tuned photo-enzymes called 'photolyases' to cope with DNA damage caused by UV-B. These photoreceptors, belonging to the cryptochrome/photolyase family (CPF), have been well characterized at the genomic and proteomic level in bacteria isolated from a wide range of environments. In this work, we go further towards studying the abundance of CPF in aquatic microbial communities from different geographic regions across the globe. Metagenomics data combined with geo-referenced solar irradiation measurements indicated that the higher the UV-B level in the microbiome's environment, the higher the abundance of CPF genes and lower the microbial diversity. A connection between CPF abundance and radiation intensity/photoperiod was found. Likewise, cryptochrome-like genes were found to be abundant in most exposed microbiomes, indicating a complementary role to standard photolyases. We observed that CPFs are more likely to be present in dominant taxa of the highly irradiated microbiomes, suggesting an evolutionary force for survival and dominance under extreme solar exposure. This work reports 3 novel CPF clades, proving the potential of global metagenomic analyses in detecting novel proteins.

KEY WORDS: UV · Photolyases · Cryptochromes · Microbiomes · Metagenomics

Resale or republication not permitted without written consent of the publisher

1. INTRODUCTION

Solar radiation is an essential factor sustaining or limiting complex life on Earth. In recent decades, a drastic reduction in stratospheric ozone has occurred as a result of increased concentrations of chlorofluorocarbons (CFCs) (Aucamp 2007) and other halogen gases in the upper atmosphere (Russell et al. 1996). This depletion of ozone has resulted in an increase of biologically harmful UV radiation at the Earth's surface, together with its detrimental effects on all life forms.

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), less than 1.0 % of UV-B (280–315 nm), and virtually no UV-C (100–280 nm) (Hu et al. 2008). Biological damage is wavelength-dependent: UV-A produces mainly reactive oxygen intermediates causing indirect damage to DNA, proteins, and lipids, while UV-B and UV-C (100–280 nm) cause both indirect and direct damage to DNA because of its strong absorption at wavelengths below 320 nm (Mitchell & Karentz 1993). In accordance, numerous studies have reported UV-B effects on plants (Searles et al. 2001, Xiong & Day 2001, Robinson et al. 2005, Ruhland et al. 2005, Jansen et al. 2010, Yan et al. 2012), animals (Robson et al. 2005, Bao et al. 2014), microorganisms (Zaller et al. 2002,

*Corresponding author: valbarracin@csnat.unt.edu.ar

Avery et al. 2004, Rinnan et al. 2005, Piccini et al. 2009), and ecosystems (Garcia-Pichel 1994, Karentz 1995, Joux et al. 1999, Ballaré et al. 2011, Häder et al. 2011).

Solar UV radiation can damage aquatic organisms and decrease the productivity of aquatic ecosystems. Beyond the targets for damage by UV radiation (DNA, lipids, protein) that are common for all biological systems, a major site of damage in primary producers is the photosynthetic machinery, including Photosystem II and the accessory pigments that funnel light energy to the reaction centers (Häder & Gao 2015). Subsequent damage will directly reduce primary production. The effects of UV radiation may also reduce the photosynthetic uptake of atmospheric carbon dioxide and affect species diversity, ecosystem stability, trophic interactions, and global biogeochemical cycles (Häder et al. 2011, Williamson et al. 2019).

Over evolutionary time, microorganisms exposed to high levels of UV and other DNA-damaging factors have developed specific and highly conserved DNA repair mechanisms. The most important are photoreactivation, nucleotide excision repair (NER), base excision repair, mismatch repair, and homologous repair. Other processes, such as damage tolerance (dimer bypass), SOS response, checkpoint activation, and programmed cell death efficiently act against DNA lesions, ensuring genomic integrity.

The finest and most efficient mechanism to repair DNA damage by UV-B is photoreactivation, executed by photoreactivating enzymes known as 'photolyases' which target the main products of UV-B, pyrimidine dimers (PDs). Such PD lesions bring polymerases to a standstill, eventually leading to cell death. Photolyases bind tightly to PDs in the dark and can be activated by light from different wavelengths, splitting the 2 PD bonds and resulting in the reformation of the 2 separate pyrimidine bases (Weber 2005). Photolyases are classified according to the type of photoproduct that is repaired: cyclobutane pyrimidine dimer (CPD-Phr) photolyases or (6-4) pyrimidine-pyrimidone photolyases (6-4-Phr). These enzymes, together with the related cryptochromes (Cry), constitute the cryptochrome/photolyase family (CPF). However, Cry have no photolyase activity, functioning as signaling molecules regulating diverse biological responses such as entrainment of circadian rhythms in plants and animals (Roenneberg & Merrow 2005, Harmer 2009). It has been argued that CPF evolved early in the history of life due to the need of primitive organisms to cope with extreme solar radiation in the time when no stratospheric ozone existed to protect the Earth (Kanai et al. 1997, Portero et al. 2019).

Investigations at the metagenome level about effects of sunlight on microbial communities have mostly involved photoreceptors such as microbial rhodopsins (Pushkarev & Béjà 2016, Pushkarev et al. 2018) and LOV-domains (Pathak et al. 2012). Likewise, Singh et al. (2009) reported abundances for several light-related genes in microbiomes from different environments. However, the abundance and diversity of CPF in microbiomes at a world-wide scale and light range have not been studied so far, even though these photoreceptors are a vital mechanism in the defense against UV-B. In this work, we studied aquatic microbiomes exposed to different intensities of UV-B under different photoperiods around the globe using metagenomic analysis of existing published DNA data sets. Our objective was to characterize the occurrence and diversity of CPF on microbiomes from diverse aquatic ecological niches. The results add a new dimension to our understanding of the short-term effects of climate change and atmospheric ozone depletion on environmental microbial communities. The focus of this work on microbial communities from extreme environments provides models for early-life research.

2. MATERIALS AND METHODS

2.1. Metagenome selection and UV-B data processing

All metagenomes were retrieved from the SRA-NCBI database. They were sequenced by a shotgun strategy through Illumina technology except the Lake Diamante metagenome, which was sequenced with 454 GS FLX Titanium chemistry. The accession numbers are as follows: Lake Diamante (DM) (ERR1824222), Socompa stromatolite (SS) (SRR3341855), Tibetan Plateau sediment (TB) (SRR3322106), Amazon River (AM) (SRR1790676), Lake Rauer (RA) (SRR6129205), Dewar Creek hot spring (HT) (SRR5580900), Greenland cryoconite (CR) (SRR5275901), Lake Montjoie (MT) (SRR5818193), Olkiluoto Island groundwater (OK) (SRR6976411), and human gut (GU) (SRR6517782).

An essential aspect of this work was to reanalyze published metagenomic data of globally distributed microbiomes taking into account their overall UV-B exposure (Fig. 1). Monthly mean UV-B data sets ('glUV'; Beckmann et al. 2014) processed using QGIS (www.qgis.org) were utilized to calculate the UV-B exposure of the samples of the selected metagenomes according to the date and location of sampling available at the biosample-linked section of each SRA

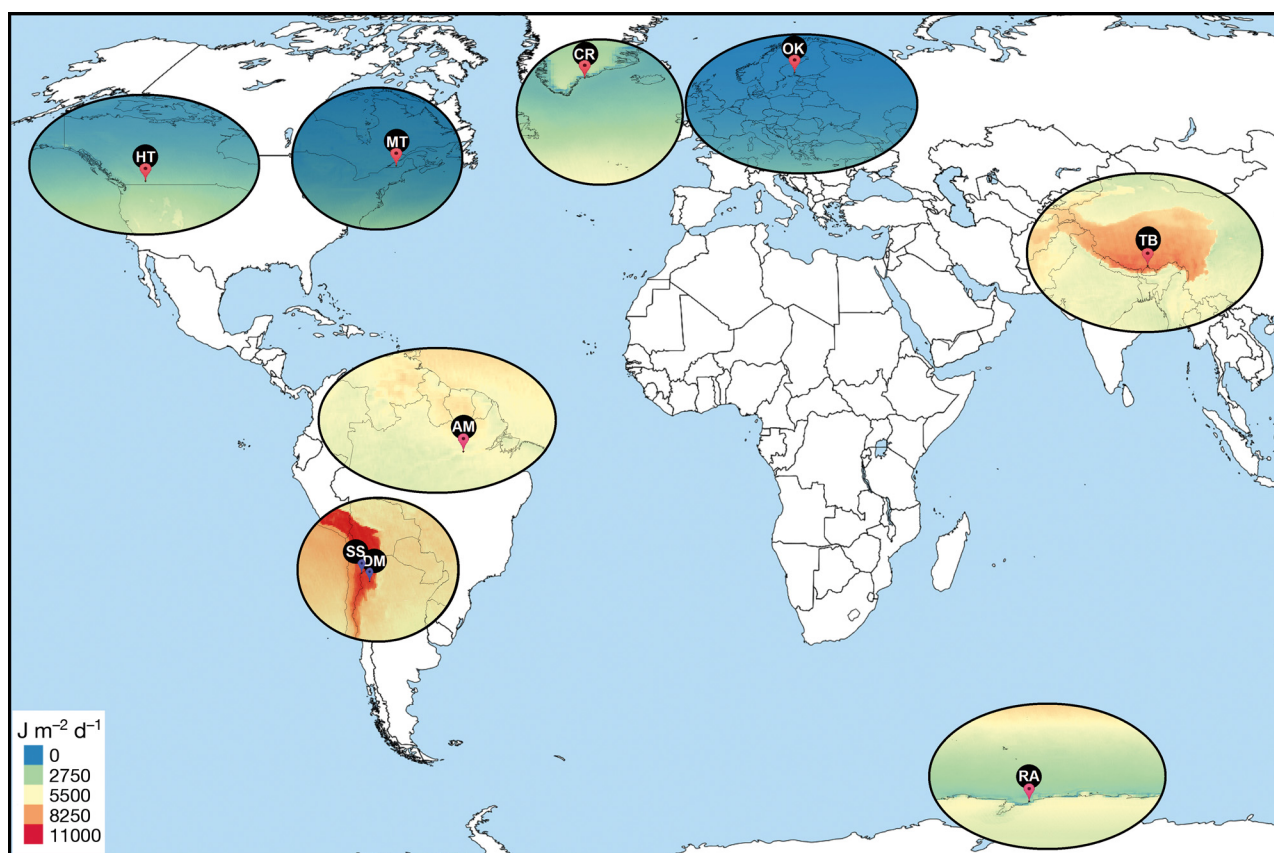


Fig. 1. World-wide metagenomic sampling linked to UV-B geo-referenced data obtained from glUV data sets (<https://www.ufz.de/gluv/>). UV-B intensities for each metagenome at the time of sampling are shown in colors. The metagenomes were selected from NCBI using biosample information: DM: Lake Diamante; 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

entry. To compensate for a lack of replicates in most samples, we performed the analysis by clustering the data sets into 3 groups according to UV-exposure regimes: low (UV_{Low}), mid (UV_{Mid}), and high-exposed (UV_{High}). We considered the UV_{Low} range as between intensities of 0 and 2000 $J m^{-2} d^{-1}$, the UV_{Mid} as between intensities of 2000 and 6000 $J m^{-2} d^{-1}$, and UV_{High} range as between intensities of 6000 and 12000 $J m^{-2} d^{-1}$. The metagenome of the GU was assumed as a UV-B-free environment, acting as a negative control. See Table 1 for the main characteristics of the environmental data sets. A brief description of the environmental conditions of each sampling site is summarized below, including citations of previous works describing the corresponding microbial communities in further detail.

DM samples were taken from red biofilms attached to the bottom of calcareous stones submerged in Lake Diamante on 15 February 2012. This lake is located inside the Galan Volcano crater, at 4589 m above sea level (a.s.l.) ($26.008^{\circ}S$, $67.043^{\circ}W$). High

pH (9–11), high arsenic concentrations (115–234 $mg l^{-1}$), high salinity (270 $g l^{-1}$, 217 $mS cm^{-1}$), high UV radiation (84 $W m^{-2}$ of UVA-B at noon), large day-night temperature range (-20 to $+20^{\circ}C$), and low O_2 pressure constitutes a set of unique extreme conditions that prevails in the lake (Rascovan et al. 2016).

An SS sample was taken from stromatolites located along the southern shore of Socompa Lake ($24.536^{\circ}S$, $68.202^{\circ}W$) at the base of the Socompa Volcano (3570 m a.s.l.). These structures were found around the border of Socompa Lake during the summer when they are partially exposed depending on the tide and hydrological regime; they are entirely submerged during winter and spring. The stromatolites are rounded, domal structures that present clear stratification in vertical sections. The sample was taken as a 5 cm deep cylinder from the top of the stromatolite. The site was exposed to air at the moment of sampling in February 2011. Extreme environmental conditions at the lake include hypersalinity, high thermal amplitude with daily tempera-

tures that range from -10 to 20°C in summer and -20 to 10°C in winter, UV solar irradiance that reaches 68 W m^{-2} , low O_2 pressure, low nutrient availability, and, primarily, high arsenic content (18.5 mg l^{-1}) (Kurth et al. 2017).

TB was collected on 1 August 2013 from sediments of the Qiangyong Glacier Lake (28.51°N , 90.12°E), located at the southern part of the Tibetan Plateau, in the Indian monsoon climate region. The temperature, pH, and electric conductivity of surface water are, on average, 6.89°C , 8.32, and $136.4\text{ }\mu\text{S cm}^{-1}$, respectively (Chen et al. 2016).

The AM water sample was collected on 19 May 2011 from the Amazon River at Obidos station (1.919°S , 55.525°W) at 33 m depth, by pumping water with a Shurflo submersible pump through a $297\text{ }\mu\text{m}$ stainless steel screen. The temperature, pH, and O_2 were 28.4°C , 6.6, and 3.2 mg l^{-1} , respectively, at the site of sampling (Satinsky et al. 2015, Doherty et al. 2017).

RA was sampled on 12 January 2015 from Rauer Lake (Torckler Island; 68.555°S , 78.191°E). Water temperature and approximate depth of water at the sampling site were 9°C and 10–15 cm. The lake was shallow ($<30\text{ cm}$). There was a crust of salt crystals on the sediment (which makes the lake look frozen from the air). There also appeared to be stratification in the lake: a top clear layer of $\sim 10\text{ cm}$ and a bottom layer of $\sim 5\text{ cm}$. These layers were not visible until they were disturbed, and a visible haze was produced when the layers were mixed (Tschitschko et al. 2018).

CR came from a cryoconite sample collected on 29 August 2013 from the Greenland ice sheet at the TAS-U-A1 site (65.41°N , 38.51°W). The pH, altitude, and depth of sampling were 5.53, 580 m, and 20 cm, respectively. Cryoconite was collected from 2–18 holes at each site, depending on their availability, using sterile 50 ml syringes (Stibal et al. 2015a,b, Hauptmann et al. 2017).

The MT microbial sample was collected from Lake Montjoie, Canada (45.409°N , 72.099°W) in February 2014; the lake was covered in ice at the time of sampling. The sample was taken from the epilimnion strata of the lake (0–3 m); pH was 6.41 (Tran et al. 2018).

The HT sediment sample was collected from the Dewar Creek hot spring, Canada (49.954°N , 116.515°W), near the source of the hot spring on 28 September 2012. The pH and temperature at the site were 7.93 and 66.4°C (Eloe-Fadrosch et al. 2016).

The OK sample was from groundwater collected from a deep subsurface site KR11_0.1 (61.241°N , 21.494°E) in Olkiluoto Island, on the south-west coast of Finland on 9 September 2016. Based on the environmental data provided by the authors of the origi-

nal research (Bomberg et al. 2016), samples were collected from 60 drill holes from different fractures at different depths (296–798 m below sea level). The groundwater is stratified, with a salinity gradient extending from fresh to brackish water to a depth of 30 m and the highest salinity concentration of 125 g l^{-1} total dissolved solids at 1000 m depth. Between 100 and 300 m depths, the groundwater originates from ancient (pre-Baltic) seawater and has high concentrations of SO_4^{4-} . The temperature rises linearly with depth, from ca. $5\text{--}6^{\circ}\text{C}$ at 50 m to ca. 20°C at 1000 m. The pH of the groundwater is slightly alkaline throughout the depth profile. The bedrock of Olkiluoto consists mainly of mica gneiss and pegmatitic granite-type rocks. The *in situ* temperature at 300 m depth in the Olkiluoto bedrock is stable at approximately 10°C and increases linearly to approximately 16°C at 800 m depth.

GU corresponds to a gut metagenome sample collected from the feces of an active Behcet's disease patient (Ye et al. 2018) and was used as a negative control with no exposure to UV.

2.2. Meta-analysis, quality control, and assembly of selected metagenomes

Adapters were removed from the Illumina raw reads using the 'fastq-mcf' tool of 'ea-utils' v.1.04.676 (Aronesty 2011). This step was not needed for the LD metagenome, as it was sequenced with 454 technologies. Quality filtering and trimming were performed with the same program, using parameters $l > 50$ and $q > 20$, where l = minimum remaining sequence length and q = quality threshold causing base removal. The program 'kneaddata' v.0.6.1 (<https://github.com/biobakery/kneaddata>) was used with the '—bypass-trim' option to clean human contaminant sequences. Paired-end reads were merged with FLASH v.1.2.11 (Magoč & Salzberg 2011) to recover unpaired longer reads. The pair-end reads with a low percentage of merging were left in the paired state for assembly. Assembly of filtered reads was performed with MEGAHIT v.1.1.2 (Li et al. 2015). Genes were predicted in assembled contigs with Prodigal v.2.6.2 (Hyatt et al. 2010), which outputted translated protein sequences.

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

2.3. Reference database building

Considering the main functional orthologous genes of the CPF group, a reference protein database was built. The 2 KEGG Orthology numbers corresponding to CPF members (K06876, K01669) were linked with the UniProt database, and then results were filtered by EC number, known molecular function or biological process, uniref 90 clusters, and sequence length >100.

2.4. Estimation of CPF gene abundance and microbial diversity

To search and quantify the abundance of genes in the different metagenomes, we aligned the pre-processed unpaired reads against the CPF reference database previously created. Protein alignment was performed with PALADIN v.1.3.1 (Westbrook et al. 2017), which outputs a table with alignment counts for each gene. Alignments were filtered by maximum quality (= 60). The outputted read counts were normalized to reads per kb per genome, and the number of genomes was computed through MicrobeCensus v.1.1.0 (Nayfach & Pollard 2015). Normalization was done in the following way:

$$\frac{\text{Read Counts} \div \text{Number of Genomes}}{\text{Gene Length}}$$

Diversity analysis of each metagenome was performed with MetaPhlAn v.2.7.7 (Truong et al. 2015) with the ‘—bt2_ps’ parameter set to ‘sensitive’. The MetaPhlAn program profiles the composition of microbial communities from metagenomic shotgun sequencing data at the species level. It relies on ~1 M unique clade-specific marker genes identified from ~17 000 reference genomes (~13 500 bacterial and archaeal, ~3500 viral, and ~110 eukaryotic). The Shannon diversity index was obtained with the R package ‘vegan’ v.2.5-2 (Oksanen et al. 2018).

2.5. Analysis of the annotated genes

The protein sequences were aligned, using Diamond v.0.9.22 software (Buchfink et al. 2015), 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 and filtered by sequence length of >400 amino acid residues.

The filtered protein sequences were used for phylogenetic analysis. The phylogenetic tree was built with ‘FastME’ v.2.0 (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 & Bork 2016). Additionally, a taxonomical identification of the sequences was performed through the BLAST web server (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

3. RESULTS AND DISCUSSION

3.1. UV-B intensity profiles on world-wide microbiomes

UV-B geo-referenced data helps us to group microbiomes according to their UV-exposure regimes as UV_{High} , UV_{Mid} , and UV_{Low} . The UV_{High} samples were DM, SS, and TB, with UV-B intensities of 9677, 9536, and 8885 $J m^{-2} d^{-1}$, respectively. In contrast, UV_{Low} microbiomes OK, MT, and CR were exposed to low UV-B intensities of 77, 719, and 1759 $J m^{-2} d^{-1}$, respectively. The UV_{Mid} microbiomes AM, RA, and HT were exposed to intensities of 5630, 2289, and 2281 $J m^{-2} d^{-1}$, respectively. It is important to note that these intensity values were registered taking into account the sampling dates; for instance, the samples of the UV_{High} group were registered during summer, when insolation is maximum at the high-altitude environments of the Argentinean Puna and Tibetan Plateau (February and August) and UV-B incidence becomes the highest on the planet. On the other hand, the UV_{Low} microbiomes were sampled during August, November, and February which corresponds to the summer, autumn, and winter seasons in the Northern Hemisphere, with lower levels of irradiation compared to the Southern Hemisphere. In turn, the UV_{Mid} samples included the AM microbiome from a tropical environment and RA and HT from high-latitude extreme cold and hot environments, respectively.

The calculated day length from the UV data sets for each geo-referenced metagenome (www.suncalc.org) revealed that 2 microbiomes experienced peculiar photoperiods at the time of sampling. While the OK microbiome was experiencing a short photoperiod (only 7.8 h), RA was sampled when Antarctica was experiencing 24 h of daylight.

We assumed that the environments had been exposed to similar conditions of solar irradiance for a

considerable period before sampling. As a consequence, the diversity of species and genes on genomes is a reflection of the ecological pressure of radiation in the environment (among other factors), since the geographic conditions (latitude, altitude, and orography) did not change considerably in those sampled regions for decades. Although not available, transcriptomic data from these sites would be an excellent tool to evaluate the reflection of environmental pressure on the dynamics of the microbial communities.

The intensity of solar irradiation over each microbiome also depends on the on-site spatial disposition, being at a maximum for samples taken from soil or from shallow waters of lakes, springs, oceans, and streams but much lower in sediments, groundwater, or deep water. Some of the microbiome samples were not directly exposed to UV, but instead were sheltered in a protective environment. Nevertheless, during the biofilm or stromatolite formation process, microbes must have been directly exposed to the harsh UV radiation without a protective shelter, and managed to survive due to their strong capacity to deal with high UV irradiation. Many species of these communities could present migration behavior, a life cycle, or an evolutionary history in which they could have been exposed to radiation.

It is well known that UV-B negatively affects microbial diversity (Balaré et al. 2011); nevertheless, this effect has not yet been assessed in a systematic way across different irradiation regimes at a global scale. Our results for species richness and the Shannon index in diverse metagenomes suggest an interesting trend (Fig. 2); TB, SS, and DM from the UV_{High} group had the lowest diversity, while the GU microbiome, which received no irradiation, had the maximum diversity. Moreover, microbiomes exposed to intensities of less than 4000 J m⁻² d⁻¹ revealed a high range of species richness (number of species: 11–103), whereas the SS microbiome with intensities above 8000 J m⁻² d⁻¹ only showed 21 species.

Species richness (Fig. 2A) was variable for microbiomes whose UV-B exposure was below a certain threshold (6000 J m⁻² d⁻¹). In microbiomes receiving higher UV doses, the number of species decreased, indicating selective pressure. Those species that do not possess efficient molecular mechanisms to defend themselves from UV-B become 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 abundance in addition to its ability to detect rare species): microbiomes of the UV_{High} group and HT (UV_{Mid}) presented the lowest Shannon index values.

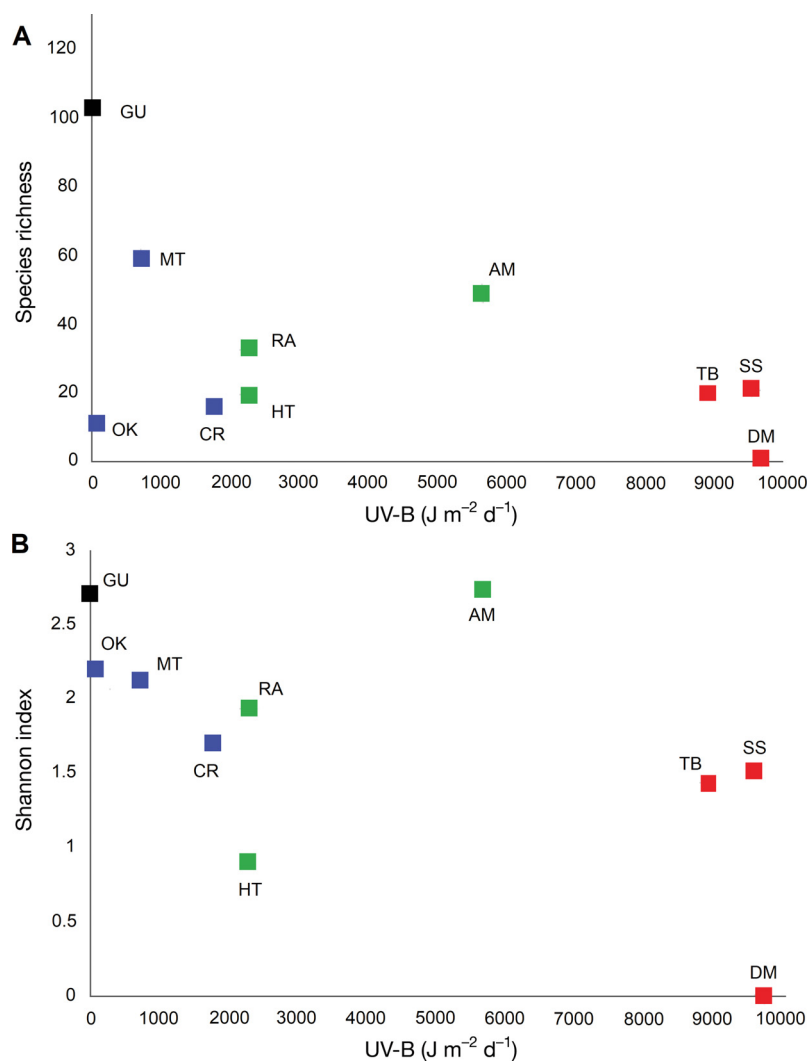


Fig. 2. Impact of UV-B on the microbial diversity in the whole metagenomic data set. Two parameters were measured: (A) species richness and (B) the Shannon index. Metagenomes were colored according to the group to which they belong: red: UV_{High}; green: UV_{Mid}; blue: UV_{Low}. DM: Lake Diamante; SS: Socoma 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

These results suggest that those micro-organisms with full capacity to defend themselves against UV-B radiation reach dominance, while less fit species vanish. Nevertheless, this correlation is not conclusive, as many other chemical and physical factors (salinity, nutrient availability, heavy metals) may cause selective pressure on microbial diversity.

3.2. The meta-analysis, quality control, and assembly of selected metagenomes

Quality filtering and merging yielded a range between 0.64 and 13.66 Gb for further analysis (Table 1). Only the TB and AM data sets reported a high percentage of merging with FLASH (Magoč & Salzberg 2011); however, the low-merged data set from OK was also used for further downstream steps, as it was a reasonable size (within the range mentioned above). MEGAHIT (Li et al. 2015) was used for assembly since that program can deal with large and complex data sets 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⁻¹, which is congruent with the high coding density expected for microbial species.

3.3. Quantitative analysis of CPF

In order to evaluate the abundance of CPF genes in each microbiome, a database was built using Uniprot sequences linked to KEGG orthology and then aligned to the query metagenomes. Our results show that the distribution of CPF genes in metagenomes followed a predicted ecological pattern; CPF genes were abundant in microbiomes with UV_{High} and UV_{Mid}, while they were insignificant or non-existent in the UV_{Low} microbiomes OK and GU, respectively. In the DM sample, CPF genes appeared profusely. Their abundance was 249 e⁻⁰² RPKG (reads per kilobase

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

Metagenome	Description	Depth of sampling (cm)	Reference	Size after merging and quality filtering (Gb)	Size of raw metagenomic data (Gb)	No. of assembled contigs (>999 bp)	Assembly size (bp)	No. of annotated proteins	No. of proteins per Mb
Lake Diamante sediment	Red biofilm from Lake Diamante, located inside the Galan Volcano crater at 4589 m a.s.l.	–	Rascovan et al. (2016)	0.74	0.64	19 537	30 047 993	47 731	1588
Socompa stromatolite	Sample from a high-altitude stromatolite in the surroundings of Lake Socompa	0–5	Kurth et al. (2017)	11.4	7.57	77 652	190 723 415	219 571	1151
Tibetan Plateau sediment	Sediment sample collected from Qiangyong Glacier Lake	–	Chen et al. (2016)	3.4	2.59	42 453	80 284 053	105 553	1314
Amazon River	Water sample from the Amazon River at a site facing Óbidos city, Brazil	0–3300	Satinsky et al. (2015)	9.1	5.7	41 430	71 904 038	108 943	1515
Lake Rauer	Saline lake microbial sample from Rauer Islands, Antarctica	10–30	Tschitschko et al. (2018)	5.1	5.01	64 101	177 770 567	216 623	1218
Dewar Creek hot	Hot spring sediment metagenome from Dewar Creek, Canada	–	–	9.9	9.47	26 840	162 485 558	169 888	1045
Greenland cryoconite	Cryoconite sample from the Greenland ice sheet	20	Hauptmann et al. (2017)	8.7	7.77	45 376	150 415 849	173 231	1151
Lake Montjoie	Freshwater microbial communities from epilimnion strata of Lake Montjoie, Canada	0–300	Tran et al. (2018)	14.2	13.66	240 051	545 748 448	693 357	1270
Olkiluoto Island groundwater	Groundwater microbial communities from the Olkiluoto Island deep subsurface site, Finland	>29 600	–	24.9	8.85	74 876	159 143 965	231 244	1453
Human gut	Metagenome of fecal sample collected from an active Behet's disease patient	–	Ye et al. (2018)	10	9.91	47 652	205 076 299	210 369	1025

per genome), which was higher than in the rest of the microbiomes.

The relationship between the abundance of CPF and the intensity/time of insolation is provided in Fig. 3A, and shows an upward trend of CPF abundance as UV-B irradiation increases. However, RA (a UV_{Mid} microbiome) completely deviates from the trend because of its high abundance. We had previously mentioned that RA had the most extended photoperiod of all the samples, so we set out to verify if there was a connection between the abundance of CPF and photoperiod. Fig. 3B shows that there is an upward trend of CPF as the photoperiod increases, with RA quite in line with the trend. Thus, both factors—intensity and photoperiod—could be influ-

encing the abundance of CPF in the microbial communities.

Our study revealed a rising trend of CPF abundance in microbial communities in environments with higher radiation levels or longer day durations. That result is likely due to an increase in DNA damage caused by UV-B, which is mainly constituted by CPD and (6-4) photoproducts. In such situations, populations could increase the expression of CPF by increasing the gene copy number, or species lacking CPF in their genome would be replaced by species that contain these genes. In either case, the overall increase in CPF copy abundance indicates that the community improves its defense capabilities against UV-B by using the highly efficient mechanism

of photoreactivation or modulating enzymatic mechanisms triggered by light-sensing Cry.

3.4. Phylogenetic analysis of CPF genes

The CPF group of proteins comprises mostly genes of photolyases with different specializations and Cry, whose functions are mostly unknown. The family has been divided into different subgroups considering phylogeny, chromophore type, specialization, host organism, and structure. Alignment, editing, and subsequent phylogenetic analysis of 214 CPF sequences found in the metagenomes configured a robust tree with numerous clades (Fig. 4). The majority of sequences grouped within well-known subfamilies: the CPD photolyases classes I, II, and III, DASH cryptochromes (cry-DASH), the FeS-BCPs group, and the thermostable CPD photolyases, a new group with an unstable position in phylogeny (Ueda et al. 2005, Portero et al. 2019). This work also reports 3 novel clades: unidentified I, II, and III (UI, UII, and UIII), which have 98, 53, and 85% of bootstrap support, respectively. The 3 new clades amount to 24.76% of the whole sequences in the analysis (Fig. 5), thus offering a large pool of candidates for new functions, specializations, or molecular specificities. Interestingly, clades UI and UII are mainly composed of se-

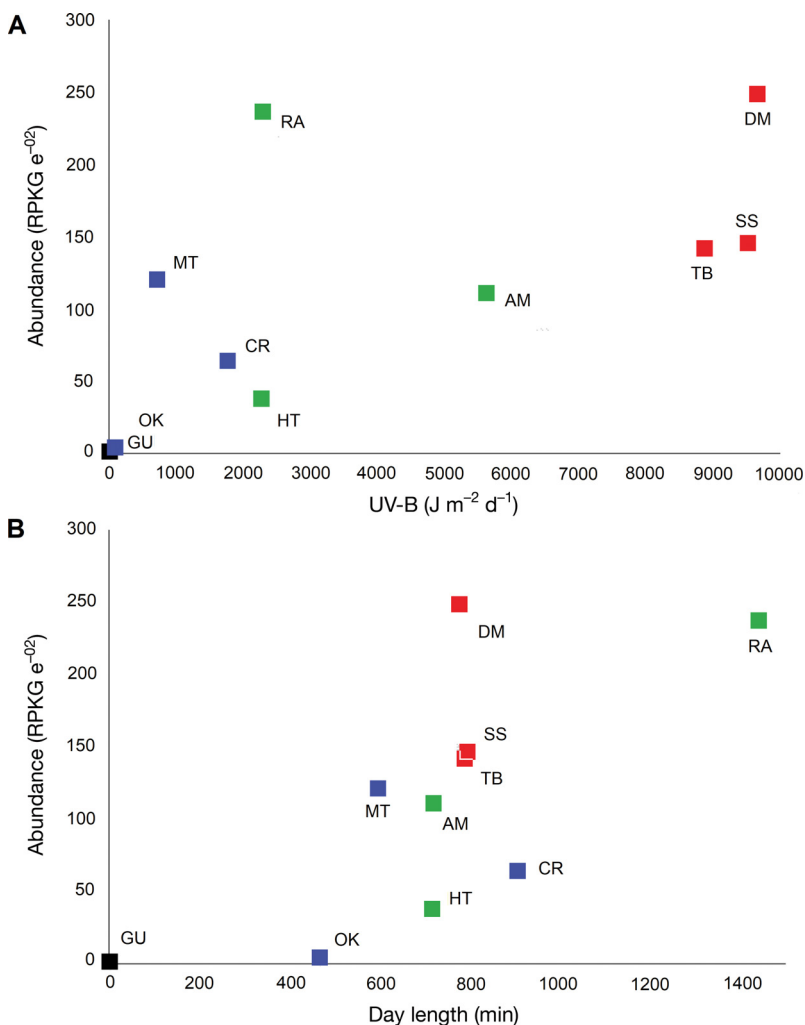


Fig. 3. Impact of (A) UV-B and (B) day length on the cryptochrome/photolyase family abundance in the whole metagenomic data set. Metagenomes were colored according the group to which they belong: red: UV_{High}; green: UV_{Mid}; blue: UV_{Low}. RPKG: reads per kilobase per genome. See Fig. 2 for sample location abbreviations

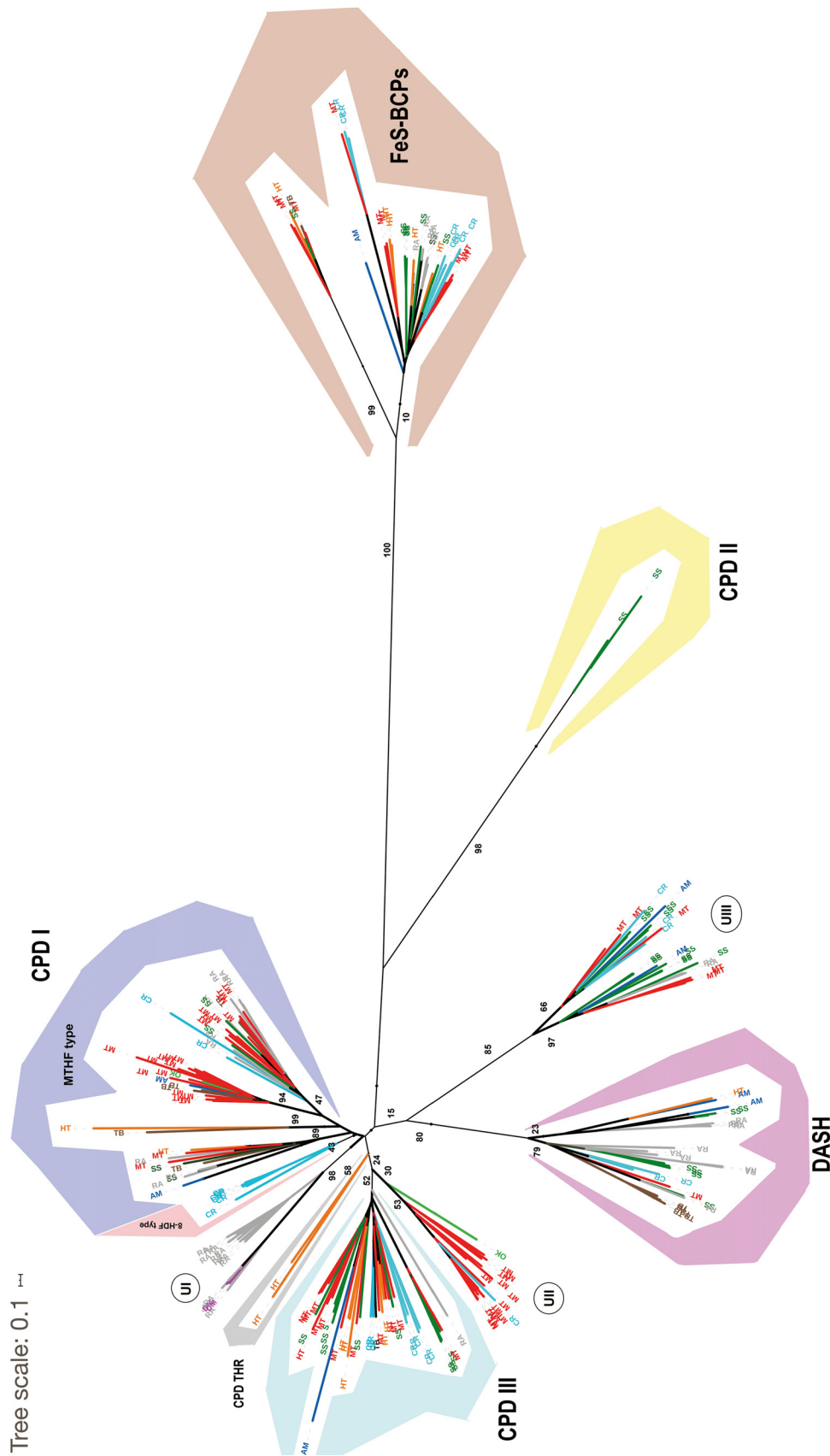


Fig. 4. Phylogenetic tree built with the cryptochrome/photolyase family sequences from the whole metagenomic data set. Bootstrap values for the main groups are shown. Each leaf is colored and labeled by sample location: DM: Lake Diamante (purple); SS: Socompa stromatolite (dark green); TB: Tibetan Plateau sediment (brown); AM: Amazon River (blue); RA: Lake Rauer hot spring (orange); CR: Greenland cryoconite (cyan); MT: Lake Monijoite (red); OK: Okluoto Island groundwater (light green). CPD I, II, and III: class I, II, and III cyclobutane pyrimidine dimer photolyases; CPD THR: thermostable CPD photolyase; DASH: DASH cryptochrome; FeS-BCPs: FeS bacterial cryptochromes and photolyases; UI, UII and UIII: unidentified group I, II and III, respectively; 8-HDF: 8-hydroxy-5-deazaflavin; MTHF: 5,10-methylenetetrahydrofolate

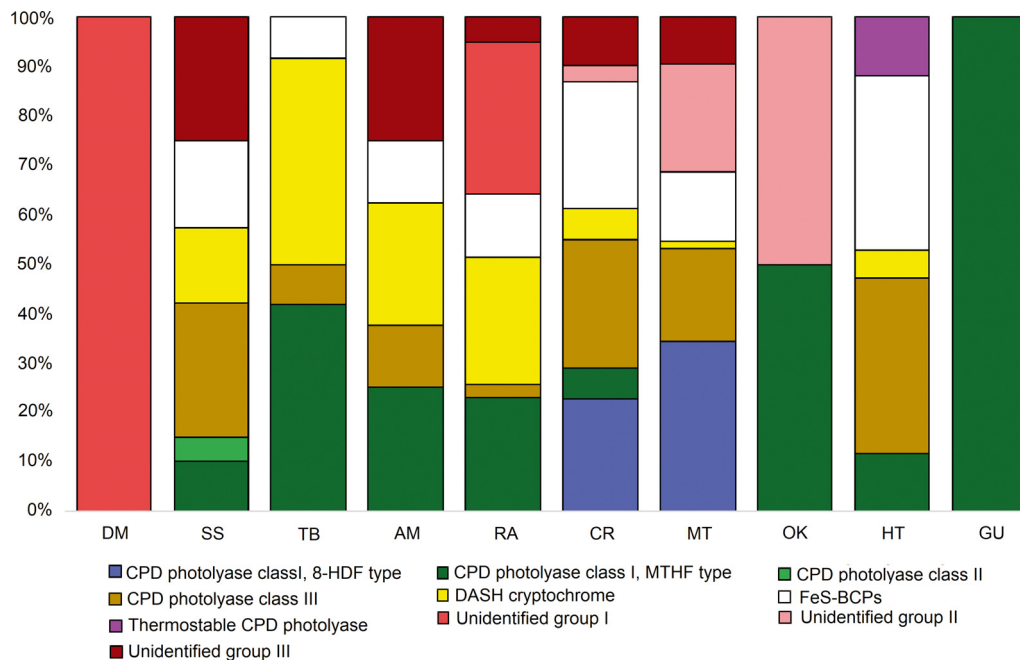


Fig. 5. Distribution of each cryptochrome/photolyase family subfamily by metagenome. See Fig. 2 for sample location abbreviations. CPD: cyclobutane pyrimidine dimer; see Fig. 4 for further definitions

quences of a single microbiome; sequences from the RA microbiome constitute 85.7 % of UI, while 87.5 % of UII corresponds to MT sequences. The UI subfamily is the only clade in DM but dominant in RA; both microbiomes displayed the highest content of CPF genes in this study. Furthermore, clades UI and UII are taxonomically homogeneous, with sequences belonging to *Halobacteria* and *Actinobacteria*, respectively, while clade UIII is formed by sequences of different taxonomic origins, including the phyla *Proteobacteria*, *Bacteroidetes*, and *Verrucomicrobia*.

The widespread presence of cryptochrome-holding subfamilies, cry-DASH and FeS-BCPs, among the microbiomes is evident (Fig. 6). The abundance of cry-DASH followed a pattern consistent with higher irradiation conditions; the highest number of cry-DASH sequences were found in the SS and TB communities (UV_{High}), AM (UV_{Mid}), and RA (the most prolonged photoperiod). It has been reported in previous work that these Cry are in fact photolyases with an affinity for single-stranded DNA and, in some cases, RNA (Selby & Sancar 2006). Cry-DASH may play a complementary role of standard photolyases with an affinity for double-stranded DNA, contributing to the global increase of photoreactivation in these microbiomes. The case for the FeS-BCPs group is intriguing, being present in almost all the communities with similar relative abundance. It has been suggested that this class complements other photolyases by per-

forming the function of a prokaryotic (6-4) photolyase (Zhang et al. 2013, Graf et al. 2015), in this way avoiding or decreasing the use of the inefficient NER system. In previous work on 3 isolated strains from the high-altitude Andean Lakes (where DM and SS originate) whose genomes were sequenced, we reported the presence of both CPD photolyases and members of the FeS-BCP subfamily in all genomes (Portero et al. 2019), even the presence of a third CPF member belonging to cry-DASH in one of the strains. Since the formation of the (6-4) photoproduct depends on the dose of UV radiation and increases with increasing dose reaching up to 40 % of the photoproduct fraction (Sancar 2003), we suggested the possibility that acquisition of prokaryotic (6-4) photolyases (included in the new group FeS-BCP) by bacteria from the high-altitude Andean Lakes is a compensatory adaptation to the high dose of UV radiation present in their original environment. Our present work provides further insight into the FeS-BCP subfamily by showing its ubiquitous and homogenous presence in world-wide microbiomes.

3.5. Relationship between CPF and taxa dominance

The assignment of CPF genes to a particular taxonomic group was also considered in this work.

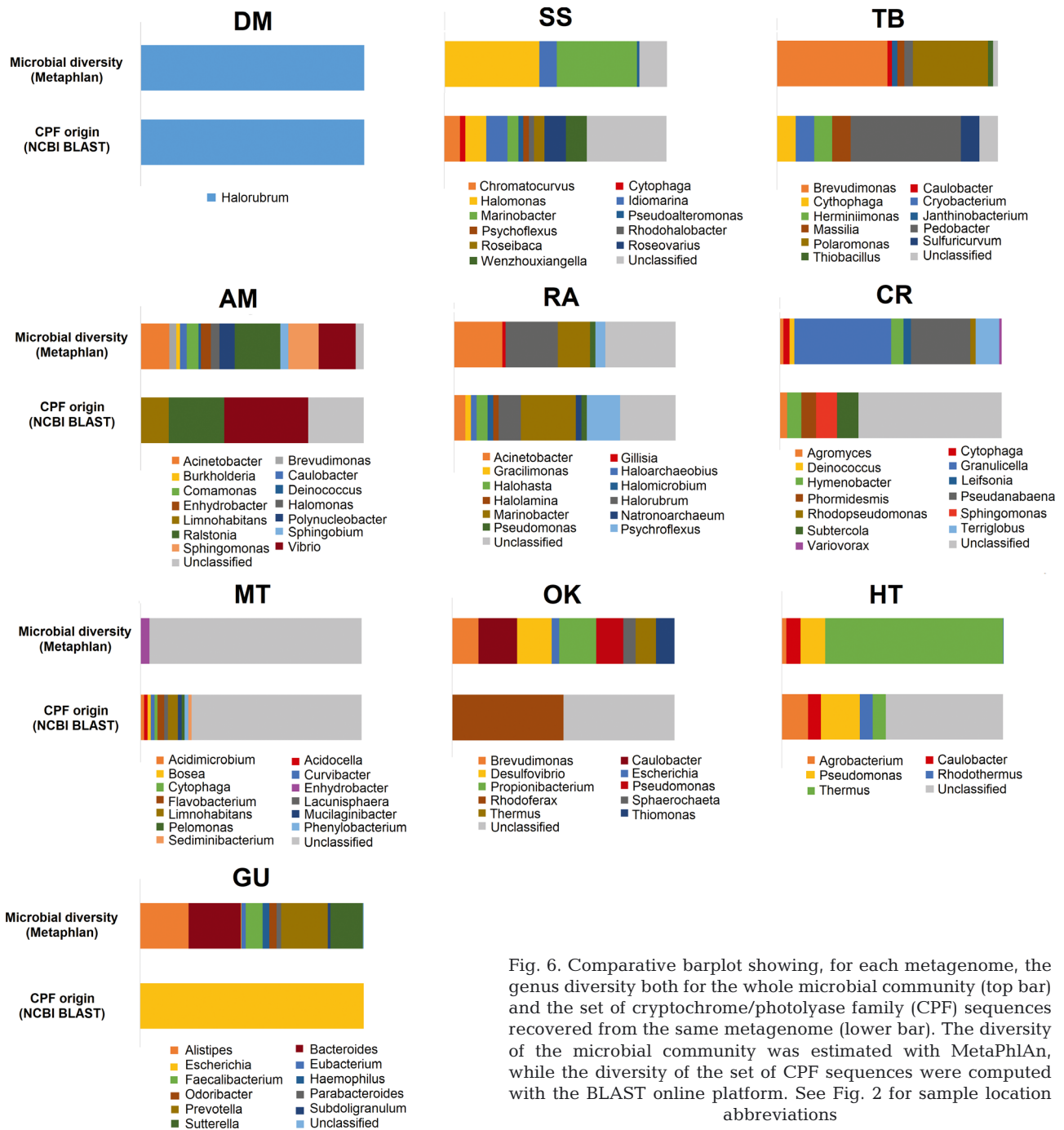


Fig. 6. Comparative barplot showing, for each metagenome, the genus diversity both for the whole microbial community (top bar) and the set of cryptochrome/photolyase family (CPF) sequences recovered from the same metagenome (lower bar). The diversity of the microbial community was estimated with MetaPhlAn, while the diversity of the set of CPF sequences were computed with the BLAST online platform. See Fig. 2 for sample location abbreviations

CPF sequences were classified by genus and paired with the same information previously obtained for each microbiome through MetaPhlAn. The CPF sequences grouped with their homolog with a BLAST identity lower than 70% clustered as an unclassified category. Genus diversity from the whole microbiome (top bar) and the set of CPF sequences recovered from that microbiome (lower bar) are shown in Fig. 6. The classification rate of

the CPF genes was generally high (>60%) in DM, SS, TB, AM, RA, and GU but lower in the rest of the samples. MT had the lowest percentage of sequences (23%) assigned to a genus. This phenomenon may be a consequence of the poorly referenced diversity of the community itself (top bar). Despite this, MT had the most substantial diversity of genera (12) with CPF hits along with SS and RA (11 each).

The most abundant taxa in the DM, SS, RA, and HT communities possess CPF genes. In AM, only *Ralstonia* and *Vibrio* displayed CPF; both genera together represented 35 % of the community. In the case of CR, only *Hymenobacter* and *Agromyces* possessed these genes, with both groups adding up to 7 %. *Pedobacter* was the unique taxon in TB with CPF genes, and it represented less than 4 % of the community. Finally, *Escherichia*, with an abundance of just 0.44 %, was the unique CPF contributor in GU. There were no matches found between the MetaPhlAn and BLAST classifications in the case of MT and OK.

These results show that CPF, and thus photoreactivation, is usually present in the most abundant organisms of the UV_{High} microbiomes (Fig. 6). Also, SS and RA show a high number of taxa in which photoreactivation is present. That suggests that photoreactivation may act as a successful system for assuring the survival and predominance of taxa in UV-stressed environments. This fact contrasts with those less isolated samples, where natural selection does not promote the diversification and dominance of taxa carrying CPF genes.

4. CONCLUSIONS

In this work, UV-B was considered for the first time as an ecological variable in a sequenced-based metagenomic study of microbial communities on a world-wide scale. Our results consistently showed concordance of high UV exposure of a given microbiome with its low microbial diversity and high CPF abundance.

We also evaluated CPF diversity in the world-wide microbiomes and reported 3 novel CPF clades not identified in previous analyses. CPF was more likely present in the most abundant organisms of the UV_{High} microbiomes, suggesting a significant evolutionary force for survival and dominance in highly irradiated environments. Also, cryptochrome-like genes were much more abundant in the most exposed microbiomes, indicating a complementary role to standard photolyses.

Finally, metagenomics proved to be an excellent tool to reveal a clear correspondence between microbiomes, UV-exposure, and UV-B resistance mechanisms, measured in the number of gene copies. Additional methods, such as metatranscriptomics and metaproteomics, should be implemented in order to unveil the molecular dynamics of the CPF upon different light conditions.

Acknowledgements. V.H.A. and M.E.F. are staff researchers from the National Research Council (CONICET) in Argentina. D.A. is a recipient of a doctoral fellowship from CONICET. The authors have produced this manuscript despite the delays and shortages 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 during the period 2016–2019. This manuscript has been released as a pre-print at bioRxiv (Alonso-Reyes et al. preprint doi:10.1101/701565).

LITERATURE CITED

- ✦ Aronesty E (2011) ea-utils: command-line tools for processing biological sequencing data. <https://expressionanalysis.github.io/ea-utils/>
- ✦ Aucamp PJ (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 ML, Thorpe PC, Thompson K, Paul ND, Grime JP, West HM (2004) Physical disturbance of an upland grassland influences the impact of elevated UV-B radiation on metabolic profiles of below-ground microorganisms. *Glob Change Biol* 10:1146–1154
- ✦ Ballaré CL, Caldwell MM, Flint SD, Robinson SA, Bornman JF (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, 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 AM, Šprtová L, von Wehrden H, Welk E, Cord AF (2014) gIUV: a global UV-B radiation data set for macroecological studies. *Methods Ecol Evol* 5:372–383
- ✦ Bomberg M, Lamminmäki T, Itävaara M (2016) Microbial communities and their predicted metabolic characteristics in deep fracture groundwaters of the crystalline bedrock at Olkiluoto, Finland. *Biogeosciences* 13:6031–6047
- ✦ Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60
- ✦ Chen B, Yuan K, Chen X, Yang Y and others (2016) Metagenomic analysis revealing antibiotic resistance genes (ARGs) and their genetic compartments in the Tibetan environment. *Environ Sci Technol* 50:6670–6679
- ✦ Doherty M, Yager PL, Moran MA, Coles VJ and others (2017) Bacterial biogeography across the Amazon River–ocean continuum. *Front Microbiol* 8:882
- ✦ Eloë-Fadrosch EA, Paez-Espino D, Jarett J, Dunfield PF and others (2016) Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat Commun* 7:10476
- ✦ 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 and others (2015) Key amino acids in the bacterial (6-4) photolyase PhrB from *Agrobacterium fabrum*. *PLOS ONE* 10:e0140955
- Häder DP, Gao K (2015) Interactions of anthropogenic stress factors on marine phytoplankton. *Front Environ Sci* 3:14
- ✦ Häder DP, Helbling EW, Williamson CE, Worrest RC (2011)

- Effects of UV radiation on aquatic ecosystems and interactions with climate change. *Photochem Photobiol Sci* 10:242–260
- ✚ Harmer SL (2009) The circadian system in higher plants. *Annu Rev Plant Biol* 60:357–377
- Hauptmann AL, Sicheritz-Pontén T, Cameron KA, Bælum J, Plichta DR, Dalgaard M, Stibal M (2017) Contamination of the Arctic reflected in microbial metagenomes from the Greenland ice sheet. *Environ Res Lett* 12:074019
- ✚ Hu B, Wang Y, 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 GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119
- ✚ Jansen MAK, Le Martret B, 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 DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8:275–282
- ✚ Joux F, Jeffrey WH, Lebaron P, Mitchell DL (1999) Marine bacterial isolates display diverse responses to UV-B radiation. *Appl Environ Microbiol* 65:3820–3827
- ✚ Kanai S, Kikuno R, Toh H, Ryo H, Todo T (1997) Molecular evolution of the photolyase–blue-light photoreceptor family. *J Mol Evol* 45:535–548
- Karentz D (1995) Considerations for evaluating ultraviolet radiation-induced genetic damage relative to Antarctic ozone depletion. *Environ Health Perspect* 102(Suppl 12): 61–64
- ✚ Kurth D, Amadio A, Ordoñez OF, Albarracín VH, Gärtner W, Farías ME (2017) Arsenic metabolism in high altitude modern stromatolites revealed by metagenomic analysis. *Sci Rep* 7:1024
- ✚ Lefort V, Desper R, Gascuel O (2015) FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol Biol Evol* 32:2798–2800
- ✚ Letunic I, 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
- ✚ Li D, Liu CM, Luo R, Sadakane K, Lam TW (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, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963
- Mitchell DL, Karentz D (1993) The induction and repair of DNA photodamage in the environment. In: Young AR, Björn LO, Moan J, Nultsch W (eds) *Environmental UV photobiology*. Springer, Boston, MA, p 345–377
- ✚ Nayfach S, Pollard KS (2015) Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol* 16:51
- ✚ Oksanen J, Blanchet FG, Friendly M, Kindt R and others (2018) vegan: community ecology package. R package version 2.5-2. <https://cran.r-project.org/web/packages/vegan/index.html>
- ✚ Pathak GP, Losi A, Gärtner W (2012) Metagenome-based screening reveals world-wide distribution of LOV-domain proteins. *Photochem Photobiol* 88:107–118
- ✚ Piccini C, Conde D, Pernthaler J, 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 LR, Alonso-Reyes DG, Zannier F, Vazquez MP, Farías ME, Gärtner W, Albarracín VH (2019) Photolyases and cryptochromes in UV-resistant bacteria from high-altitude Andean lakes. *Photochem Photobiol* 95:315–330
- ✚ Pushkarev A, Béjà O (2016) Functional metagenomic screen reveals new and diverse microbial rhodopsins. *ISME J* 10:2331–2335
- ✚ Pushkarev A, Inoue K, Larom S, Flores-Urbe J and others (2018) A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* 558: 595–599
- ✚ Rascovan N, Maldonado J, Vazquez MP, Farías ME (2016) Metagenomic study of red biofilms from Diamante Lake reveals ancient arsenic bioenergetics in haloarchaea. *ISME J* 10:299–309
- ✚ Rinnan R, Keinänen MM, Kasurinen A, Asikainen J and others (2005) Ambient ultraviolet radiation in the Arctic reduces root biomass and alters microbial community composition but has no effects on microbial biomass. *Glob Change Biol* 11:564–574
- ✚ Robinson SA, Turnbull JD, Lovelock CE (2005) Impact of changes in natural ultraviolet radiation on pigment composition, physiological and morphological characteristics of the Antarctic moss, *Grimmia antarctici*. *Glob Change Biol* 11:476–489
- ✚ Robson TM, Pancotto VA, Scopel AL, Flint SD, Caldwell MM (2005) Solar UV-B influences microfaunal community composition in a Tierra del Fuego peatland. *Soil Biol Biochem* 37:2205–2215
- ✚ Roenneberg T, Merrow M (2005) Circadian clocks—the fall and rise of physiology. *Nat Rev Mol Cell Biol* 6:965–971
- ✚ Ruhland CT, Xiong FS, Clark WD, Day TA (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 JM III, Luo M, Cicerone RJ, Deaver LE (1996) Satellite confirmation of the dominance of chlorofluorocarbons in the global stratospheric chlorine budget. *Nature* 379:526–529
- ✚ Sancar A (2003) Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem Rev* 103:2203–2237
- ✚ Satinsky BM, Fortunato CS, Doherty M, Smith CB and others (2015) Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011. *Microbiome* 3:39
- ✚ Searles PS, Flint SD, Caldwell MM (2001) A meta-analysis of plant field studies simulating stratospheric ozone depletion. *Oecologia* 127:1–10
- ✚ Selby CP, Sancar A (2006) A cryptochrome/photolyase class of enzymes with single-stranded DNA-specific photolyase activity. *Proc Natl Acad Sci USA* 103:17696–17700
- ✚ Singh AH, Doerks T, Letunic I, Raes J, Bork P (2009) Discovering functional novelty in metagenomes: examples from light-mediated processes. *J Bacteriol* 191:32–41
- ✚ Stibal M, Gözdereliler E, Cameron KA, Box JE and others (2015a) Microbial abundance in surface ice on the Greenland ice sheet. *Front Microbiol* 6:225

- ✦ Stibal M, Schostag M, Cameron KA, Hansen LH, Chandler DM, Wadham JL, Jacobsen CS (2015b) Different bulk and active bacterial communities in cryoconite from the margin and interior of the Greenland ice sheet. *Environ Microbiol Rep* 7:293–300
- ✦ Tran P, Ramachandran A, Khawasik O, Beisner BE, Rautio M, Huot Y, Walsh DA (2018) Microbial life under ice: metagenome diversity and *in situ* activity of *Verrucomicrobia* in seasonally ice-covered lakes. *Environ Microbiol* 20:2568–2584
- ✦ Truong DT, Franzosa EA, Tickle TL, Scholz M and others (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 12:902–903
- ✦ Tschitschko B, Erdmann S, DeMaere MZ, Roux S and others (2018) Genomic variation and biogeography of *Antarctic haloarchaea*. *Microbiome* 6:113
- ✦ Ueda T, Kato A, Kuramitsu S, Terasawa H, Shimada I (2005) Identification and characterization of a second chromophore of DNA photolyase from *Thermus thermophilus* HB27. *J Biol Chem* 280:36237–36243
- Weber S (2005) Light-driven enzymatic catalysis of DNA repair: a review of recent biophysical studies on photolyase. *Biochim Biophys Acta* 1707:1–23
- ✦ Westbrook A, Ramsdell J, Bergeron RD, Normington L, MacManes MD, Thomas WK, Schuelke T (2017) PALADIN: protein alignment for functional profiling whole metagenome shotgun data. *Bioinformatics* 33:1473–1478
- ✦ Williamson CE, Neale PJ, Hylander S, Rose KC and others (2019) The interactive effects of stratospheric ozone depletion, UV radiation, and climate change on aquatic ecosystems. *Photochem Photobiol Sci* 18:717–746
- ✦ Xiong FS, Day TA (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, Feng H (2012) The response of trichome mutants to enhanced ultraviolet-B radiation in *Arabidopsis thaliana*. *J Photochem Photobiol B* 113: 29–35
- ✦ Ye Z, Zhang N, Wu C, Zhang X and others (2018) A metagenomic study of the gut microbiome in Behcet's disease. *Microbiome* 6:135
- ✦ Zaller JG, Caldwell MMD, Flint S, Scopel AL, Salo OE, Ballaré CL (2002) Solar UV-B radiation affects below-ground parameters in a fen ecosystem in Tierra del Fuego, Argentina: implications of stratospheric ozone depletion. *Glob Change Biol* 8:867–871
- Zhang F, Scheerer P, Oberpichler I, Lampartner T, Krauß N (2013) Crystal structure of a prokaryotic (6-4) photolyase with an Fe-S cluster and a 6, 7-dimethyl-8-ribityllumazine antenna chromophore. *Proc Natl Acad Sci USA* 110: 7217–7222

Editorial responsibility: Michaela Salcher,
České Budějovice, Czech Republic

Submitted: May 6, 2020; Accepted: August 17, 2020
Proofs received from author(s): October 18, 2020